Cite this: *Soft Matter*, 2021,
17, 9772

A generic force field for simulating native protein structures using dissipative particle dynamics†

Rakesh Vaiwala *^a and K. Ganapathy Ayappa ^{ab}

A coarse-grained force field for molecular dynamics simulations of the native structures of proteins in a dissipative particle dynamics (DPD) framework is developed. The parameters for bonded interactions are derived by mapping the bonds and angles for 20 amino acids onto target distributions obtained from fully atomistic simulations in explicit solvent. A dual-basin potential is introduced for stabilizing backbone angles, to cover a wide spectrum of protein secondary structures. The backbone dihedral potential enables folding of the protein from an unfolded initial state to the folded native structure. The proposed force field is validated by evaluating the structural properties of several model peptides and proteins including the SARS-CoV-2 fusion peptide, consisting of α -helices, β -sheets, loops and turns. Detailed comparisons with fully atomistic simulations are carried out to assess the ability of the proposed force field to stabilize the different secondary structures present in proteins. The compact conformations of the native states were evident from the radius of gyration and the high intensity peaks of the root mean square deviation histograms, which were found to be within 0.4 nm. The Ramachandran-like energy landscape on the phase space of backbone angles (θ) and dihedrals (ϕ) effectively captured the conformational phase space of α -helices at $\sim(\phi = 50^\circ, \theta = 90^\circ)$ and β -strands at $\sim(\phi = \pm 180^\circ, \theta = 90-120^\circ)$. Furthermore, the residue-residue native contacts were also well reproduced by the proposed DPD model. The applicability of the model to multidomain complexes was assessed using lysozyme and a large α -helical bacterial pore-forming toxin, cytolysin A. Our study illustrates that the proposed force field is generic, and can potentially be extended for efficient *in silico* investigations of membrane bound polypeptides and proteins using DPD simulations.

Received 15th August 2021,
Accepted 29th September 2021

DOI: 10.1039/d1sm01194d

rsc.li/soft-matter-journal

1 Introduction

Molecular dynamics (MD) simulations have been extensively used to uncover the understanding of processes that occur at the molecular scale. The simulation technique has emerged as an indispensable computational tool to study a wide variety of systems with atomistic and mesoscopic details. Although accurate in its description, studying the structure and dynamics of large biomolecular complexes for longer times with atomistic resolution using all-atom simulations is computationally prohibitive. Coarse-grained (CG) approaches, with a reduction in the number of degrees of freedom to enhance conformational sampling, allow assessing processes that occur at physiologically relevant spatial and temporal scales. Consequently, there has been a rapid development of CG models and multiscale methods

in the last few decades for simulating simple fluids, ionic liquids, biomembranes, proteins and transmembrane protein assemblies.^{1–11}

With an increase in virulent bacterial and viral infections, CG modelling of biological systems involving membrane-protein complexes has attracted considerable attention.^{11–14} As a result, numerous CG protein models have been developed at various levels of reduced resolution ranging from one to five CG beads mapped to represent an amino acid residue.^{15,16} Depending on the level of description used to characterize the protein and solvent, CG models vary in their ability to capture protein folding, conformational dynamics and prediction of the native states. These CG models of proteins are structure-based,^{17–24} knowledge-based,^{25–30} and physics-based,^{31–37} with varying degrees of structural accuracy, model transferability and thermodynamic consistency. MARTINI, OPEP, PRIMO, UNRES, and SIRAH are among some of the popular protein force fields for CG-MD simulations.^{5,15,16,38–40}

Dissipative particle dynamics (DPD) is a particle based simulation method, which allows one to examine the underlying system at mesoscopic length and time scales.^{41,42} DPD has been extensively used to study biomembranes, polymers and

^a Department of Chemical Engineering, Indian Institute of Science, Bangalore 560012, India. E-mail: ayappa@iisc.ac.in

^b Centre for BioSystems Science and Engineering, Indian Institute of Science, Bangalore 560012, India

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1sm01194d

polyelectrolytes, and it has been shown to predict the scaling behaviour of several systems.^{43–51} However, DPD has been used to a lesser extent to study systems involving peptides and proteins. The DPD simulation framework is attractive for the simulation of proteins, since dynamics are naturally evolved in the canonical ensemble in the presence of explicit solvent particles.

In an early study, Vishnyakov *et al.*⁵² introduced a novel approach to model the secondary structures of polypeptides using DPD simulations. The α -helices and β -sheet structures were stabilized using dissociable Morse bonds between 1–3 and 1–5 amino acid residues in the protein sequence. The stiffness constant for the Morse bonds, the charge on the residues and the hydrophobicity were varied to capture a given secondary structure without the need for amino acid specific interaction potentials. The authors also pointed out that the presence of multiple Morse bonds due to their spherical symmetry could cause peptides to adopt unphysical conformations.⁵² Following the MARTINI scheme for assigning bead types to protein residues, a CG mapping scheme for amino acids has been proposed for simulating antimicrobial peptides using the DPD methodology.^{53,54} The repulsive force parameters for backbone beads are specific to the secondary structures being simulated, and the Morse potential between pairs of backbone beads has been used to stabilize α -helices and β -sheet structures.

The knowledge-based protein force field proposed by Peter *et al.*⁵⁵ is a polarizable model for simulating proteins in the DPD framework. A pair of oppositely charged Drude particles, which are transparent to other CG beads, has been employed to mimic electrostatic polarization of the peptide backbone. Being a polarizable force field, the protein model is computationally expensive. In a DPD based model developed by Truszkowski *et al.*⁵⁶ a protein force field has been presented with a CG mapping scheme based on underlying molecular fragments. For instance, the side chains of serine and valine were represented by the CG beads that correspond to methanol and propane fragments, respectively. In the absence of anisotropic interactions in the model, the structural rigidity of the protein was maintained by harmonic bonds between the backbone beads, similar to the elastic dynamic network models used in MARTINI simulations,⁵⁷ adding to the computational overhead. This hinders simulating protein folding from an unfolded state and studying secondary structure transitions.

In a very recent DPD study on the folding of polyaniline into its native structure, the α -helix was stabilized by optimizing the repulsive force constant for the alanine residues that form the helical structure using the native contacts derived from atomistic simulations.⁵⁸ The dihedral potential was used to stabilize the helical structure. Nevertheless, the parameter set is polyaniline specific, and therefore the parameters need to be re-optimized for different proteins with varying secondary structures.

The above studies reveal that although DPD models have been used to model specific protein structures, an efficient and generic protein force field for simulating native protein structures in the DPD formalism is yet to be developed. Additionally, it is

desirable to develop a protein force field which can be integrated with the existing DPD models for phospholipid membranes.^{43,53,59–63} In an attempt to bridge this gap, we present a complete set of force field parameters for 20 amino acids for simulating polypeptides and proteins using DPD. The native conformations, namely α -helices, rigid turns, flexible loops, β -hairpins and other extended structures, are preserved by introducing backbone dihedrals. Not restricted to a specific protein or protein structure, the proposed force field for bonded and non-bonded interactions is generic and can be applied to a wide range of peptides and protein structures. Unlike previous force fields developed for DPD simulations, the force field presented here does not require *a priori* knowledge of the presence of distinct secondary structures such as α -helices and β -sheets. Additionally the model is computationally inexpensive when compared with the polarizable protein models and the elastic network based models.^{55,56}

2 Simulation methodology

In this section the protocols for simulating the amino acid sequences in atomistic and CG frameworks are described. The peptides chosen for parametrizing and validating the DPD force field for proteins are depicted in Fig. 1. These peptides include the standard 20 amino acids, encompassing distinct structures such as α -helices, β -sheets, turns and flexible loops. Some of these peptides such as Trpzip (1LE1), Trpcage (1L2Y), 1HDN, 1J4M and lysozyme have been previously used for developing various protein force fields.^{7,55,58,64–67} The peptides shown in Fig. 1A are used to parametrize the DPD

Fig. 1 The peptides simulated for (A) parametrization and (B) validation of the CG force field are illustrated with their protein data bank (PDB) entries. Lysozyme is a multidomain protein with 129 residues comprising helical segments, H1 (pink), H2 (blue), H3 (cyan), H4 (red), H5 (purple), and H6 (silver), and an extended sheet structure (green). Cytolysin A (ClyA) is a pore-forming toxin containing structural segments, namely α A1 (cyan), α A2 (violet), α B (green), α C (orange), β -tongue (yellow), α F (blue), and α G (pink).

force field, while the peptides and proteins in Fig. 1B are used to test the robustness of the proposed force field. In addition to the smaller peptides we carried out DPD simulations of complex multidomain bacterial toxin cytolysin A (1QOY) and lysozyme (3TXJ) to demonstrate the versatility of the proposed force field (Fig. 1B). The monomer of cytolysin A (ClyA), a 34 kDa protein expressed by *Escherichia coli*, takes part in membrane assisted oligomerization during pore formation. We also modify the force field to describe disulfide bonds between cysteine residues while simulating the lysozyme. We used VMD software for molecular visualization.⁶⁸

2.1 Atomistic simulations

The peptides shown in Fig. 1 were simulated in water with the all-atom (AA) CHARMM36 force fields using the GROMACS (2018.2) MD engine.^{69,70} The CHARMM TIP3P water model was used for the aqueous environment, and periodic boundary conditions were used in the Cartesian directions. The details of the different systems investigated are given in Table 1, and data were acquired over 200 ns long simulations after equilibration for 10 ns. The simulations were performed at room temperature (303.15 K) using a Nosé–Hoover thermostat with a coupling time constant of 1 ps.⁷¹ The system pressure was maintained at 1 bar using a Parrinello–Rahman barostat with an isotropic pressure coupling time constant of 5 ps and a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$, which corresponds to water compressibility under ambient conditions.⁷² The bonds with H-atoms were constrained using the LINCS algorithm.⁷³ The equations of motion were integrated using the leap-frog algorithm with a step size of 2 fs. The non-bonded van der Waals interactions were computed within a cutoff distance of 1.2 nm, and the forces were smoothly truncated to zero over a distance from 1.0 to 1.2 nm. The long range electrostatic interactions for charged species were evaluated using the Particle Mesh Ewald (PME) method with a real-space cutoff distance of 1.2 nm.⁷⁴ Na^+ and Cl^- ions were added to neutralize charges on the peptides.

2.2 Dissipative particle dynamics simulations

We use the standard formulation of the DPD method,^{41,42,59,75} where suitable molecular fragments or beads represent the level of coarse-graining. The DPD beads interact with each other *via* the centrally directed pairwise forces. The soft repulsive conservative force f^C between DPD particles i and j positioned

at $\mathbf{r}_i$ and $\mathbf{r}_j$ is given by,

$$\mathbf{f}_{ij}^C = a_{ij} \left(1 - \frac{r_{ij}}{R_c} \right) \mathbf{e}_{ij} \quad (1)$$

where a_{ij} is the repulsive force constant and R_c is the cutoff radius. The radial separation between the particles i and j is $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, and the unit vector $\mathbf{e}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)/r_{ij}$. The DPD particles also interact through pairwise dissipative f^D and stochastic f^R forces given by,

$$\mathbf{f}_{ij}^D = -\gamma w_D(\mathbf{e}_{ij} \cdot \mathbf{v}_{ij}) \mathbf{e}_{ij} \quad \text{and} \quad \mathbf{f}_{ij}^R = \sigma w_R \zeta_{ij} \mathbf{e}_{ij}. \quad (2)$$

The relative velocity $\mathbf{v}_{ij}$ is $\mathbf{v}_i - \mathbf{v}_j$, while ζ_{ij} is a random variable drawn from a Gaussian distribution with zero mean and unit variance. The weight function $w_R(r) = 1 - (r/R_c)$. The strengths of the dissipative and stochastic forces are γ and σ respectively, and they are governed by the fluctuation–dissipation theorem,⁷⁶

$$\sigma^2 = 2\gamma kT \quad \text{and} \quad w_D = w_R^2 \quad (3)$$

resulting in a thermostat for the DPD simulation. Accordingly, the noise amplitude $\sigma = 3.0$ and the frictional strength $\gamma = 4.5$ are typically used for the DPD thermostat. The energy scale is kT , where k and T denote the Boltzmann constant and absolute temperature, respectively. The equations of motion for the DPD particles are integrated using the velocity Verlet scheme.^{42,77}

2.3 Electrostatics in DPD simulations

The charges assigned to the side chains for some of the amino acids are illustrated in Fig. 2. In addition, peptides were assigned a charge of $+1e$ at the N-terminus and $-1e$ at the C-terminus.^{55,64} Electroneutrality is maintained by addition of ions (see Table 1). Electrostatic interactions among the charged DPD particles can be evaluated either by locally solving the electrostatic field on spatially distributed grids⁷⁵ or by using the Ewald summation.^{78–80} A comparative assessment of these methods is detailed elsewhere.⁸¹ We computed the electrostatic potential Φ by solving iteratively the Poisson equation on real-space grids,⁷⁵

$$\nabla \cdot (\epsilon_r(\mathbf{r}) \nabla \Phi(\mathbf{r})) = -\rho_c(\mathbf{r})/\epsilon_0. \quad (4)$$

The constants ϵ_0 and ϵ_r are the dielectric permittivity of vacuum and local relative dielectric constant, respectively. We used a uniform dielectric constant, $\epsilon_r = 78$. The computational domain is discretized into three dimensional cells with a mesh size R_c . The Coulomb charges on the DPD particles are smeared out over the grid points within a distance of $R_e = 1.6R_c$ using a linear charge distribution function,

$$f(r) = \frac{3}{\pi R_c^3} \left(1 - \frac{r}{R_c} \right). \quad (5)$$

Following Groot,⁷⁵ the grid-based electrostatic potential, $\Phi(r)$, is obtained by solving the discretized version of the Poisson equation (eqn (4)) using the successive relaxation method. The electrostatic force on each particle is calculated

Table 1 Details for different systems simulated using all-atom and DPD descriptions

Peptide	All-atom simulation		DPD simulation	
	Peptide atoms	Water (ions)	Peptide beads	CG water (ions)
1HDN	233	6944 (1 Na^+)	38	2961 (+1)
1J4M	236	6944 (2 Cl^-)	34	2964 (−2)
6VSB	410	6887 (2 Na^+)	64	2934 (+2)
1L4X	260	6939	41	2959
Trpzip	218	6947 (1 Cl^-)	34	2965 (−1)
Trpcage	304	6929 (1 Cl^-)	42	2957 (−1)
Lysozyme	1960	55 963 (8 Cl^-)	301	9816 (−8)
ClyA	4792	24 787 (7 Na^+)	749	31188 (+7)

Fig. 2 Schematic showing the mapping of amino acids onto the CG beads illustrated as shaded boxes. The backbone CG beads are centered on the C_α atoms of the amino acid residues and charges are illustrated for some of the side chains. The residues GLY and PRO are represented by a single bead. The bead type ascribed to the backbone bead is P2, except for the PRO residue, which has particle type P3. The other side chain beads have particle types, P4–P17 as labelled in the schematic. The repulsive force parameters for the different particle types are given in Table 2.

by interpolating the gradient of the potential using the same distribution function $f(r)$.

The system details including the number of DPD water beads are provided in Table 1. All DPD simulations were carried out at room temperature and constant volume with periodic boundary conditions in three Cartesian directions. A time step of 0.01τ was used for integrating the equations of motion. The time scale of $\tau \sim 88$ ps was deduced by mapping to the diffusion coefficient of water.⁵⁹ In reduced DPD units the length R_c , time τ , energy kT and masses m of the CG particles were set to unity. Each DPD trajectory for peptides 1HDN, 1J4M and 6VSB, which are used for parametrization, consists of 2×10^5 time steps with a step size of 0.01τ , while each DPD simulation for 1L4X, Trpzip and Trpcage is comprised of 10^6 time steps. We used the last half of the trajectory for analysis in each case. The DPD force field developed using the different peptides (Fig. 1) is given next.

3 DPD models for water and proteins

The DPD bead for the representation of water effectively comprises of three water molecules with the length scale $R_c = 0.646$ nm, which is equivalent to the volume of three water molecules at room temperature.⁵⁹ With this three-to-one CG mapping scheme, the force constant for soft repulsive interactions between water–water is $a_{ww} = 78 kT/R_c$, which corresponds to the isothermal compressibility of water at room temperature.⁵⁹

Amino acids are the building blocks of peptides and proteins, whose secondary structures are highly correlated with the backbone dihedral angles.^{82–84} Therefore, in the reduced

representation shown in Fig. 2, we map the backbone atoms of the protein onto the CG backbone beads (B), with the centers of these beads located at the positions of the C_α atoms.⁶⁴ The mapping scheme we follow in fragmenting the side chains into CG beads allows us to directly adopt the Flory–Huggins (χ) parameters used by Truszkowski *et al.*⁵⁶ for interactions between the CG beads. The different bead types for the amino acids are labelled from P2–P17 as illustrated in Fig. 2 where the backbone bead ‘B’ corresponds to the bead type P3 for PRO and P2 for all other residues. For notational convenience we refer to the side chain beads as ‘S1’, ‘S2’ and ‘S3’ labelled progressively outward from the backbone bead. The residues GLY and PRO consist only of backbone beads, while TRP with bulky aromatic groups is represented by one backbone bead and three side chain beads. For example P4, P14 and P12 in the TRP residue represent the side chain beads S1, S2 and S3, respectively, while the remaining 17 amino acid residues have either one (S1) or two (S1 and S2) side chain beads.

The bonded interactions between the different beads are modelled using harmonic potentials with appropriate spring constants,

$$U^B(r) = \frac{1}{2}k_b(r - r_0)^2 \quad (6a)$$

$$U^A(\theta) = k_\theta(\theta - \theta_0)^2 \quad (6b)$$

$$U_{\text{backbone}}^A(\theta) = k_\theta(\theta - \theta_1)^2(\theta - \theta_2)^2 \quad (6c)$$

where $U^B(r)$ is the bonded potential with equilibrium bond length r_0 , $U^A(\theta)$ is the angle bending potential with equilibrium angle θ_0 , while the corresponding spring constants are k_b and k_θ respectively. Here $U_{\text{backbone}}^A(\theta)$ represents a dual-basin angle

potential (eqn (6c)) between three consecutive beads (B–B–B) along the backbone with the equilibrium angles $\theta_1 = 90^\circ$ and $\theta_2 = 120^\circ$, which correspond to the backbone angles for α -helix and β -sheet structures, respectively. Covering a wider range of backbone angles and having a negligible energy barrier between two conformational minima on the energy landscape, the dual-basin force field potentially allows secondary structure transitions in proteins. Moreover, it energetically favours a protein to attain the folded native structure from an unfolded state. This parametrization is an improvement and a departure from the existing DPD force fields reported in the literature. Additionally we incorporate the backbone dihedral potential,

$$U^D(\phi) = k_\phi(1 - \cos(\phi - \phi_0)) \quad (7)$$

where the torsion angles along the protein backbone are maintained at the native state ϕ_0 using a cosine potential with the stiffness constant k_ϕ . The dihedral potential has recently been used to stabilize the α -helical structure in polyaniline.⁵⁸

4 Model parametrization

The repulsive force constants a_{ij} among the protein beads and protein–water interactions are computed using the Flory–Huggins energy parameters χ_{ij} according to,⁴²

$$a_{ij} = a_{\text{ww}} + 3.497\chi_{ij}. \quad (8)$$

As mentioned earlier, we have adopted the Flory–Huggins parameters for the DPD beads of proteins from the work Truszkowski *et al.*,⁵⁶ where molecular fragments were simulated in an aqueous solvent using the COMPASS force field.⁸⁵ The full interaction matrix of the repulsive force constant evaluated from eqn (8) is tabulated in Table 2, where P_i s represent the different bead types as described in Fig. 2. We note that ions interact with the protein and other CG beads with the same repulsive strength as the CG water, and hence ions are assigned the P1 type bead.

Since many DPD models for lipid membranes use a repulsive force with $a_{\text{ww}} = 78 \text{ kT}/R_c$, our model for the protein can directly be integrated with the existing phospholipid force fields to investigate protein–membrane complexes.^{53,54,59,86–89}

Using a bottom-up approach for coarse-graining, the bonded parameters, namely the equilibrium bond lengths and angles (r_0 and θ_0 in eqn (6)), are deduced from the centers of reference distributions for the virtual CG bonds and angles derived from AA trajectories. We used an iterative procedure to optimize the bonded interactions. In order to use a higher time step in DPD simulations, the bond stiffness k_b is restricted to $2000 \text{ kT}/R_c^2$, and a minimum value of k_b is set to $1500 \text{ kT}/R_c^2$ to control the spread of the distributions. The stiffness for the angle bending potential is chosen to be 10, 15 or 20 kT rad^{-2} depending on the spread of the angles in the reference distributions. During the iterative procedure, bond lengths ($X = r_0$) and the angles ($X = \theta_0$) in CG distributions (X^{CG}) are updated, according to $X = X + (X^{\text{AA}} - X^{\text{CG}})$, with a convergence tolerance of 0.005 nm in bond lengths and a maximum 5° deviation in angles from the centers of the corresponding reference distributions (X^{AA}). The optimized CG bond lengths and the bond stiffness values are given in Table 3. With $R_c = 0.646 \text{ nm}$, the CG bond length of $0.58R_c$ between two consecutive backbone beads (B–B) corresponds to a distance of 0.38 nm between two consecutive C_α atoms in the AA protein model. The bond length between S1 and S3 beads of TRP is also constrained to $0.64R_c$ for controlling angles in the TRP residue.

Fig. 3 shows the distributions for bonds in the different amino acid residues compared with the corresponding target distributions obtained from atomistic simulations. The CG distributions obtained in DPD simulations are inherently broader due to the soft repulsive nature of the DPD interactions. However, an excellent overlap is observed with the mean values coinciding well with the AA target distributions. For residues LYS and ILE, the bond distributions are bimodal in the AA simulations, while the harmonic bond potential in DPD captures one of the peaks.

Table 2 Repulsive force constants, a_{ij} , evaluated from eqn (8) for interactions between the different protein beads, P2–P17, as illustrated in Fig. 2 and the water bead, P1

↓ Bead type black →	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17
P1	78	78	78	91	93	94	90	73	70	77	74	91	76	83	68	77	76
P2		78	72	80	77	77	76	72	75	71	74	75	71	72	72	75	72
P3			78	78	75	74	73	73	73	72	72	73	67	69	69	74	71
P4				78	73	74	73	80	82	81	83	76	80	78	96	101	81
P5					78	72	72	78	82	78	81	73	78	75	95	100	79
P6						78	72	78	82	78	81	72	77	74	97	101	79
P7							78	76	81	77	80	71	75	73	92	93	77
P8								78	72	71	73	77	69	72	63	68	73
P9									78	75	72	81	72	78	66	71	73
P10										78	75	76	72	73	72	73	73
P11											78	80	71	76	67	73	73
P12												78	75	72	93	94	76
P13													78	73	71	72	68
P14														78	82	80	71
P15															105	56	70
P16																105	74
P17																	78

Table 3 Bonded interaction parameters between the backbone beads, B, and the side chain beads S1, S2 and S3 for the different amino acid residues illustrated in Fig. 2

Residue	Bonds		Angles	
	(B-S1, S1-S2)		(B-S1-S2)	
	Bond length r_0 (R_c)	Stiffness k_b ($kT R_c^{-2}$)	Angle θ_0 (degree)	Stiffness k_θ ($kT \text{ rad}^{-2}$)
ALA	0.24	2000	—	—
ARG	0.38, 0.56	1500, 1500	140	10
ASN	0.40	1500	—	—
ASP	0.39	1500	—	—
CYS	0.37	1500	—	—
GLN	0.24, 0.41	2000, 1500	90	15
GLU	0.24, 0.40	2000, 1500	140	15
HIS	0.24, 0.43	2000, 1500	110	20
ILE	0.31, 0.39	1500, 1500	80	20
LEU	0.24, 0.33	2000, 1500	120	20
LYS	0.37, 0.48	1500, 1500	140	15
MET	0.31, 0.40	1500, 1500	110	15
PHE	0.24, 0.48	2000, 1500	110	20
SER	0.31	1500	—	—
THR	0.31	1500	—	—
TRP	0.24, 0.43	2000, 1500	115	20
TRP	0.35 (S2-S3)	1500	108	20 (S1-S2-S3)
TRP	0.64 (S1-S3)	1500	—	—
TYR	0.24, 0.55	2000, 1500	110	20
VAL	0.31	1500	—	—
Backbone	0.58 (B-B)	1500	115	15 (B-B-S1)

Fig. 4 indicates the distributions of angles within the different amino acid residues, namely B-S1-S2, and the B-B-S1

angle across two consecutive residues. The B-B-S1 angle is set to 115° . The S1-B-B angle (not shown) comprising first two residues at the N-terminus is constrained to $\theta_0 = 100^\circ$ with the stiffness constant $k_\theta = 15 \text{ kT rad}^{-2}$. On account of having three CG beads in its side chain, the TRP residue has an additional angle S1-S2-S3 peaked at $\theta_0 \sim 108^\circ$. According to eqn (6c), the energy landscape for the backbone angle B-B-B captures dual basins, 90° and 120° , encompassing the α -helices as well as β -extended structures which will be illustrated later in the text. The stiffness constant k_θ is set to 15 kT rad^{-4} for the B-B-B backbone angle. Table 3 provides the angle bending parameters.

Stabilization of the native conformations of peptides is achieved through the backbone dihedrals. The torsion angles ϕ_0 between four consecutive C_α atoms along the protein backbone are computed from the native crystal structure, and maintained in DPD simulations using the torsion potential (eqn (7)) with the stiffness constant $k_\phi = 10 \text{ kT}$. The values for ϕ_0 are provided in Table S1 (ESI†).

5 Model validation

The proposed DPD parameters for protein simulations are tested by evaluating various protein structural metrics such as the root mean square deviation (RMSD), radius of gyration (R_g), residue-residue contacts and the Ramachandran-like maps.

Fig. 3 Distributions for bond lengths in amino acid residues. The solid and dotted curves indicate the distributions obtained in AA and DPD simulations, respectively, for peptides 1HDN, 1J4M and 6VSB. The B-S1 (red) bond refers to the bond between the backbone bead B and side chain bead S1, while S1-S2 (blue) is the bond between S1 and S2 beads in the side chains. The TRP residue comprises of three beads in the side chain, and thus has an additional S2-S3 bond (maroon). The distribution along the peptide backbone is B-B (red). The scale on the x-axis is the bond length in nm.

Fig. 4 Distributions for angles in amino acid residues obtained from AA (—) and DPD (---) simulations of peptides 1HDN, 1J4M and 6VSB. Within amino acids, there are B-S1-S2 (blue) and S1-S2-S3 (red in TRP) angles. The distribution for the angle across two consecutive residues, B-B-S1, is obtained by averaging over amino acids in a peptide.

The DPD force field is validated using peptides 1L4X (α -helix), Trpzp (β -hairpin) and Trpcage (helix + loop), where we make a direct comparison with the atomistic simulations thereby providing a stringent test to the accuracy of the proposed force field.

5.1 Structural similarity

To measure the similarity between the protein conformations and their native structure, we have calculated the root mean

square deviation, $\text{RMSD}(\mathbf{r}, \mathbf{r}') = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i - \mathbf{r}'_i)^2}$, where N

Fig. 5 Color maps for the joint probability histograms P for the root mean square deviation (RMSD) and radius of gyration on the log-normal scale of $-\log(P/P_{\max})$, where $P_{\max}$ is the maximum value of P in histograms. The histograms (A–C) refer to the AA simulations for peptides 1L4X (A), Trpzp (B) and Trpcage (C), while their DPD counterparts are shown in panels D, E, and F respectively.

denotes the number of amino acid residues in the peptide sequence. While $\mathbf{r}$ and $\mathbf{r}'$ are the position vectors of backbone beads (C_α atoms for the AA model) in the simulated trajectory and in the native PDB structure, respectively. The ability of the DPD model to retain the level of compactness of the protein structure close to that of the native structure is determined by

computing the radius of gyration, $R_g(\mathbf{r}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i - \mathbf{r}_m)^2}$,

where $\mathbf{r}_m$ is the center of mass coordinates of the protein backbone.

Fig. 5 shows the joint probability histograms of RMSD and R_g for peptides 1L4X (α -helix), Trpzp (β -hairpin) and Trpcage (helix + loop). The color maps are represented as $-\log(P/P_{\max})$ where $P_{\max}$ is the maximum value of P in a given distribution. Hence the intensity values which are close to zero indicate higher probabilities in the distributions. From the high probability regions the radius of gyration is accurately captured in DPD simulations. For instance, the distribution for the Trpzp peptide peaks at $R_g \sim 0.6 \pm 0.001$ nm in the DPD model (Fig. 5E), and this agrees well with $R_g \sim 0.58 \pm 0.001$ nm in the high probability region for the AA model (Fig. 5B).

The RMSD variations show some interesting trends across the different proteins. For the α -helical peptide 1L4X, the range of the RMSD variations is similar, with the DPD model displaying a slightly lower RMSD when compared with the AA data. In the case of Trpzp (β -hairpin) and Trpcage (helix + loop), the RMSD values obtained from the AA simulations are smaller when compared with the DPD simulations.

Although the RMSD of the peptides in the DPD simulations shows a slightly higher spread when compared to a relatively narrow range observed in the AA simulations, the conformations sampled at RMSD ranging from 0.1–0.3 nm represent the energetically favourable configurations that are sampled to a greater extent (Fig. 5). Furthermore, if one examines the most probable values for RMSD, the differences are within 0.2 nm between DPD results and the reference AA RMSD data. Additionally, the range of the RMSD sampling is superior to a previous polarizable model,⁵⁵ which reported RMSD in the range of 0.5–0.7 nm for peptides Trpzp and Trpcage. For peptides 1HDN, 1J4M and 6VSB, which were used for parametrization, the RMSD and R_g data are provided in Fig. S1 (ESI†).

5.2 Native contacts

The success of a CG force field for proteins relies on its ability to retain the residue–residue contacts that characterize a given native structure. This can be achieved by constraining the contacts using the harmonic potential (Gö and elastic network models) or the Lennard-Jones potential (Gö-MARTINI).^{17,38,57,90} In the DPD model presented here, the secondary structures are stabilized with the dihedral (eqn (7)) and the dual-basin angle bending potentials (eqn (6c)). Unless stated otherwise, we computed the contact maps using a distance criterion between a pair of residues lying within 1 nm for both the AA and DPD simulations.

As illustrated in Fig. 6, where contact maps with the color bars indicate the fraction of a trajectory having stabilized

Fig. 6 The residue–residue contacts with a cutoff distance of 1.0 nm are shown for peptides 1L4X (A and D), Trpzp (B and E) and Trpcage (C and F). The contact maps obtained from the all-atom simulations are indicated in panels (A–C), while those derived from the DPD simulations are shown in panels D–F.

Fig. 7 Ramachandran-like plots for the model peptides 1L4X (A and D), Trpzip (B and E) and Trpcage (C and F), showing in color maps the probability density $P(\phi, \theta)$ as $-\log(P(\phi, \theta)/P_{\max})$ on the phase space of backbone angle θ and dihedral angle ϕ . The maximum density is $P_{\max}$. The top panels (A–C) are constructed from AA simulations, while the panels at the bottom (D–F) are the results from the DPD simulations. The angles are represented in degrees.

contacts, the residue–residue contacts are well captured in the helical peptides (Fig. 6A and D). The comparisons for the peptides with β -hairpins (Fig. 6B and E) and the flexible loop structures (Fig. 6C and F) are also satisfactory. The discrepancies in contact maps between the AA and DPD models for these structures are attributed to the non-local excluded volume interactions, which are soft repulsive in the DPD force

field (eqn (1)). The soft repulsive nature of the DPD force field is able to capture all the dominant contact pairs, albeit with reduced intensity in some cases. For example in the Trpzip structure (Fig. 6B and E) the contacts between residues 12 with 1–4 are captured with a lower intensity in the DPD simulations. Although an increased cutoff for the contacts in the DPD model can reduce the differences, we prefer to make comparisons with

Fig. 8 Folding of peptides (A) α -helix 1L4X, and (B) β -hairpin Trpzip to their native states. The insets show the initial unfolded and the final folded structures. The backbone beads are colored in blue, while the side chain beads are illustrated in orange. The contacts between the backbone beads were calculated using a distance cutoff of 1.3 times the distance ' d ' between C_α atoms in the native AA structure. A slightly higher cutoff of $1.3d + 0.2$ nm allows the β -hairpin to recover more than 90% of the native contacts.

the AA simulations with a similar cutoff criterion. Fig. S2 in the ESI† provides the contact maps for peptides 1HDN, 1J4M and 6VSB. The native contacts are further quantified by computing the order parameter⁹¹ $Q = \frac{1}{n} \sum_{j \geq i+3}^n \exp\left(-\left(\frac{|r_{ij} - r_{ij}^{\text{native}}|}{3}\right)^2\right)$.

The summation runs over all non-local contacts between pairs of residues i and $j \geq i + 3$, with distances r_{ij} in simulated trajectories and r_{ij}^{native} in native PDB structures. The time averaged Q values from the DPD simulations are 0.79, 0.63 and 0.66 for 1L4X, Trpzip and Trpcage, respectively (Fig. S3 in the ESI†). The Q values lying in the range of $0.55 \leq Q \leq 1$ provide additional confirmation for the inherent stability of the structures in the proposed DPD force field.

5.3 Ramachandran conformational space

Equivalent to an energy surface on the $\phi - \psi$ plane formed by two dihedral angles in atomistic models, the Ramachandran-like plot in a reduced CG representation is a two-dimensional density plot of the backbone angle θ and torsion angle ϕ .⁹² The Ramachandran-like maps for peptides 1HDN, 1J4M and 6VSB are given in Fig. S4 of the ESI†. Referring to Fig. 7A and D,

dominant sampling of the conformations occurs at $\theta = 90^\circ$ and $\phi = 50^\circ$ for the helical peptide 1L4X as well as for the peptide Trpcage (Fig. 7C and F), which have helices at the N-terminus (residues 1–9). The increased sampling centered around $\phi = -120^\circ$ accounts for the turn regions in peptides Trpzip (Fig. 7B and E) and Trpcage (Fig. 7C and F). The presence of β -sheets in peptide Trpzip (Fig. 7B and E) is evident from high density spots at $\phi \sim \pm 180^\circ$. Correlating the conformational space mapped onto the ϕ and θ angles with the observed secondary structures, our data are consistent with the density maps in these reduced coordinates given in earlier CG-MD studies.⁹²

5.4 Protein folding

The robustness of the proposed force field is tested by allowing the peptide sequences to fold back to their three dimensional native structures. We carried out folding simulations on peptides 1L4X (α -helix) and Trpzip (β -hairpin). The initial unfolded configurations were obtained by switching off the dihedral potentials and setting the equilibrium backbone angles to 180° . With ' d ' being the distance between C_α atoms in the

Fig. 9 Lysozyme simulation results. (A) The residue–residue contact map in DPD simulation with a cutoff distance of 1 nm, showing the contacts especially in the structural regions, namely six helices H1–H6 and the extended sheet. The arrows indicate the non-local contacts in the vicinity of the disulfide linkages. (B) The contact map obtained from the AA simulation of the lysozyme with a distance cutoff of 1 nm. (C) The radius of gyration (R_g) in the DPD simulation stabilizes at $R_g \sim 1.5$ nm, while it is quite stable at $R_g \sim 1.4$ nm in the AA simulation shown in the inset. The time in DPD simulations is in units of τ . (D) The simulation snapshot of the lysozyme (backbone) is illustrated with structural segments, and the color scheme is similar to that used for the lysozyme in Fig. 1.

native state in the AA model, a pair of residues in the DPD model is said to be in contact when the distance between the backbone beads is within $1.3d$.⁹⁰ However, due to the purely repulsive nature of the DPD forces and relatively higher flexibility of the β -hairpin in the absence of hydrogen bonding, a cutoff of $1.3d + 0.2$ nm was able to capture more than 90% of the native contacts in the β -hairpin structure. Fig. 8 illustrates the time evolution of native contacts averaged over 10 independent folding simulation trajectories for each peptide. It is clear from the folding curves in Fig. 8 that the characteristic folding times for α -helices and the β -hairpins are of similar order ($\sim 10\tau$), which we attribute to the equilibration time scale for the dihedral angles. Unlike the ten-fold faster folding of α -helices over the β -hairpin structures observed for atomistic as well as Gō-MARTINI models in previous studies,^{90,93} the folding for the α -helix is only slightly faster than that of the β -hairpin in the present DPD model.

5.5 Multidomain proteins

The generality of the proposed DPD protein model for the *in silico* study of multidomain proteins is further examined by simulating larger protein structures. We chose the lysozyme and

cytolysin A (ClyA) for this purpose, and describe simulations first for the lysozyme followed by ClyA in this section. The simulations were carried out for 10^6 time steps after an initial equilibration period of 10^5 steps, and the last half of the trajectory was used for calculating the structural properties for both the lysozyme and ClyA. The lysozyme (PDB: 3TXJ) is employed here to illustrate the manner in which we incorporate the disulfide linkages present between pairs of CYS residues. Lysozyme comprises of 129 amino acid residues, with six small helical domains, H1 (residues: 5–15), H2 (25–36), H3 (80–83), H4 (88–102), H5 (109–115), and H6 (120–124), as well as extended structures (residues: 43–45, 51–53, and 58–59) as depicted in Fig. 1 and 9D. The four pairs of CYS residues (residues 6 and 127, 30 and 115, 64 and 80, and 76 and 94) are involved in forming four disulfide linkages in the lysozyme. In order to incorporate the disulfide bonds in the DPD model, we used a harmonic potential (eqn (6a)) between these pairs of residues with a bond length of $r_0 = 0.31R_c$ (0.2 nm in the AA reference model) and a bond stiffness, $k_b = 1500 kT R_c^{-2}$.

The time evolution of the RMSD of the structural domains of the lysozyme (Fig. S5 in the ESI†) indicates that the six helices and the extended strands retain their native structures

Fig. 10 The performance of DPD force field for the cytolysin A (ClyA) monomer in water and its comparison with AA simulation results. (A) The residue–residue contact map with a cutoff distance of 1.2 nm between pairs of backbone beads in the DPD simulation. (B) Contact maps constructed from AA simulations for the pairs of C_α atoms with a distance cutoff of 1.0 nm. (C) The final snapshot from the DPD simulation indicates the backbone of ClyA, with the color scheme similar to that used in Fig. 1 for the C_α atoms of ClyA. (D) Root mean square fluctuations of the backbone bead positions for ClyA from both the DPD and AA simulations.

at RMSD ~ 0.3 nm in the DPD simulation. Due to the presence of flexible loops and turns, the overall structure of the entire lysozyme showed a higher RMSD of 0.937 ± 0.003 nm when compared with a RMSD of 0.117 ± 0.003 nm obtained in the atomistic simulation. However, the RMSD observed in our simulation is nearly 30% lower than the RMSD of the lysozyme reported in earlier DPD simulations.⁵⁸ Interestingly, as indicated in Fig. S5 in the ESI,[†] the root mean square fluctuations (RMSF data) in the DPD simulation largely follow the trends observed in RMSF data obtained from the AA simulation of the lysozyme, lying well within ~ 0.5 nm for most of the regions in the protein.

Fig. 9A shows the residue–residue contacts within a cutoff distance of 1 nm in the lysozyme. Our DPD force field efficiently stabilizes the residue contacts in all six helical and extended sheet regions of the protein when compared with the contact maps derived from atomistic simulations (Fig. 9B). Furthermore, the contacts between residues forming the disulfide linkages are captured both in AA and in DPD results. For instance, the contacts around the pairs of residues, namely 6 and 127, 30 and 115, 64 and 80, and 76 and 94, are captured in DPD simulations. The R_g value of 1.549 ± 0.001 nm from the DPD simulation is in good agreement with the R_g of 1.393 ± 0.001 nm obtained in the AA simulation (Fig. 9C).

ClyA is an α pore-forming toxin secreted by *Escherichia coli*, which assembles on the target plasma membrane to form a dodecameric transmembrane channel.^{94–96} The structure of the ClyA monomer (PDB: 1QOY, 303 residues in Fig. 1) is predominantly α -helical containing the segments α A1 (residues 1–35), α A2 (residues 36–46), α B (residues 56–101), α C (residues 106–176), α F (residues 208–259) and α G (residues 269–291). The segment with residues 177–203 is referred to as the β -tongue. The residues in the β -tongue are flexible and highly hydrophobic in nature. The overall length of the protein is ~ 10 nm forming a 34 kDa protein. An energy minimized atomistic configuration of the ClyA monomer was obtained from the work of Sathyanarayana *et al.*⁹⁷ where homology modelling was used to determine the structure of the N and C terminal segments. The equivalent CG-mapped structure was simulated in water using the DPD force field. The residue–residue contact maps are depicted in Fig. 10A and B for the DPD and AA models of ClyA. The DPD model is able to faithfully reproduce most of the contacts observed in the AA data. We also observed additional contacts between residues 50–100 (α B) and 200–250 (α F) in the DPD simulations (Fig. 10A) which form part of the helical bundle in the ClyA monomer.

For most of the ClyA residues, the RMSD (Fig. S6 in the ESI[†]) and the RMSF (Fig. 10D) are within 0.5 nm, which is lower than the length scale of the CG beads. In the AA model, an increased RMSF is observed in the loop regions that connect the different helices. In addition a higher RMSF is observed in the β -tongue residues where large conformational changes occur upon binding with a phospholipid membrane. Interestingly our DPD model is able to qualitatively capture the trends in RMSF observed in the AA simulation for ClyA; however the base line fluctuations are higher in the DPD model. Fig. 10C illustrates

the secondary structures preserved in the DPD simulation of ClyA with an overall compactness similar to that observed in the crystal structure (Fig. 1). Overall our results indicate that the DPD parametrization is able to capture and reliably retain the secondary structure of ClyA when compared with the AA model.

6 Conclusions

We propose a CG force field for simulating peptides and proteins within the DPD framework. The force field is developed using the Flory–Huggins parameters for 20 standard amino acids. The bonded interactions in the DPD force field are parametrized using representative small peptides consisting of an α -helix, β -sheet and the fusion peptide of SARS-CoV-2 having an extended loop structure. The performance of the proposed DPD force field is extensively tested and compared with AA simulations using other short peptides, a large 34 kDa protein with 303 residues found in the pore-forming protein ClyA and the lysozyme protein having 129 residues. When compared with fully atomistic simulations, the proposed DPD force field is found to reliably preserve the protein structure and compactness as measured by the RMSD and R_g respectively. Excellent agreement of the native contacts with AA simulations further validates the accuracy of the proposed DPD force field. Additionally, the conformations sampled on two-dimensional Ramachandran-like θ – ϕ maps are in agreement with similar maps used in a previous CG-MD study⁹² for the respective secondary structures.

With the ability to retain the native protein structure, the model captures protein folding from unfolded states. The dihedral potential is a key factor in stabilizing the secondary conformations of the different protein structures. The dual-basin potential for the protein backbone angles, which is an improvement over previous DPD models, effectively covers the entire spectrum of secondary structures consisting of helices, loops and extended conformations.

When compared with the existing DPD models for proteins,^{52–56,58} our proposed DPD force field reliably stabilizes a variety of protein secondary structures, reproduces native contacts, and captures protein folding to the native state. Our model alleviates the limitations present in previous Morse potential based DPD models^{52–54} that have been used to predict the secondary structures of proteins, and obviates the need to reparametrize potentials for specific protein structures. Comparing our model with polarizable DPD force fields,⁵⁵ improved RMSD values are obtained for Trpcage (helix-coil motif) and Trpzip (β -sheet motif) using the proposed DPD force field. Unlike an elastic network model that preserves the structural integrity of proteins by the use of harmonic potentials between pairs of backbone beads,⁵⁶ our protein model retains the backbone flexibility by incorporating backbone angles and dihedral potentials parametrized with AA reference models. We therefore expect our model to be computationally efficient, especially for simulating large proteins and membrane–protein domains. The trimeric spike protein of SARS-CoV-2 can also be simulated using our proposed DPD force field.

Combined with the backbone angle and dihedral potentials, our proposed DPD force field is generic and not limited to any specific secondary structure. Additionally the strengths of the repulsion parameters used in our study are compatible with the existing repulsion parameters used for phospholipid membranes, allowing one to study protein-membrane systems without significant reparametrization. We finally mention that the proposed force field can potentially be modified to include dual-basin dihedral potentials similar to those used in structure based Gō models^{24,98} to enable the study of secondary structure transitions in proteins.

Conflicts of interest

The authors declare no conflict of interests.

Acknowledgements

We acknowledge the Department of Science and Technology supported Thematic Unit of Excellence on Computational Materials Science (TUE-CMS) and the National Supercomputing Mission for the computational resources used in this work. One of us, Rakesh Vaiwala, acknowledges support from Unilever Research and Development (Bengaluru, India). We also thank Rochish Thakkar and Soumava Palit from Indian Institute of Technology Bombay (IITB) for initial discussions on protein simulations using the DPD methodology. We also thank Avijeet Kulshrestha from our laboratory for providing the all-atom simulation data for ClyA.

References

- 1 J.-W. Chu, G. Ayton, S. Izvekov and G. Voth, *Mol. Phys.*, 2007, **105**, 167–175.
- 2 W. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das and H. C. Andersen, *J. Chem. Phys.*, 2008, **128**, 244114.
- 3 W. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen and G. A. Voth, *J. Chem. Phys.*, 2008, **128**, 244115.
- 4 A. Arkhipov, Y. Yin and K. Schulten, *Biophys. J.*, 2008, **95**, 2806–2821.
- 5 M. Cascella and M. Dal Peraro, *CHIMIA Int. J. Chem.*, 2009, **63**, 14–18.
- 6 C. Czaplewski, A. Liwo, M. Makowski, S. Oldziej and H. A. Scheraga, *Multiscale Approaches to Protein Modeling*, Springer, 2011, pp. 35–83.
- 7 A. N. Adhikari, K. F. Freed and T. R. Sosnick, *Phys. Rev. Lett.*, 2013, **111**, 028103.
- 8 P. G. Lafond and S. Izvekov, *J. Chem. Theory Comput.*, 2016, **12**, 5737–5750.
- 9 S. L. Bore, G. Milano and M. Cascella, *J. Chem. Theory Comput.*, 2018, **14**, 1120–1130.
- 10 O. Conway, Y. An, K. K. Bejagam and S. A. Deshmukh, *Mol. Syst. Des. Eng.*, 2020, **5**, 675–685.
- 11 R. Desikan, A. Behera, P. K. Maiti and K. G. Ayappa, *Methods Enzymol.*, 2021, **649**, 461–502.
- 12 R. Vaiwala, P. Sharma, M. Puranik and K. G. Ayappa, *J. Chem. Theory Comput.*, 2020, **16**, 5369–5384.
- 13 X. Jiang, Y. Sun, K. Yang, B. Yuan, T. Velkov, L. Wang and J. Li, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 3885–3891.
- 14 A. Yu, A. J. Pak, P. He, V. Monje-Galvan, L. Casalino, Z. Gaieb, A. C. Dommer, R. E. Amaro and G. A. Voth, *Biophys. J.*, 2021, **120**, 1097–1104.
- 15 S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid and A. Kolinski, *Chem. Rev.*, 2016, **116**, 7898–7936.
- 16 V. Tozzini, *Curr. Opin. Struct. Biol.*, 2005, **15**, 144–150.
- 17 H. Taketomi, Y. Ueda and N. Gō, *Int. J. Pept. Protein Res.*, 1975, **7**, 445–459.
- 18 N. Gō and H. Taketomi, *Int. J. Pept. Protein Res.*, 1979, **13**, 447–461.
- 19 J. Trylska, V. Tozzini and J. A. McCammon, *Biophys. J.*, 2005, **89**, 1455–1463.
- 20 P. Maragakis and M. Karplus, *J. Mol. Biol.*, 2005, **352**, 807–822.
- 21 J.-W. Chu and G. A. Voth, *Biophys. J.*, 2007, **93**, 3860–3871.
- 22 E. Lyman, J. Pfandtner and G. A. Voth, *Biophys. J.*, 2008, **95**, 4183–4192.
- 23 K. Hamacher and J. A. McCammon, *J. Chem. Theory Comput.*, 2006, **2**, 873–878.
- 24 V. H. Giri Rao, R. Desikan, K. G. Ayappa and S. Gosavi, *J. Phys. Chem. B*, 2016, **120**, 12064–12078.
- 25 T. Bereau and M. Deserno, *J. Chem. Phys.*, 2009, **130**, 06B621.
- 26 A. Koliński, *Acta Biochim. Pol.*, 2004, **51**, 349–371.
- 27 E.-H. Yap, N. L. Fawzi and T. Head-Gordon, *Proteins: Struct., Funct., Bioinf.*, 2008, **70**, 626–638.
- 28 D. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 2544–2549.
- 29 A. Mukherjee and B. Bagchi, *J. Chem. Phys.*, 2003, **118**, 4733–4747.
- 30 L. Darré, M. R. Machado, A. F. Brandner, H. C. González, S. Ferreira and S. Pantano, *J. Chem. Theory Comput.*, 2015, **11**, 723–739.
- 31 L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S.-J. Marrink, *J. Chem. Theory Comput.*, 2008, **4**, 819–834.
- 32 F. Sterpone, S. Melchionna, P. Tuffery, S. Pasquali, N. Mousseau, T. Cragolini, Y. Chebaro, J.-F. St-Pierre, M. Kalimeri and A. Barducci, *Chem. Soc. Rev.*, 2014, **43**, 4871–4893.
- 33 M. Pasi, R. Lavery and N. Ceres, *J. Chem. Theory Comput.*, 2013, **9**, 785–793.
- 34 P. Kar, S. M. Gopal, Y.-M. Cheng, A. Predeus and M. Feig, *J. Chem. Theory Comput.*, 2013, **9**, 3769–3788.
- 35 R. Das and D. Baker, *Annu. Rev. Biochem.*, 2008, **77**, 363–382.
- 36 A. Liwo, M. Baranowski, C. Czaplewski, E. Gołaś, Y. He, D. Jagieła, P. Krupa, M. Maciejczyk, M. Makowski and M. A. Mozolewska, *J. Mol. Model.*, 2014, **20**, 1–15.
- 37 N. Basdevant, D. Borgis and T. Ha-Duong, *J. Chem. Theory Comput.*, 2013, **9**, 803–813.
- 38 M. I. Mahmood, A. B. Poma and K.-I. Okazaki, *Front. Mol. Biosci.*, 2021, **8**, 1–10.

- 39 M. Baaden and S. J. Marrink, *Curr. Opin. Struct. Biol.*, 2013, **23**, 878–886.
- 40 L.-W. Yang and C.-P. Chng, *Bioinf. Biol. Insights*, 2008, **2**, BBI-S460.
- 41 P. Hoogerbrugge and J. Koelman, *Europhys. Lett.*, 1992, **19**, 155.
- 42 R. D. Groot and P. B. Warren, *J. Chem. Phys.*, 1997, **107**, 4423–4435.
- 43 M. Kranenburg, M. Venturoli and B. Smit, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2003, **67**, 060901.
- 44 M. Laradji and P. Sunil Kumar, *J. Chem. Phys.*, 2005, **123**, 224902.
- 45 F. M. Thakkar, P. K. Maiti, V. Kumaran and K. Ayappa, *Soft Matter*, 2011, **7**, 3963–3966.
- 46 F. M. Thakkar and K. Ayappa, *J. Chem. Phys.*, 2011, **135**, 104901.
- 47 R. Vaiwala, S. Jadhav and R. Thaokar, *J. Chem. Theory Comput.*, 2019, **15**, 5737–5749.
- 48 Y. H. Feng, X. P. Zhang, Z. Q. Zhao and X. D. Guo, *Mol. Pharming*, 2020, **17**, 1778–1799.
- 49 J. Wang, Y. Han, Z. Xu, X. Yang, S. Ramakrishna and Y. Liu, *Macromol. Mater. Eng.*, 2021, **306**, 2000724.
- 50 N. Spenley, *Europhys. Lett.*, 2000, **49**, 534.
- 51 R. Vaiwala and R. Thaokar, *Europhys. Lett.*, 2018, **120**, 48001.
- 52 A. Vishnyakov, D. S. Talaga and A. V. Neimark, *J. Phys. Chem. Lett.*, 2012, **3**, 3081–3087.
- 53 L. Chen, X. Li, L. Gao and W. Fang, *J. Phys. Chem. B*, 2015, **119**, 850–860.
- 54 X. Li, M. Wan, L. Gao and W. Fang, *Sci. Rep.*, 2016, **6**, 1–17.
- 55 E. K. Peter, K. Lykov and I. V. Pivkin, *Phys. Chem. Chem. Phys.*, 2015, **17**, 24452–24461.
- 56 A. Truszkowski, K. van den Broek, H. Kuhn, A. Zielesny and M. Eppel, *J. Chem. Inf. Model.*, 2015, **55**, 983–997.
- 57 I. Bahar, A. R. Atilgan and B. Erman, *Folding Des.*, 1997, **2**, 173–181.
- 58 C. K. Choudhury and O. Kuksenok, *J. Phys. Chem. B*, 2020, **124**, 11379–11386.
- 59 R. D. Groot and K. Rabone, *Biophys. J.*, 2001, **81**, 725–736.
- 60 F. M. Thakkar and K. Ayappa, *J. Phys. Chem. B*, 2010, **114**, 2738–2748.
- 61 G. Illya, R. Lipowsky and J. Shillcock, *J. Chem. Phys.*, 2006, **125**, 114710.
- 62 M. Kranenburg and B. Smit, *J. Phys. Chem. B*, 2005, **109**, 6553–6563.
- 63 M. Laradji and P. S. Kumar, *Computer Simulation Studies in Condensed-Matter Physics XIX*, Springer, 2009, pp. 119–133.
- 64 R. D. Hills Jr, L. Lu and G. A. Voth, *PLoS Comput. Biol.*, 2010, **6**, e1000827.
- 65 F. Ding, S. V. Buldyrev and N. V. Dokholyan, *Biophys. J.*, 2005, **88**, 147–155.
- 66 P. Robustelli, S. Piana and D. E. Shaw, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E4758–E4766.
- 67 T. Sanyal, J. Mittal and M. S. Shell, *J. Chem. Phys.*, 2019, **151**, 044111.
- 68 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
- 69 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19–25.
- 70 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.
- 71 G. J. Martyna, M. L. Klein and M. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635–2643.
- 72 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
- 73 B. Hess, H. Bekker, H. J. Berendsen and J. G. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
- 74 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 75 R. D. Groot, *J. Chem. Phys.*, 2003, **118**, 11265–11277.
- 76 P. Espanol and P. Warren, *Europhys. Lett.*, 1995, **30**, 191.
- 77 M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, Oxford university press, 2017.
- 78 M. González-Melchor, E. Mayoral, M. E. Velázquez and J. Alejandre, *J. Chem. Phys.*, 2006, **125**, 224107.
- 79 K. A. Terrón-Meja, R. López-Rendón and A. G. Goicochea, *J. Phys.: Condens. Matter*, 2016, **28**, 425101.
- 80 R. Vaiwala, S. Jadhav and R. Thaokar, *J. Chem. Phys.*, 2017, **146**, 124904.
- 81 C. Ibergay, P. Malfreyt and D. J. Tildesley, *J. Chem. Theory Comput.*, 2009, **5**, 3245–3259.
- 82 C. Ramakrishnan and G. Ramachandran, *Biophys. J.*, 1965, **5**, 909–933.
- 83 S. Brown, N. J. Fawzi and T. Head-Gordon, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 10712–10717.
- 84 K. M. Saravanan and S. Selvaraj, *J. Biol. Phys.*, 2017, **43**, 265–278.
- 85 H. Sun, *J. Phys. Chem. B*, 1998, **102**, 7338–7364.
- 86 S. Burgess, A. Vishnyakov, C. Tsovko and A. V. Neimark, *J. Phys. Chem. Lett.*, 2018, **9**, 4872–4877.
- 87 T. Sugii, S. Takagi and Y. Matsumoto, *Int. J. Numer. Methods Fluids*, 2007, **54**, 831–840.
- 88 V. Ortiz, S. O. Nielsen, D. E. Discher, M. L. Klein, R. Lipowsky and J. Shillcock, *J. Phys. Chem. B*, 2005, **109**, 17708–17714.
- 89 M. Choudhary and S. Kamil, *ACS Omega*, 2020, **5**, 22891–22900.
- 90 A. B. Poma, M. Cieplak and P. E. Theodorakis, *J. Chem. Theory Comput.*, 2017, **13**, 1366–1374.
- 91 S. Takada, Z. Luthey-Schulten and P. G. Wolynes, *J. Chem. Phys.*, 1999, **110**, 11616–11629.
- 92 V. Tozzini, W. Rocchia and J. A. McCammon, *J. Chem. Theory Comput.*, 2006, **2**, 667–673.
- 93 J. Kubelka, J. Hofrichter and W. A. Eaton, *Curr. Opin. Struct. Biol.*, 2004, **14**, 76–88.
- 94 M. Mueller, U. Grauschopf, T. Maier, R. Glockshuber and N. Ban, *Nature*, 2009, **459**, 726–730.
- 95 P. Sathyanarayana, S. S. Visweswariah and K. G. Ayappa, *Acc. Chem. Res.*, 2020, E7323–E7330.
- 96 P. Sathyanarayana, S. Maurya, A. Behera, M. Ravichandran, S. S. Visweswariah, K. G. Ayappa and R. Roy, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E7323–E7330.
- 97 P. Sathyanarayana, R. Desikan, K. G. Ayappa and S. S. Visweswariah, *Biochemistry*, 2016, **55**, 5952–5961.
- 98 Q. Lu and J. Wang, *J. Am. Chem. Soc.*, 2008, **130**, 4772–4783.