Supplementary Information for: Adsorption of Amino Acids on Graphene: Assessment of Current Force Fields

Siva Dasetty,[a] John K. Barrows,[b] and Sapna Sarupria*[a]

1 Differences in the parameterization of partial charges across force fields

Amber03w/TIP4P/2005 is a variant of Amber03 that was optimized to work with TIP4P/2005 water model. It was optimized to capture the changes in helix-coil transition obtained in nuclear magnetic resonance (NMR) experiments for peptides in a broad range of temperatures.[1] The optimization was performed on the Fourier coefficients (specifically ψ) of the backbone dihedral angles. Amber03 differs from Amber family-based (Amberff99SB-ILDN and its predecessors) force fields in only estimating the point charges of each amino acid and the backbone dihedral parameters.[2] The partial charges of the amino acids in Amber family-based force fields were obtained by quantum mechanics (QM) simulations using HF/6-31G* level of theory in gas phase. The usage of resulting partial charges in condensed phase simulations were justified by the statistical similarity of gas phase and condensed phase bond-dipole moments.[3]

Amber03 used density functional theory method (B3LYP/cc-pVTZ//HF/6-31G**) with continuum solvent model to estimate partial charges of amino acids. The dielectric constant was chosen to mimic the interior of protein and not water.[2] The resulting partial charges were lower in magnitude than those of Amber-family based charges. The two-sets of partial charges were strongly correlated with each other but the amino acids represented by Amber03 are less polar overall than Amber-family based charges.[2] The differences in the polarity was quantified by the slopes of the trend lines that best fit the charges obtained in the two methods. CHARMM- and OPLS- based force fields used charge groups for each chemical fragment and separate charges for each atom type, respectively. These two force fields differ from the Amber (Amber-family and Amber03) force fields in estimating the non-bonded parameters. CHARMM optimizes the partial charges using a combination of experimental data and interaction energies between amino acid analogs and mTIP3P (hydrogens have non-zero LJ parameters) by quantum simulations. The complexes were constructed by optimizing the positions of water molecules near each site of a model compound using HF/6-31G(d) level of theory.[4] This is one of the reason for the tight coupling between water models and CHARMM force field parameters. On the other hand, partial charges in OPLS were obtained by fitting the thermodynamic properties of organic liquids determined by molecular simulations to the available experimental data.[5]

2 Differences in the revision protocols of force fields

The backbone parameters of Amber03w/TIP4P/2005 were determined by refining the Fourier coefficient of ψ in Amber03 to fit the temperature variation of helix-coil transition of polyalanine (Ala_5) and α-helix forming peptide (ACE-(Ala-Ala-Gln-Ala-Ala)$_3$-NH_2). In Amberff99SB-ILDN, the side chain dihedral parameters of Ile, Leu, Asp, and Asn were optimized by using the potential energy surface (PES) of the side chain dihedral space of capped amino acids (capping groups are ACE and NMA) determined by gas phase QM simulations. These amino acids were selected because the sampling of the side chain rotamer distribution observed in crystal structures and molecular simulations differ more for these amino acids than other amino acids. In the optimization of the side chain dihedral parameters, the backbone dihedral parameters were constrained to extended conformation in Amberff99SB-ILDN. Subsequently, the final parameters were validated against the J- and residual dipolar- coupling constants observed in NMR experiments of globular proteins (ubiqitin, hen egg white lysozyme, bovine pancreatic trypsin, and B4 domain of protein G).[6]

In CHARMM36, the side chain dihedral terms and crossterms or grid-based energy correction maps (CMAP) terms implemented in CHARMM22/CMAP[7] were updated using a combination of gas phase QM simulations and condensed phase experimental data. CMAP is a correction term in the potential energy equation incorporated to minimize the differences in the backbone dihedral sampling obtained via molecular mechanics and QM simulations.[8] The backbone CMAP parameters of residues excluding Gly and Pro were updated using NMR data of the same peptides used in Amber03w/TIP4P/2005 parameterization. For Gly and Pro, PESs obtained by QM were used for optimizing the backbone parameters. The side chain dihedral space was sampled by constraining the backbone dihedral angles to α-helix, β-strand, and α'-helix regions (nomenclature defined in main text). The final set of parameters were validated against experimental data spanning short peptides to globular proteins.[9] Similarly, the dihedral parameters of OPLS-AA[10] were updated to generate OPLS-AA/M using updated gas phase QM simulations data of capped amino acids. The side chain dihedral space was scanned by constraining the backbone dihedral angles to α-helix and β-sheet regions.

[a] Department of Chemical & Biomolecular Engineering, Clemson University, Clemson, SC 29634, USA; E-mail: ssarupr@g.clemson.edu
[b] Department of Biochemistry & Molecular Biology, Medical University of South Carolina, Charleston, SC 29425, USA

3 Effect of system size

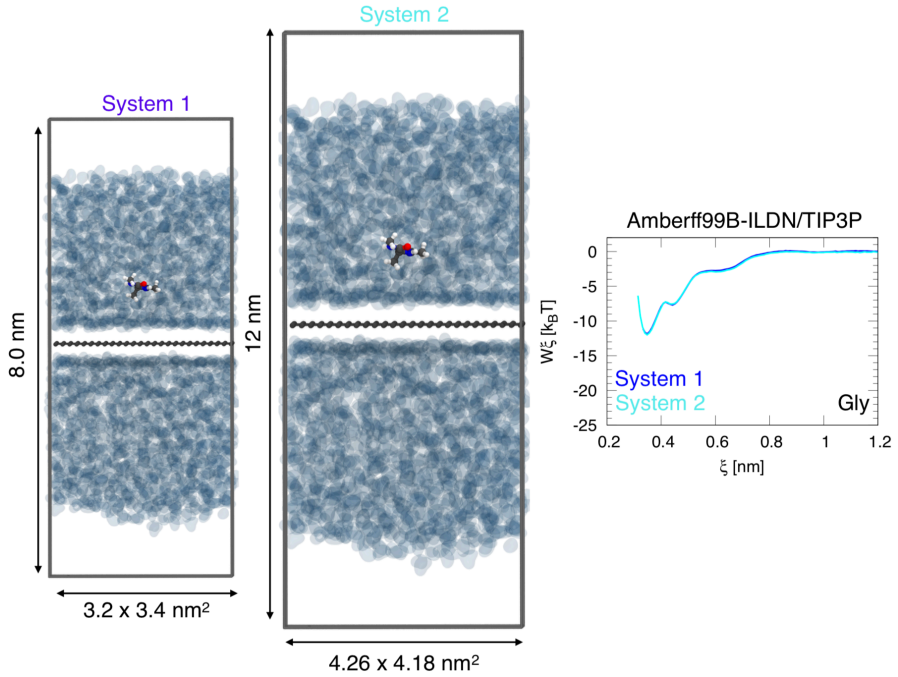


Fig. S1 A larger system (system 2) was constructed by increasing the dimensions of the box by ∼1 nm along the graphene surface, and 4 nm in the direction normal to graphene surface. Umbrella sampling calculations were performed for Gly–graphene system with the larger system containing 4462 water molecules. The resulting PMF obtained from a single run with the larger system was compared with the average PMF from the smaller system (system 1). Similar PMFs indicate no system size effects in our results.

4 Comparison of PMF profile obtained with Amber03w/TIP3P and other force fields

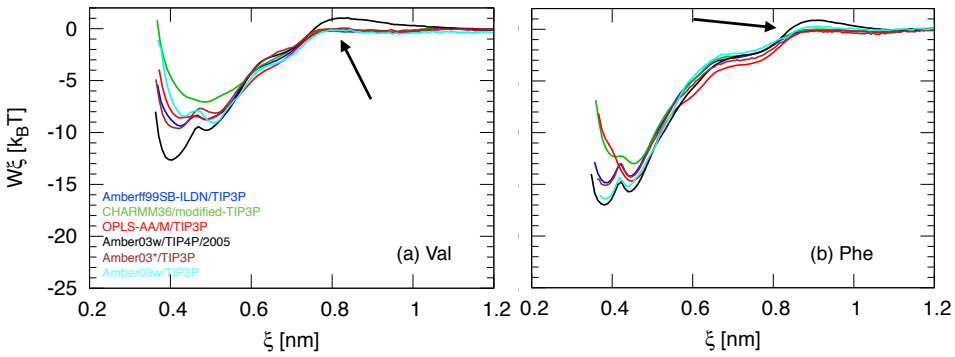


Fig. S2 PMF profile of (a) Val and (b) Phe obtained with Amber03*/TIP3P (compatible pair) and Amber03w/TIP3P in a single run compared with the average PMFs obtained with other force fields. The arrows point towards the difference in the PMF profile between Amber03w/TIP4P/2005 and other force fields. These PMFs are shown to illustrate the impact of water model on the PMF profile for 0.8 nm ξ < 1.0 nm.

5 Unbiased MD simulations

The unbiased MD simulations of the amino acids–graphene system were performed for analyzing the non-bonded interaction energies, and the conformations of the amino acid in the graphene adsorbed state. The simulations were performed by placing the amino acids at ξ = 0.6 nm and using the same simulation procedure described in Sections IIA and IIB of main text. Similarly, we performed the unbiased MD simulations of the amino acids in water (without graphene) in a NPT ensemble at T = 300 K and P = 1 bar to analyze the conformations of amino acids in bulk. The initial configurations of the systems were prepared by solvating the capped amino acids with ~1935 water molecules in a box of volume $3.2\times3.4\times6$ nm^3. Temperature and pressure in the production simulations were controlled by using Nosé-Hoover thermostat and Parrinello-Rahman barostat, respectively. The pressure in the system was maintained isotropically with a time constant of 1 ps and compressibility of 4.5×10^{-5} bar^{-1}. The remaining parameters were same as those used during PMF calculations (Section IIB). The unbiased MD simulations were performed for 10 ns and repeated for 5 times using different initial velocities to generate 10000 configurations for analysis. Both umbrella sampling and unbiased MD simulations were performed using GROMACS-5.0.2.[11]

6 Change in free energy of adsorption with definition of graphene adsorbed and bulk states

We estimated the ΔA_{ads} for all the amino acid–graphene systems using Equation 1, where k_B is the Boltzmann constant, and T is the temperature. The term in the parentheses is the ratio of the probabilities of the system to be in graphene adsorbed state (S_{ads}) relative to bulk state (S_{bulk}).

$$\Delta A_{ads} = -k_B T \ln\left\{\frac{\int_{S_{ads}} \exp\left(\frac{-W(\xi)}{k_B T}\right) dS}{\int_{S_{bulk}} \exp\left(\frac{-W(\xi)}{k_B T}\right) dS}\right\} \tag{1}$$

We considered three approaches to define S_{ads} and S_{bulk}. (i) All the configurations with ξ <0.5 nm as S_{ads}. (ii) We used the maximum value of the mean force along ξ that is closest to the graphene surface as the upper bound for S_{ads}. The mean force along ξ was obtained by determining the negative of the slope of W(ξ) with respect to ξ. (iii) The upper bound was defined as the value of ξ for which the cumulative distribution function (CDF) of ξ is 0.95 in the unbiased simulations (5). In all the three definitions, we used the first data point of ξ in the PMF/mean force/CDF as the lower bound of S_{ads}. For S_{bulk}, we used 2.0 nm$-\delta_{ads} \leq \xi \leq$ 2.0 nm as the bulk state, where δ_{ads} is the difference between upper and lower bounds of S_{ads}. We found that the values of the amino acids for all force fields were insensitive to our definition of S_{ads}. Therefore, we used definition (iii) to compare the changes in the ΔA_{ads} with amino acids across force fields.

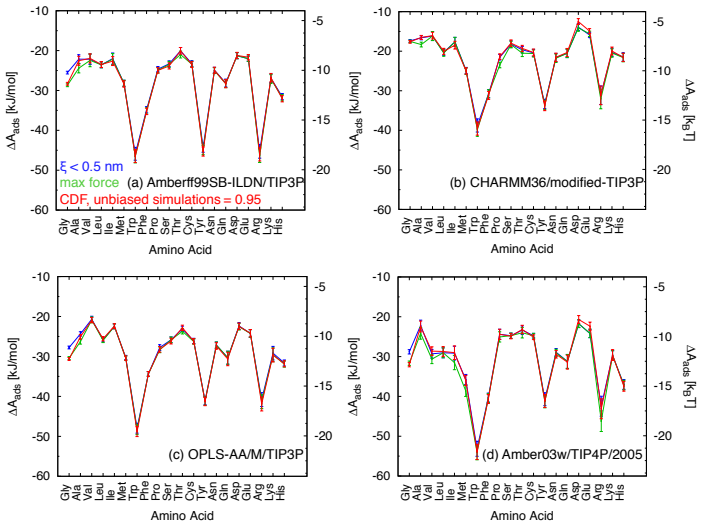


Fig. S3 Difference in the free energies of adsorption with different methods to determine the limits of ξ for the graphene adsorbed and bulk states. Blue, green, and red indicate definitions (i), (ii), and (iii), respectively.

Table S1 Free energy of adsorption (ΔA_{ads})

Amino acid	Amberff99SB-ILDN		CHARMM36		OPLS-AA/M		Amber03w	
Water model	TIP3P		modified TIP3P		TIP3P		TIP4P/2005	
	ΔA_{ads}	SE	ΔA_{ads}	SE	ΔA_{ads}	SE	ΔA_{ads}	SE
	[kJ/mol]		[kJ/mol]		[kJ/mol]		[kJ/mol]	
Gly	-28.315	0.357	-17.636	0.337	-30.628	0.36	-32.04	0.542
Ala	-22.547	1.128	-16.509	0.505	-25.077	0.564	-22.453	1.371
Val	-21.959	1.146	-16.064	0.849	-20.904	0.648	-28.631	1.086
Leu	-23.494	0.737	-20.04	0.761	-25.576	0.529	-28.631	1.234
Ile	-22.632	1.117	-18.472	0.992	-22.396	0.582	-29.014	1.701
Met	-28.367	0.801	-24.996	0.795	-30.414	0.539	-35.984	1.264
Trp	-46.475	1.615	-39.559	1.666	-48.815	1.243	-53.952	1.9
Phe	-35.247	0.807	-31.173	1.008	-34.381	0.592	-40.655	1.115
Pro	-24.972	0.679	-21.366	0.703	-28.215	0.719	-24.405	1.278
Ser	-23.848	0.767	-17.957	0.811	-26.157	0.721	-24.865	0.673
Thr	-20.055	0.825	-19.288	0.767	-22.818	0.652	-23.187	1.047
Cys	-23.506	0.726	-20.32	0.875	-26.206	0.656	-25.133	0.6
Tyr	-45.315	1.151	-33.828	1.206	-41.278	1.039	-41.158	1.631
Asn	-24.933	0.764	-21.542	1.051	-27.379	0.808	-29.423	0.947
Gln	-28.315	1.011	-20.318	1.022	-30.632	1.682	-31.399	1.75
Asp	-21.22	0.778	-12.502	0.739	-22.455	0.823	-20.683	0.972
Glu	-21.658	0.666	-14.901	0.613	-24.188	0.928	-22.368	1.033
Arg	-46.019	1.669	-31.178	2.356	-41.894	1.751	-43.115	2.505
Lys	-26.799	1.064	-19.959	1.003	-29.654	1.718	-29.603	1.352
His	-32.088	0.774	-21.554	1.054	-31.845	0.756	-37.461	1.233

7 Correlations between ΔA_{ads} and atomic hydrophobic scale, molecular weight, and components of potential energy

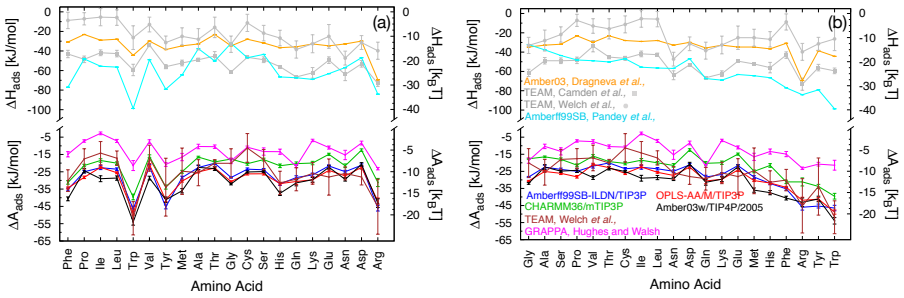


Fig. S4 Change in ΔA_{ads} with (a) atomic hydrophobic scale [12] and (b) molecular weight. Atomic hydrophobic scale is based on the charge distributions of each atom type in OPLS-AA [13] force field.

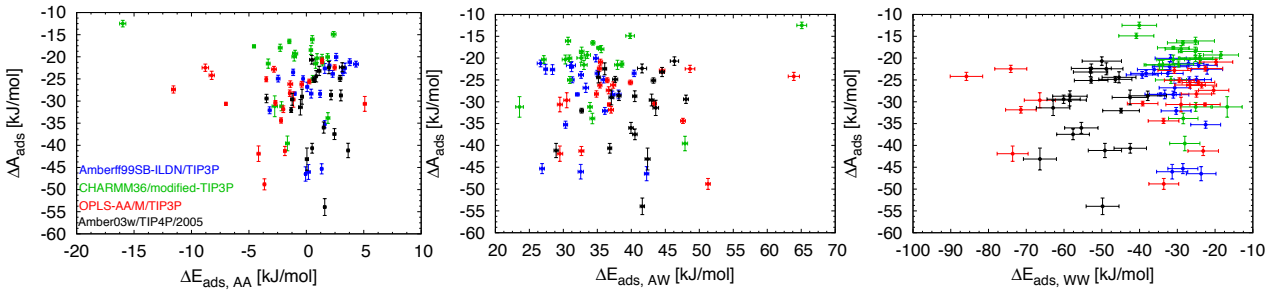


Fig. S5 Correlations between ΔA_{ads} and the interaction energy between amino acid–amino acid (AA), amino acid–water (AW), and water–water (WW).

8 ϕ-ψ distributions in bulk and graphene adsorbed states

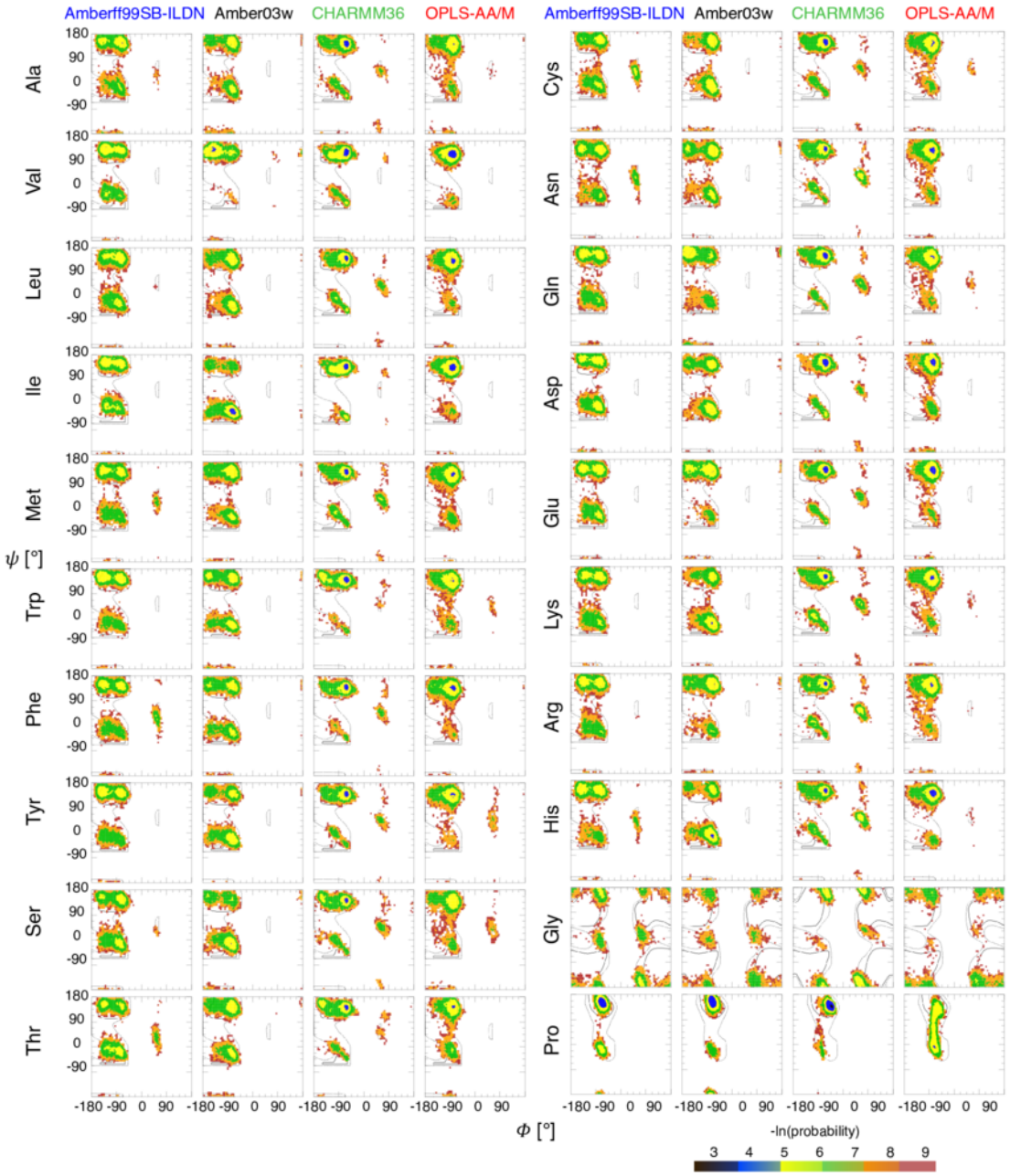


Fig. S6 Ramachandran angles of the amino acids observed with different force fields in bulk state. Water models are not included in the legend. For Amberff99SB-ILDN and OPLS, TIP3P is used. TIP4P/2005 and mTIP3P are used with Amber03w and CHARMM36, respectively. The background contours represent the sterically preferred Ramachandran angles[14]. Cool to warm colors in the heat map indicate probabilities from high to low, respectively.

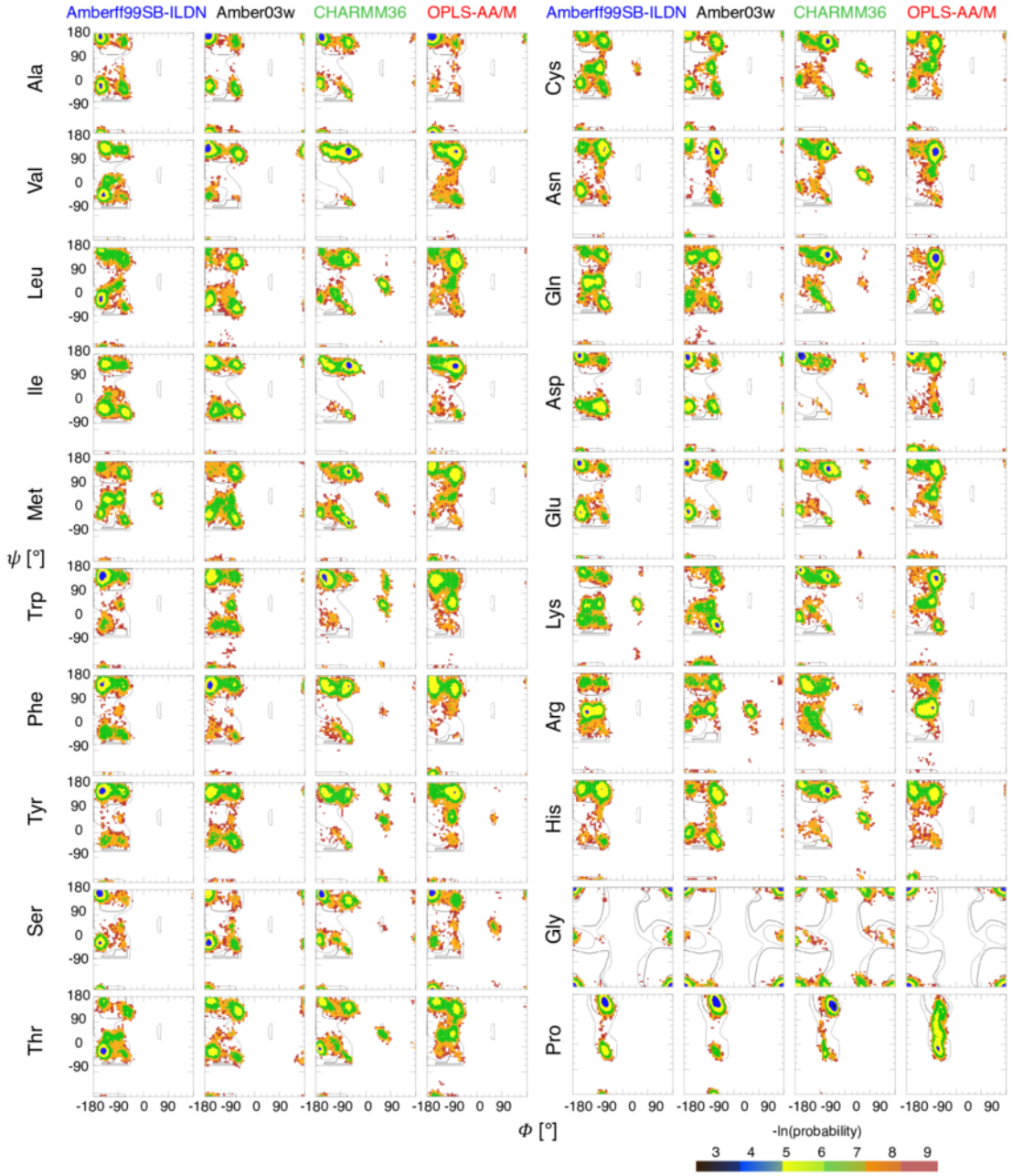


Fig. S7 Ramachandran angles of the amino acids observed with different force fields in graphene adsorbed state. Color scheme is same as that in Fig S6.

9 JS distance between the ϕ-ψ distributions

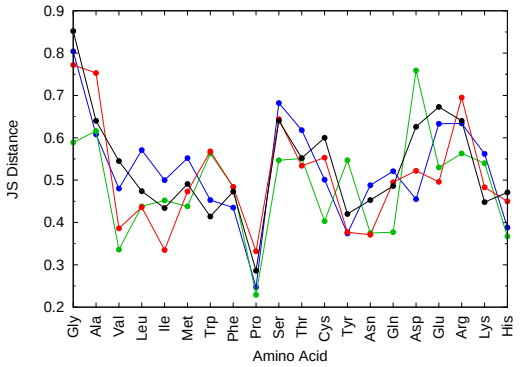


Fig. S8 Variation of JS distance between graphene adsorbed and bulk states. If JS distance is 0, then the distributions are considered identical. Color code is same as that in Fig. S5

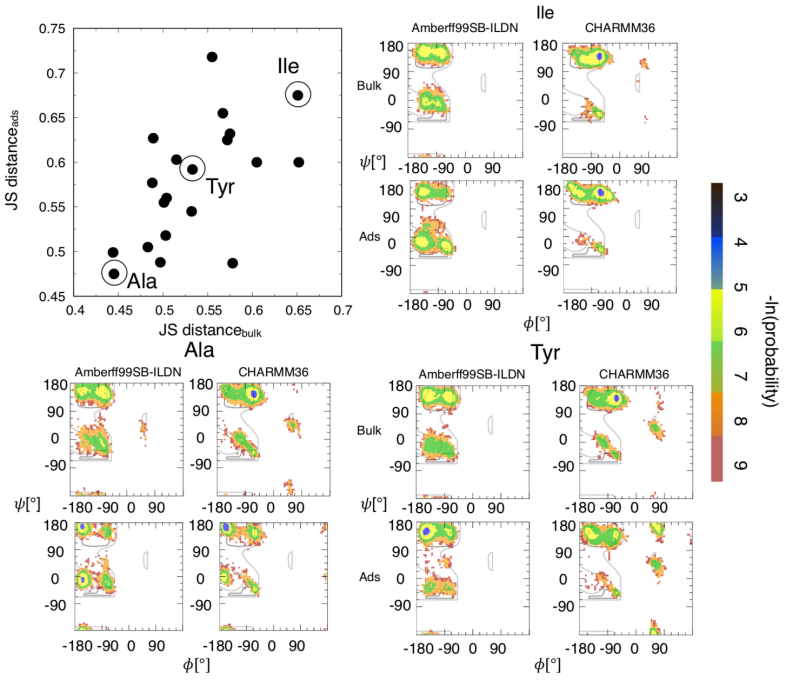


Fig. S9 Variation of JS distance between Amberff99SB-ILDN/TIP3P and CHARMM36/mTIP3P observed in bulk and adsorbed states. If JS distance is 0, then the distributions are considered identical. Color scheme of the ϕ-ψ plots is same as that in Figure S6.

10 Clusters analysis

We employed the recently developed accelerated hierarchical density-based spatial clustering of applications [15] with noise (HDBSCAN*) algorithm by McInnes and Healy [16,17] for clustering structures of amino acids adsorbed on graphene. In comparison to standard partition (k-means) and hierarchical (e.g. single linkage) cluster algorithms, HDBSCAN* provides a robust set of clusters with minimum information from the user. For instance, the clusters and the cluster centers produced in standard k-means algorithms vary depending on the random initial assignment of cluster centers, and the choice of the number of clusters. HDBSCAN* requires no *a priori* information about the cutoff distances for identifying points in a cluster. Instead, there are two key parameters that the user controls — minimum cluster size and minimum number of samples. These two parameters are used for determining an optimum value of the cutoff distance that will result in well-separated clusters. The first parameter signifies the minimum number of data points that can form a separate cluster. To identify the core clusters i.e., densest clusters, the second parameter can be adjusted. The second parameter can be

interpreted as the uncertainty in the population of a cluster. These two parameters can be subjective in nature and indirectly control the number of clusters produced. HDBSCAN* can be useful for extracting configurations representing the most probable structures in a ensemble, which suits our objective. The clustering method followed by HDBSCAN* is based on the local density of points. The boundaries of each cluster are determined by the user provided parameters.

The efficiency of any cluster analysis algorithm depends on the data and the distance metric used for determining clusters. To identify clusters of amino acid structures with similar orientation with respect to graphene and intramolecular energy, we used a combination of the backbone dihedral angles, heavy atom normal distance from graphene, and non-bonded intramolecular energies of amino acids. Since a combination of distances, angles, and energies would result in highly correlated data, determining clusters with low variance within each cluster would be challenging. For minimizing the effects of correlations within the data, we performed principal component analysis (PCA).[18] We applied HDBSCAN* algorithm on the data projected[18] onto eigenvectors that captures 95% of the variance in the data. Prior to PCA, we normalized the distances and energies with the respective maximum absolute values in the dataset to keep the scale, and the range of values uniform between angles (in radians), distances, and energies. The dihedral angles were normalized by 2π radians and we used wrapped Ramachandran angles proposed by Hollingsworth and Karplus[19] for removing periodicity of angles. Our preliminary analysis indicated Euclidean (or l_2 norm) distance metric with a minimum cluster size of 50 and number of samples as 5 produces clusters that correspond to critical regions (α-helix, β-strand, and P_{II} region) in the dihedral space. These parameters results in <15 reproducible clusters, which makes visual analysis less cumbersome. Furthermore, we observed a magnitude of <0.05-0.2 nm for the root mean square deviation (RMSD) of the heavy atoms between the structures in each of the identified clusters. The snapshots of the superimposed structures are generated using VMD-1.9.3.[20]

Illustration of similarity in the structures within clusters:

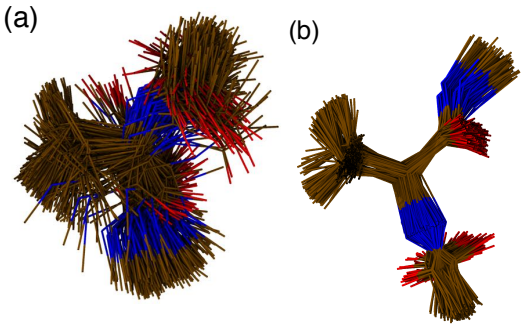
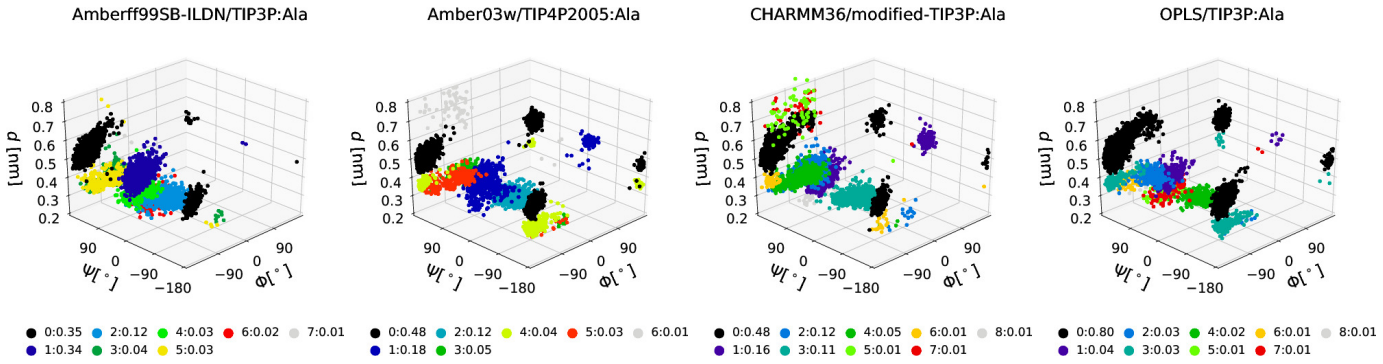
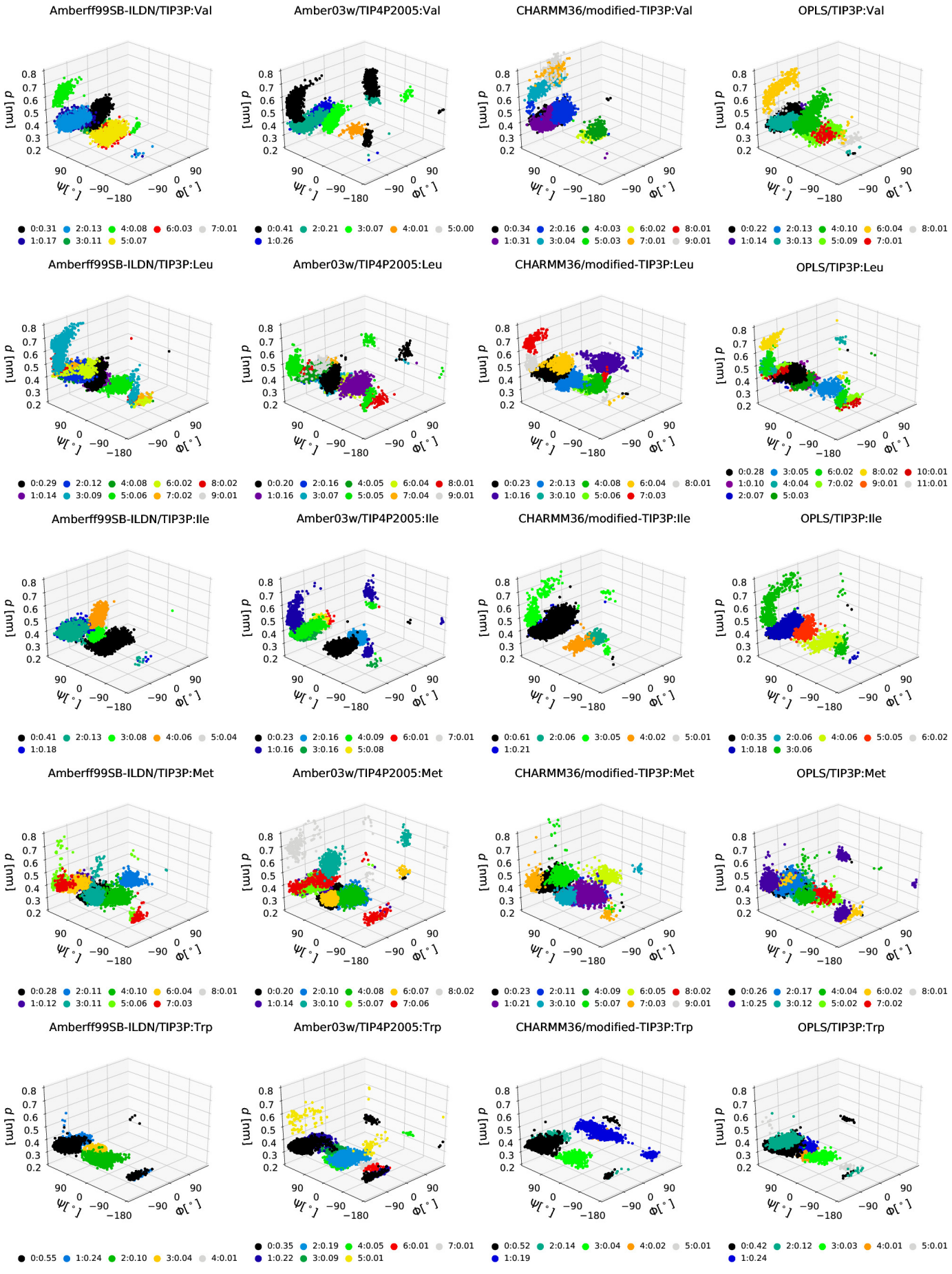


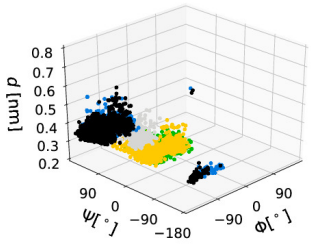
Fig. S10 Superimposed structures after performing a least squares fit (a) along the graphene plane, and (b) 3D fit of cluster 0 identified in Val with CHARMM36/mTIP3P.

Clusters in ϕ-ψ-d subspace: d is the z-distance between the center of mass of side chain and the graphene surface. Each color in each plot represents a different cluster. The cluster labels are shown in the legend below each plot. The numbers next to each color in the legend indicate the fraction of total structures in that cluster.



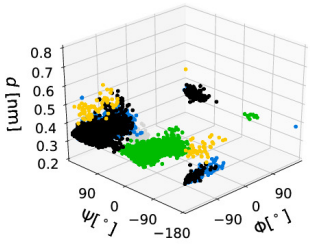

Amberff99SB-ILDN/TIP3P:Val
Amber03w/TIP4P2005:Val
CHARMM36/modified-TIP3P:Val
OPLS/TIP3P:Val
Amberff99SB-ILDN/TIP3P:Leu
Amber03w/TIP4P2005:Leu
CHARMM36/modified-TIP3P:Leu
OPLS/TIP3P:Leu
Amberff99SB-ILDN/TIP3P:Ile
Amber03w/TIP4P2005:Ile
CHARMM36/modified-TIP3P:Ile
OPLS/TIP3P:Ile
Amberff99SB-ILDN/TIP3P:Met
Amber03w/TIP4P2005:Met
CHARMM36/modified-TIP3P:Met
OPLS/TIP3P:Met
Amberff99SB-ILDN/TIP3P:Trp
Amber03w/TIP4P2005:Trp
CHARMM36/modified-TIP3P:Trp
OPLS/TIP3P:Trp
d [nm]
Ψ[°]
Φ[°]

Amberff99SB-ILDN/TIP3P:Phe



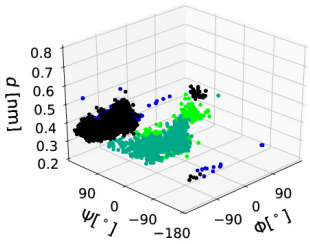
0:0.41 1:0.26 2:0.12 3:0.11 4:0.03

Amber03w/TIP4P2005:Phe



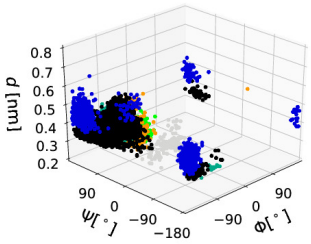
0:0.58 1:0.23 2:0.11 3:0.01 4:0.01

CHARMM36/modified-TIP3P:Phe



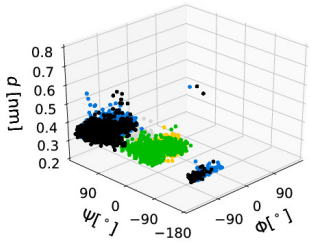
0:0.48 2:0.14 3:0.05 4:0.01 5:0.01
1:0.23

OPLS/TIP3P:Phe



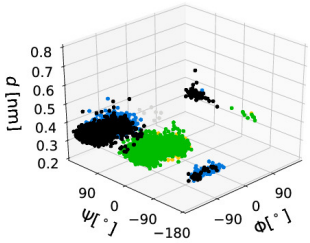
0:0.53 2:0.15 3:0.03 4:0.02 5:0.01
1:0.20

Amberff99SB-ILDN/TIP3P:Tyr



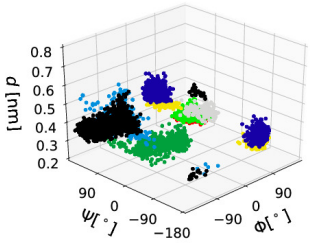
0:0.44 1:0.37 2:0.07 3:0.07 4:0.02

Amber03w/TIP4P2005:Tyr



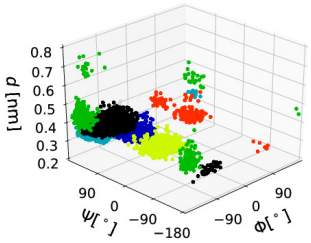
0:0.36 1:0.28 2:0.25 3:0.04 4:0.02

CHARMM36/modified-TIP3P:Tyr



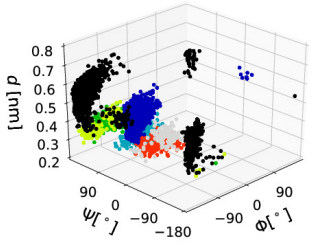
0:0.41 2:0.15 4:0.04 6:0.03 7:0.01
1:0.21 3:0.08 5:0.03

OPLS/TIP3P:Tyr



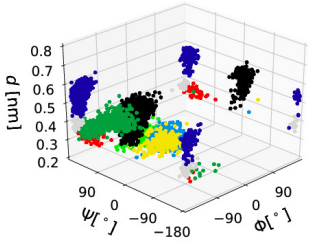
0:0.58 2:0.07 4:0.05 5:0.02 6:0.01
1:0.11 3:0.05

Amberff99SB-ILDN/TIP3P:Ser



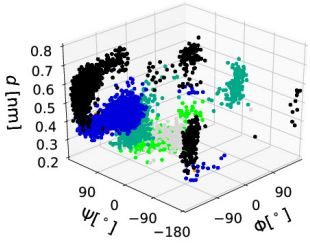
0:0.53 2:0.04 4:0.02 5:0.02 6:0.01
1:0.29 3:0.03

Amber03w/TIP4P2005:Ser



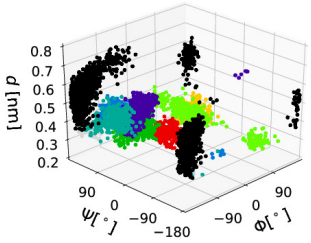
0:0.46 2:0.06 4:0.04 6:0.02 7:0.01
1:0.21 3:0.05 5:0.02

CHARMM36/modified-TIP3P:Ser



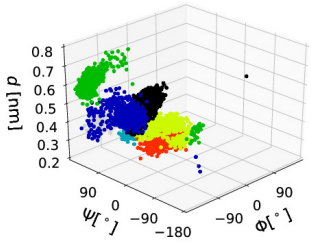
0:0.43 2:0.16 3:0.06 4:0.02 5:0.02
1:0.25

OPLS/TIP3P:Ser



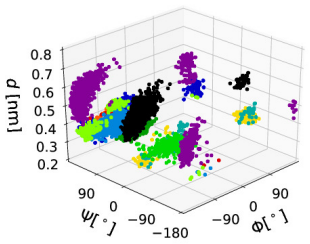
0:0.59 2:0.06 4:0.03 6:0.03 8:0.01
1:0.09 3:0.04 5:0.03 7:0.03

Amberff99SB-ILDN/TIP3P:Thr



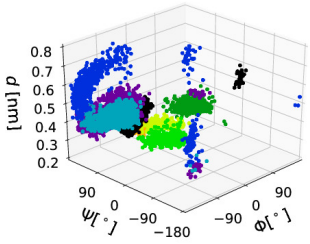
0:0.43 2:0.10 4:0.05 5:0.03 6:0.02
1:0.11 3:0.09

Amber03w/TIP4P2005:Thr



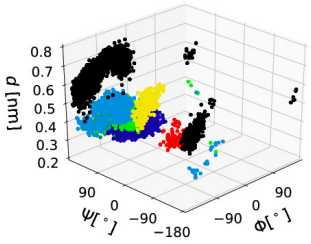
0:0.23 3:0.10 6:0.02 8:0.01 10:0.01
1:0.20 4:0.05 7:0.02 9:0.01 11:0.01
2:0.18 5:0.03

CHARMM36/modified-TIP3P:Thr



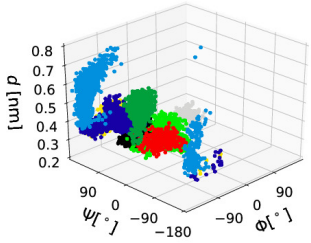
0:0.21 2:0.16 4:0.08 6:0.02 8:0.01
1:0.17 3:0.13 5:0.03 7:0.02 9:0.01

OPLS/TIP3P:Thr



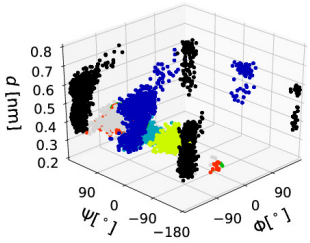
0:0.29 2:0.07 4:0.06 6:0.01 7:0.01
1:0.28 3:0.07 5:0.03

Amberff99SB-ILDN/TIP3P:Cys



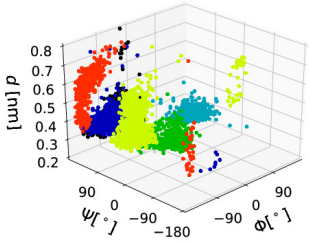
0:0.22 2:0.15 4:0.09 6:0.03 7:0.02
1:0.18 3:0.13 5:0.06

Amber03w/TIP4P2005:Cys



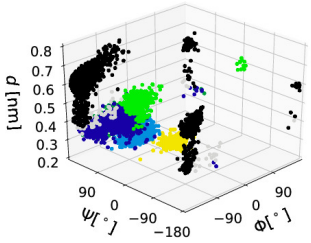
0:0.25 2:0.13 4:0.10 5:0.08 6:0.04
1:0.20 3:0.13

CHARMM36/modified-TIP3P:Cys



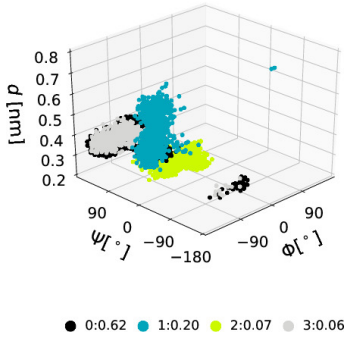
0:0.28 2:0.14 4:0.13 5:0.05 6:0.01
1:0.21 3:0.13

OPLS/TIP3P:Cys



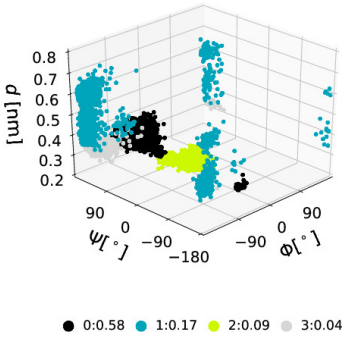
0:0.30 2:0.16 4:0.07 6:0.01 7:0.01
1:0.20 3:0.08 5:0.02

Amberff99SB-ILDN/TIP3P:Asn



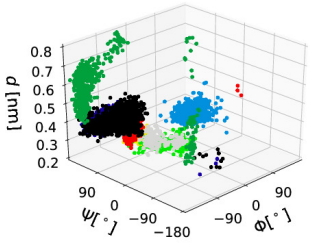
0:0.62 1:0.20 2:0.07 3:0.06

Amber03w/TIP4P2005:Asn



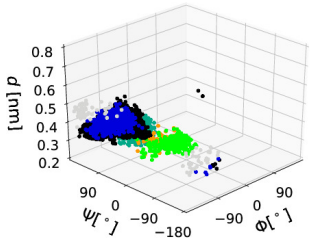
0:0.58 1:0.17 2:0.09 3:0.04

CHARMM36/modified-TIP3P:Asn



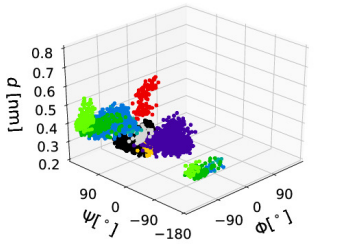
0:0.35 2:0.18 4:0.03 6:0.02 7:0.02
1:0.19 3:0.07 5:0.03

OPLS/TIP3P:Asn



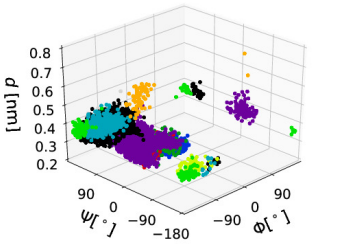
0:0.70 2:0.08 3:0.03 4:0.01 5:0.01
1:0.08

Amberff99SB-ILDN/TIP3P:Gln



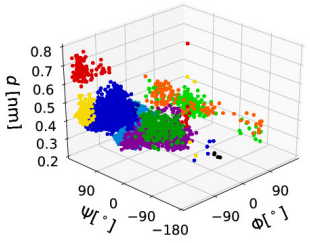
0:0.40 2:0.12 4:0.03 6:0.01 8:0.01
1:0.18 3:0.10 5:0.02 7:0.01

Amber03w/TIP4P2005:Gln



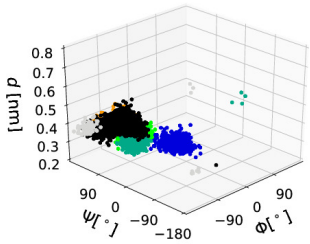
0:0.43 2:0.09 4:0.04 6:0.01 8:0.01
1:0.25 3:0.06 5:0.02 7:0.01 9:0.01

CHARMM36/modified-TIP3P:Gln



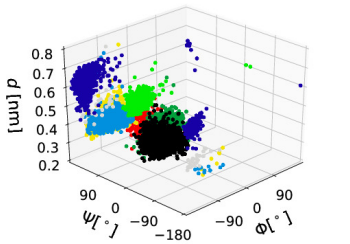
0:0.25 3:0.09 6:0.02 8:0.01 10:0.01
1:0.17 4:0.02 7:0.01 9:0.01 11:0.01
2:0.14 5:0.02

OPLS/TIP3P:Gln



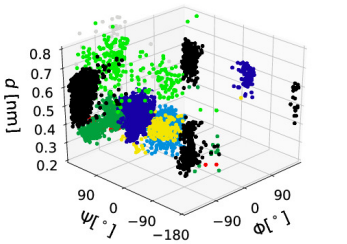
0:0.68 2:0.07 3:0.04 4:0.02 5:0.01
1:0.14

Amberff99SB-ILDN/TIP3P:Asp



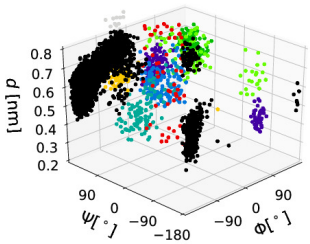
0:0.36 2:0.09 4:0.08 6:0.03 7:0.01
1:0.29 3:0.08 5:0.03

Amber03w/TIP4P2005:Asp



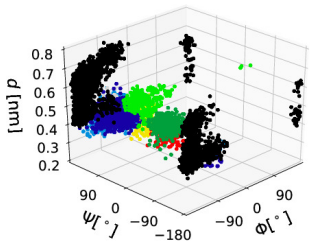
0:0.45 2:0.13 4:0.04 6:0.01 7:0.01
1:0.24 3:0.06 5:0.03

CHARMM36/modified-TIP3P:Asp



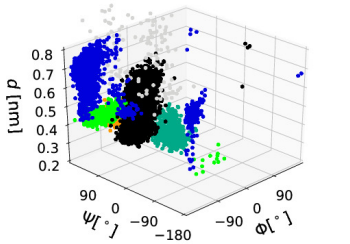
0:0.86 2:0.02 4:0.01 6:0.01 8:0.01
1:0.02 3:0.01 5:0.01 7:0.01

OPLS/TIP3P:Asp



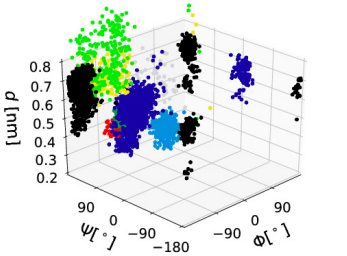
0:0.47 2:0.08 4:0.03 6:0.01 7:0.01
1:0.21 3:0.06 5:0.03

Amberff99SB-ILDN/TIP3P:Glu



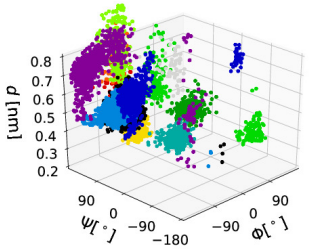
0:0.42 2:0.06 3:0.06 4:0.03 5:0.02
1:0.37

Amber03w/TIP4P2005:Glu



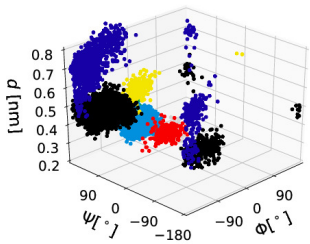
0:0.38 2:0.04 4:0.03 6:0.02 7:0.01
1:0.36 3:0.03 5:0.02

CHARMM36/modified-TIP3P:Glu



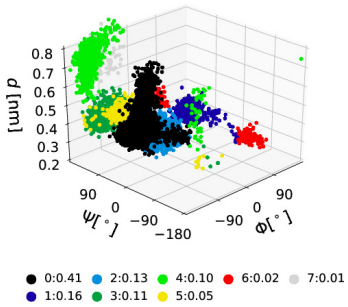
0:0.35 3:0.11 6:0.04 8:0.01 10:0.01
1:0.14 4:0.05 7:0.02 9:0.01 11:0.01
2:0.12 5:0.04

OPLS/TIP3P:Glu



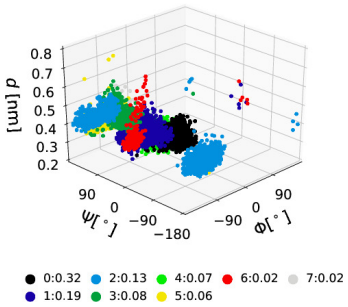
0:0.36 2:0.15 4:0.05 6:0.02 7:0.02
1:0.17 3:0.05 5:0.02

Amberff99SB-ILDN/TIP3P:Lys



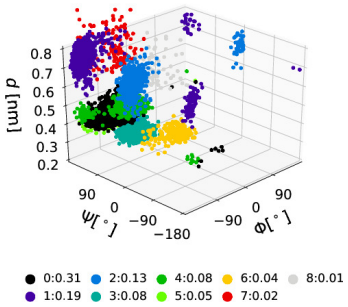
0:0.41 2:0.13 4:0.10 6:0.02 7:0.01
1:0.16 3:0.11 5:0.05

Amber03w/TIP4P2005:Lys



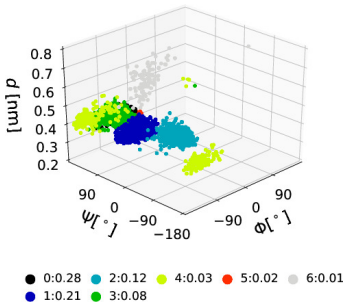
0:0.32 2:0.13 4:0.07 6:0.02 7:0.02
1:0.19 3:0.08 5:0.06

CHARMM36/modified-TIP3P:Lys



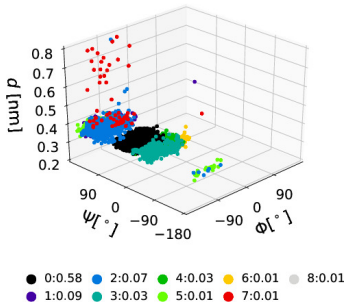
0:0.31 2:0.13 4:0.08 6:0.04 8:0.01
1:0.19 3:0.08 5:0.05 7:0.02

OPLS/TIP3P:Lys



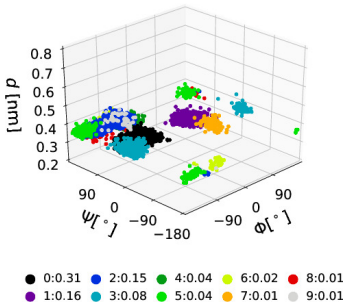
0:0.28 2:0.12 4:0.03 5:0.02 6:0.01
1:0.21 3:0.08

Amberff99SB-ILDN/TIP3P:Arg



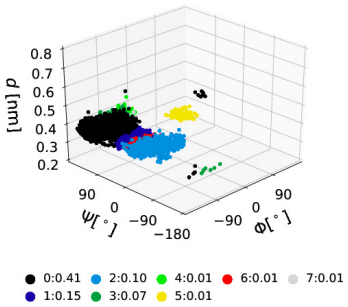
0:0.58 2:0.07 4:0.03 6:0.01 8:0.01
1:0.09 3:0.03 5:0.01 7:0.01

Amber03w/TIP4P2005:Arg



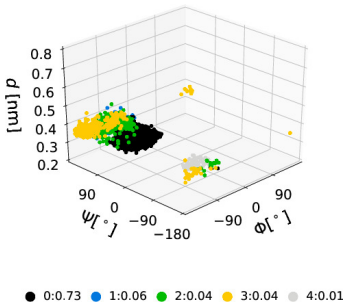
0:0.31 2:0.15 4:0.04 6:0.02 8:0.01
1:0.16 3:0.08 5:0.04 7:0.01 9:0.01

CHARMM36/modified-TIP3P:Arg



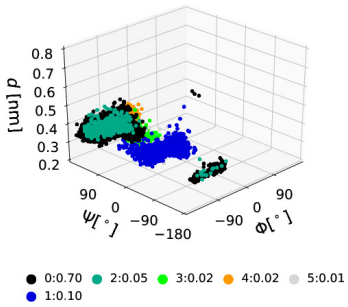
0:0.41 2:0.10 4:0.01 6:0.01 7:0.01
1:0.15 3:0.07 5:0.01

OPLS/TIP3P:Arg



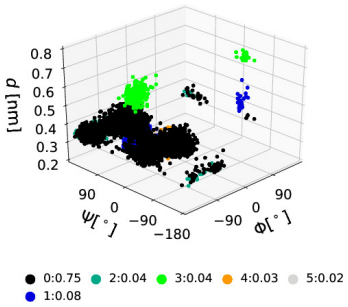
0:0.73 1:0.06 2:0.04 3:0.04 4:0.01

Amberff99SB-ILDN/TIP3P:His



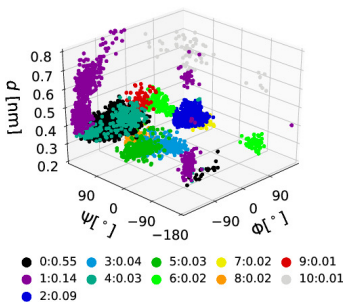
0:0.70 2:0.05 3:0.02 4:0.02 5:0.01
1:0.10

Amber03w/TIP4P2005:His



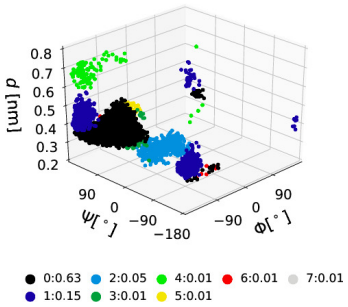
0:0.75 2:0.04 3:0.04 4:0.03 5:0.02
1:0.08

CHARMM36/modified-TIP3P:His



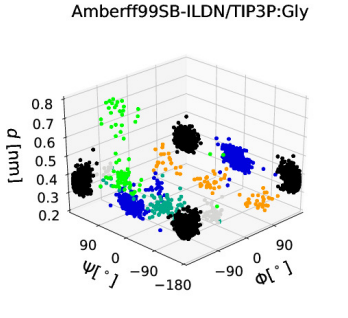
0:0.55 3:0.04 5:0.03 7:0.02 9:0.01
1:0.14 4:0.03 6:0.02 8:0.02 10:0.01
2:0.09

OPLS/TIP3P:His



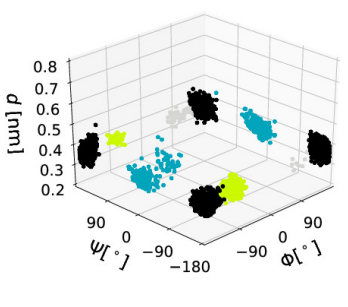
0:0.63 2:0.05 4:0.01 6:0.01 7:0.01
1:0.15 3:0.01 5:0.01

Amberff99SB-ILDN/TIP3P:Gly



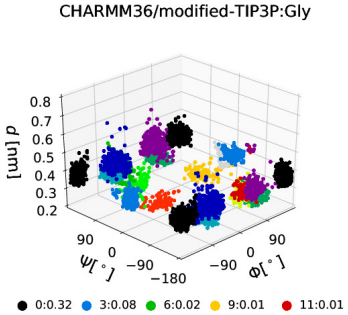
0:0.82 2:0.01 3:0.01 4:0.01 5:0.01
1:0.12

Amber03w/TIP4P2005:Gly



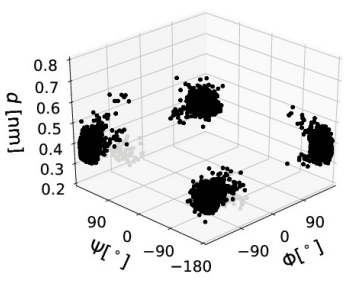
0:0.81 1:0.05 2:0.03 3:0.01

CHARMM36/modified-TIP3P:Gly

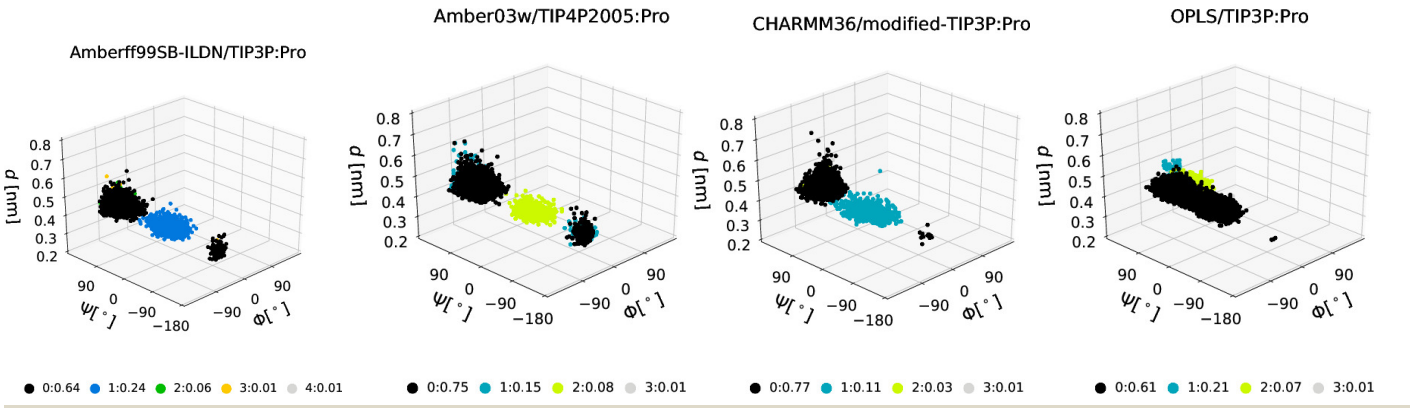


0:0.32 3:0.08 6:0.02 9:0.01 11:0.01
1:0.10 4:0.05 7:0.01 10:0.01 12:0.01
2:0.10 5:0.05 8:0.01

OPLS/TIP3P:Gly



0:0.98 1:0.01



References

1 R. B. Best and J. Mittal, *J. Phys. Chem. B*, 2010, **114**, 14916–14923.

2 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *Journal of computational chemistry*, 2003, **24**, 1999–2012.

3 J. W. Ponder and D. A. Case, *Force fields for protein simulations*, Elsevier, 2003, vol. 66, pp. 27–85.

4 A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. R. Stote, J. Straub, J. Watanabe, M.and Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

5 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

6 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Structure, Function, and Bioinformatics*, 2010, **78**, 1950–1958.

7 A. D. Mackerell Jr, M. Feig and C. L. Brooks III, *J. Comput. Chem.*, 2004, **25**, 1400–1415.

8 A. D. MacKerell Jr, *J. Comput. Chem.*, 2004, **25**, 1584–1604.

9 R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell Jr, *J. Chem. Theory Comput.*, 2012, **8**, 3257–3273.

10 M. J. Robertson, J. Tirado-Rives and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2015, **11**, 3499–3509.

11 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19–25.

12 L. H. Kapcha and P. J. Rossky, *J. Mol. Biol.*, 2014, **426**, 484–498.

13 W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.

14 O. Carugo and K. Djinović-Carugo, *Acta. Crystallogr. D.*, 2013, **69**, 1333–1341.

15 R. J. Campello, D. Moulavi, A. Zimek and J. Sander, *ACM T Knowl. Discov. D.*, 2015, **10**, 5.

16 L. McInnes, J. Healy and S. Astels, *The Journal of Open Source Software*, 2017, **2**, 205.

17 L. McInnes and J. Healy, Data Mining Workshops (ICDMW), 2017 IEEE International Conference on, 2017, pp. 33–42.

18 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

19 S. A. Hollingsworth and P. A. Karplus, *Biomolecular concepts*, 2010, **1**, 271–283.

20 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.