

Simulations of nucleic acids and lipids

Initial MD simulation studies were on globular proteins in vacuum, showing the structures to be reasonably stable in those conditions. Early studies on nucleic acids and lipids, however, were not successful, with the structures becoming highly distorted. This was due to the balance of physical interactions responsible for nucleic acid and lipid structure and dynamics being in a very subtle balance. To achieve this balance improvements in theoretical methodology (e.g. Particle Mesh Ewald, MD in the constant pressure, constant volume ensemble), computational resources (i.e. allowing for simulations with explicit solvent) and in force fields have been made, allowing successful simulations of these systems to be performed.

Solvation

- periodic boundary conditions

- waterball/cylinder (stochastic boundary conditions)

Counterions

Proper Equilibration

EcoRI periodic boundary simulation preparation

PME treatment of electrostatics with 12-10-8 Å truncation scheme, with switch (irrelevant) of electrostatics and vswitch of vdW interactions.

12 Å: nonbond list

10 Å: truncation of vdW interactions

8 Å: truncation of real space electrostatic calculation and initiation of vdW switching function

1) Solvent boxes (**na_duplex/waterbox_prep.inp**, need large version of CHARMM)

CHARMM TIP3P model

Rectangle (orthorhombic)

Pure water or with counterions (MgCl₂ etc. possible)

Use preequilibrated water or saltwater box

2) Initial DNA structure (**na_duplex/gen_overlay.inp**)

A) canonical forms (modeling software packages, CGCGAATTCGCG dodecamer, canonical B form)

B) crystal structures from the Nucleic Acid Database(Berman *et al.*, 1992)

C) convert nucleotide and residue names to CHARMM convention

D) align DNA along axis

E) input correct nucleotide and atom names for CHARMM convention (versus pdb convention)

Ade -> A etc,

3) Solvation of DNA (**na_duplex/gen_overlay.inp**)

A) generate large solvent box to solvate DNA by minimum of 8 Å in all directions (based on real space cutoff distance)

B) delete solvent molecules with non-hydrogen atom < 1.6 Å from DNA non-hydrogen atom

C) add or delete ions to obtain a electrically neutral system (important for EWALD)

DNA has -22 total charge

Add required sodiums at random positions in solvent region

MacKerell, nucleic acids and lipid notes

Delete ions which are furthest from the DNA

4) Solvent equilibration (**dyn_harm_1a.inp**)

A) apply mass weight harmonic constraints of 5 kcal/mol to DNA non-hydrogen atoms

B) minimize for 50 ABNR steps w/ SHAKE, DNA fixed and harmonic constraints (2.0) on the water oxygens

E) 20 ps NVT simulation w/ SHAKE

5) Production simulations (**dyn_pme_1a.inp**)

A) minimize for 100 ABNR steps w/SHAKE

B) NPT dynamics simulation

Langevin Piston (corresponds to Gibbs free energy)

300 K

2 fs timestep

nonbond list updated heuristically

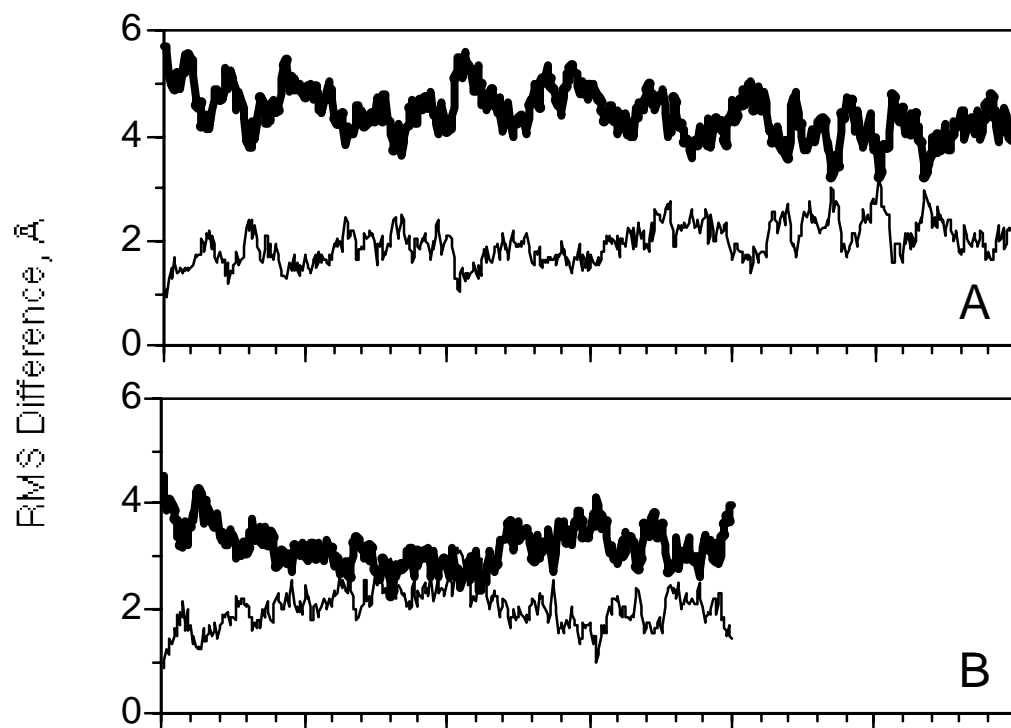
image centering

coordinates saved every 50 to 500 steps for analysis

C) duration

initial 500 ps is considered equilibration

1500 ps of sampling



RMS difference of the non-hydrogen DNA atoms from the EcoRI (A) and CATT decamer MD simulations. Results versus the canonical A (thick line) and B (thin line) are shown. Structures were least-square fit prior to determining the rms differences.

Selected helicoidal parameters from the MD simulations, canonical structures and a survey of crystal structures in the Nucleic Acids Database (see **FREEHELIX**(Dickerson, 1998) **program developed to read CHARMM trajectories**)

Source	Rise	Inclination	Roll	Twist	X-disp	Prop. Twist	Slide
Canonical, fibre diffraction							
A-DNA	2.56	20.7	0.0	32.7	-5.28	-7.5	0.0
B-DNA	3.38	1.5	0.0	36.0	0.0	-13.3	0.0
NDB survey							
A-DNA	2.8±0.4	14.0±4.9	6.9±4.7	32.3±3.8	-4.1±0.8	-9.2±6.3	-1.4±0.4
B-DNA	3.4±0.5	2.8±7.7	1.9±6.0	35.7±5.5	0.0±1.4	-11.3±8.2	0.4±0.7
RNA	2.5±0.8	14.6±13.0	8.1±17.7	29.9±12.9	-3.6±3.6	-3.2±27.1	-2.6±1.7
MD simulations							
EcoRI	3.4±0.4	8.7±3.0	5.6±3.0	34.9±2.5	-0.9±0.4	-11.4±5.4	0.1±0.3
CATTT	3.4±0.2	7.3±1.9	5.2±4.6	33.4±2.3	-1.2±0.1	-11.1±5.4	-0.2±0.5
Hexamer, A	2.5±0.4	25.3±7.7	10.8±13.1	35.8±3.1	-3.5±0.4	-12.8±7.9	-1.0±0.2
Hexamer, B	3.4±0.5	7.8±2.1	5.3±6.3	36.3±3.2	-0.3±0.0	-13.3±0.6	0.3±0.1
RNA	2.6±0.2	14.5±3.8	2.8±6.5	31.3±3.4	-2.6±2.7	-5.4±5.3	-2.3±2.0

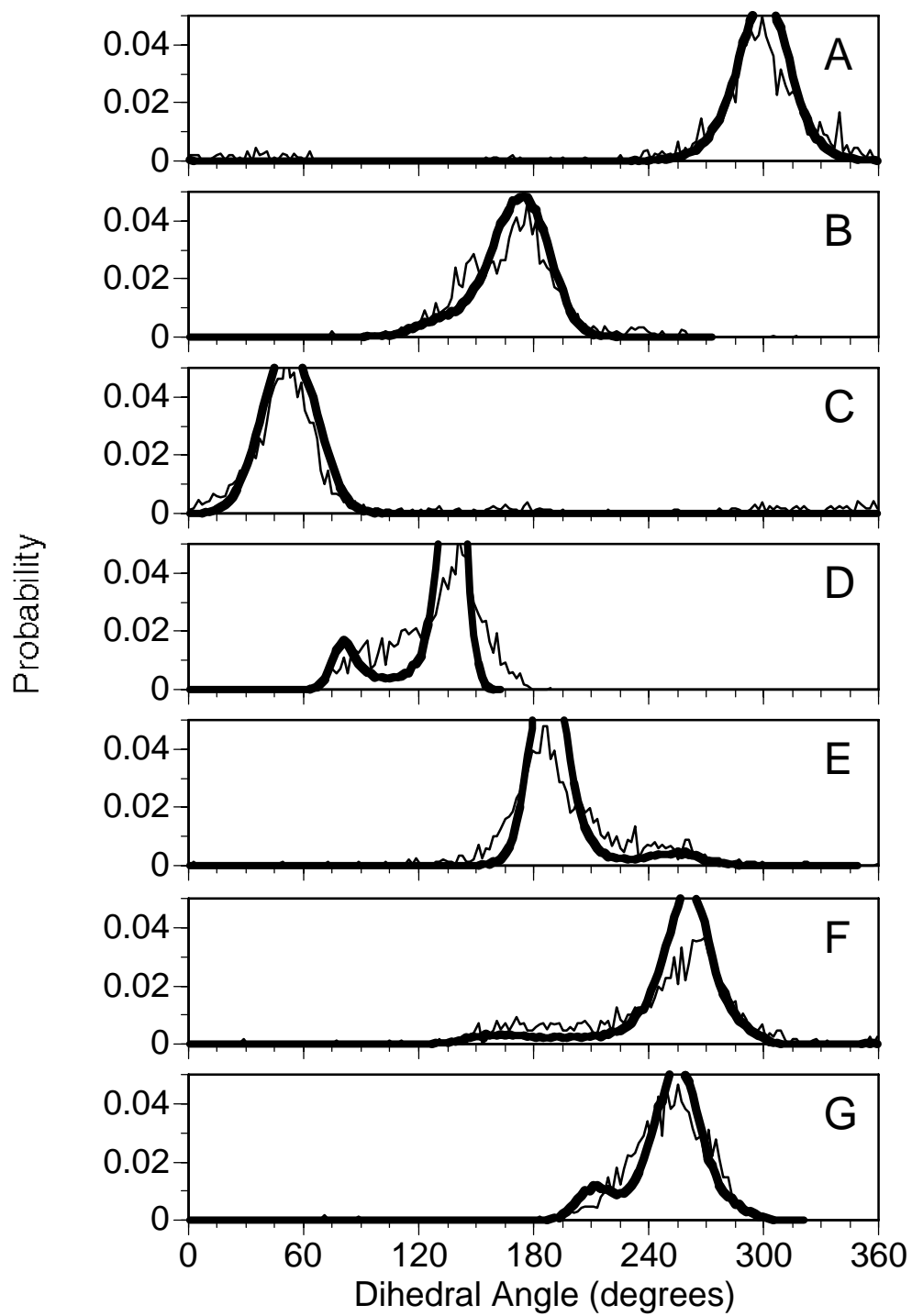
Terms calculated with FREEHELIX are as follows: Rise and twist are associated with strand 1 and inclination, roll, twist, X-displacement, propeller and slide are calculated for the base pairs. The use of only strand 1 is due to a large number of crystal structures having a single strand as the asymmetric unit. See original references for more details. Averages and standard deviations from the MD simulations were over all nonterminal bases or basepairs in the duplexes with the values for the individual bases or base pairs determined from averages of five individual block averages.

Also see Curves(Lavery & Sklenar, 1989) **(Dials and Windows)** (Ravishanker *et al.*, 1989) **and others**(Lu *et al.*, 1999).

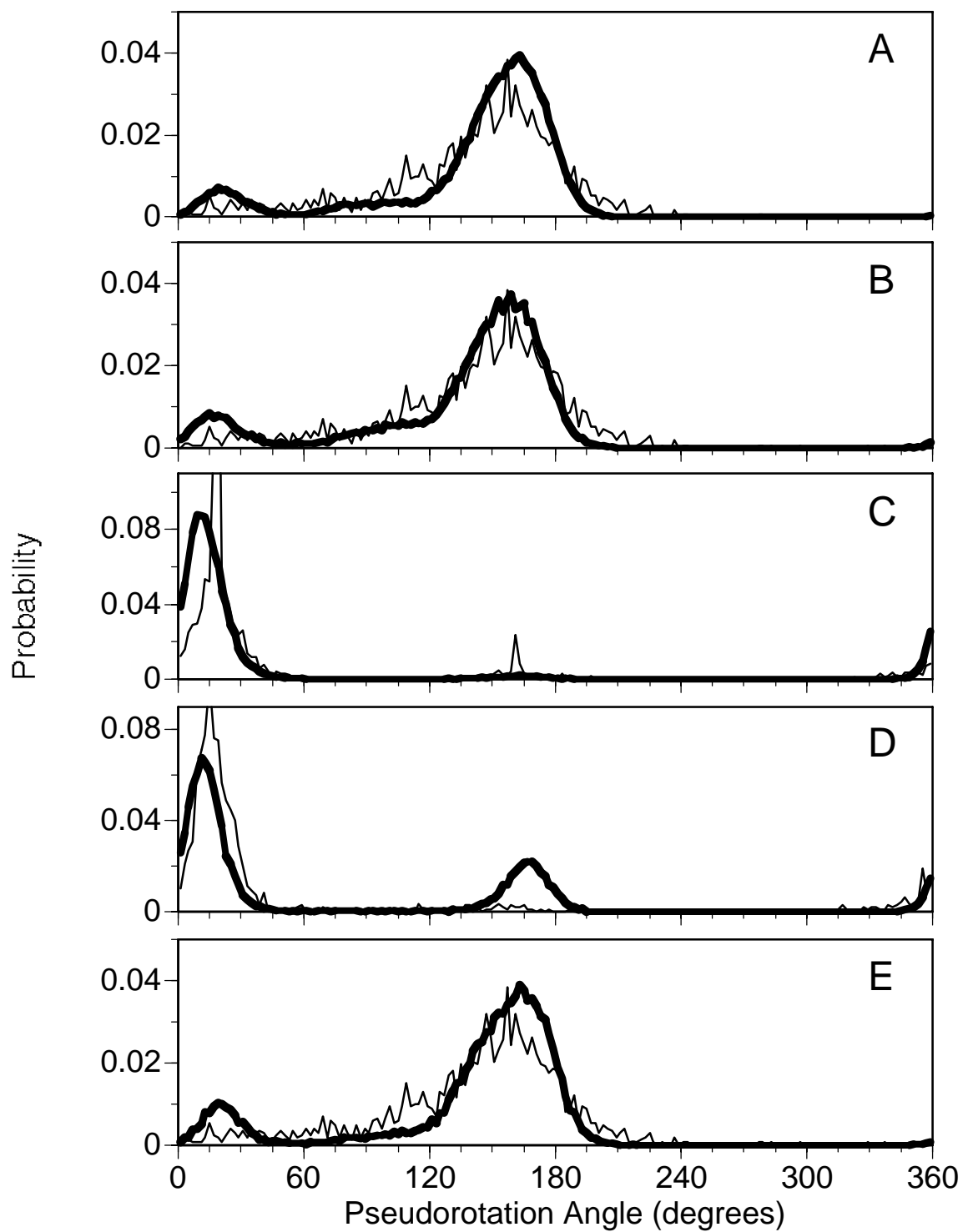
Hydration numbers for selected atoms from the EcoRI, CATTTCATC and RNA UAAGGAGGUGUA solution simulations (see **na_duplex/solv_dna_inter_1.inp**).

Atoms	EcoR1		CATTTCATC		RNA	
	#/atom	#/base	#/atom	#/base	#/atom	#/base
DNA	0.50±0.00	10.22	0.51±0.00	10.38	0.54±0.00	11.40
Bases						
Pyrimidines	0.39±0.00	3.30	0.39±0.00	3.43	0.45±0.00	3.64
Purines	0.37±0.00	3.87	0.35±0.00	3.59	0.40±0.00	4.22
Minor	0.83±0.00	2.49	0.79±0.00	2.38	0.82±0.01	2.47
Major	0.79±0.01	2.38	0.79±0.01	2.36	1.12±0.01	3.37
C8	0.69±0.01		0.58±0.01		0.68±0.01	
C6	0.49±0.01		0.39±0.01		0.52±0.00	
Backbone						
Sugar O4'	1.26±0.01		1.23±0.01		0.96±0.01	
Ester O	0.92±0.01	1.83	0.94±0.01	1.88	0.88±0.01	1.75
Anionic O	3.02±0.03	6.04	3.04±0.01	6.08	2.80±0.01	5.60
Sugar O2'					2.51±0.01	

Values represent the hydration numbers and standard errors from five 500 ps blocks from the EcoRI simulation and five 300 ps blocks from the CATTTCATC and RNA simulations. Hydration numbers per atom (#/atom) were determined based on all waters within 3.5 Å of the selected atoms and averaged over values from windows every 10 ps in each block. Values are normalized with respect to the time frame and the the number of DNA atoms. Hydration number per base (#/base) was obtained by multiplying the #/atom values by the total number of atoms and dividing by the associated number of bases, except in the case of the minor and major grooves where #/base represents the total number of waters hydrating the basepairs comprising the grooves. DNA represents all non-hydrogen atoms, Base represents all base non-hydrogen atoms, Minor represents the purine N2,N3 or pyrimidine O2 atoms in the minor groove and Major represents the purine O6,N6,N7 or pyrimidine N4,O4 in the major groove



See `na_duplex/bkb_dihedral_corr.inp`



See `na_duplex/ana_dyn_puck.inp`

Base pairing: see `na_duplex/dist_bp_1.inp`

DNA-Protein systems

The majority of protein-DNA systems are too large for periodic boundary simulations. Accordingly, explicit solvent simulations supplemented with reaction field methods (e.g. stochastic boundary dynamics (Brooks & Karplus, 1983)) are appropriate. These approaches allow for the region of interest of the system to be studied in detail while minimizing computer costs. Such approaches are also appropriate for studying enzyme catalyzed chemical reactions, where the action is focussed on the active site.

Cytosine-5-methyltransferase from *HhaI* (M.*HhaI*)-S-adenosylhomocysteine (SAH)-DNA ternary complex.

Atom truncation via 14-12-10 Å distances

14 Å: nonbond list

12 Å: electrostatic (fshift) and vdW (vswitch) cutoff distance

10 Å: initiation of vdW switching function

Alternative would be the use of Extended electrostatic, where the electrostatic contributions beyond the cutoff distance are included via a multipole expansion (Stote *et al.*, 1991).

1) Solvent boxes (**na_duplex/waterbox_prep.inp**, need large version of CHARMM)

CHARMM TIP3P model

Create sphere of desired size to solvate region of interest

Neutral system not required, but may be desirable

Specific ions may impact property of interest, so be careful

2) Initial Protein-DNA structure (**na_ternary/gen_hhamt_ternary.inp**)

A) crystal structures from the PDB or NDB (Berman et al., 1992)

B) convert nucleotide and residue names to CHARMM convention

Ade -> A etc,

C) generate DNA, followed by protein and coenzyme

Swap xtal DNA coordinates with canonical DNA coordinates

3) Solvation of DNA (**na_ternary/gen_hhamt_ternary.inp**)

A) translate system so region of interest is at the origin

B) overlay large solvent sphere onto solute (sphere size should be large enough to avoid edge effects; distances from “essential region” to edge should be > nonbond truncation distance)

C) delete solvent molecules overlapping solute

with non-hydrogen atom < **1.6 Å** from DNA non-hydrogen atoms

with non-hydrogen atom < **2.5 Å** from protein non-hydrogen atoms

C) add or delete ions to obtain an electrically neutral system (if desired)

Add required sodiums at random positions in solvent region

Delete ions which are furthest from the DNA

4) Production simulations (**dyn_hhamt_1.inp**): constraints allow for more rigorous system preparation versus that for DNA alone.

A) fix protein residues with no atoms < 25 Å from center of water sphere

B) harmonic constraints (force 2.0, mass weighted)

protein residues with no atoms < 21 Å from center and not fixed

DNA terminal nucleotides

C) apply potential to maintain water density

Mean Field Molecular Potential (MMFP) (Beglov & Roux, 1994)

Stochastic boundary (Brooks & Karplus, 1983)

D) 50 step Steepest-Decent (SD) minimization

E) NVT MD simulation based on Nosé-Hoover Thermostats

Preparation of lipid simulations

Bilayers(Tobias *et al.*, 1997)

1) Do it yourself

Pack preequilibrated lipids along a hexagonal array

spacing based on surface area per lipid

Initial equilibration aliphatic tails with head group positions fixed

gradual minimization to remove bad contacts

MD simulation

Prepare solvent/counterions (if required) as with nucleic acids

Overlay lipid bilayer

Remove solvent in the bilayer interior if desired

2) Have somebody else do it.

Snapshots from previous MD simulations of bilayers are available

previously equilibrated and ready to go.

DPPC (see **lipid/gen_dppc_bilayer.inp**) (Feller *et al.*, 1997a; Venable *et al.*, 1993)

DOPC (see **lipid/gen_dopc_bilayer.inp**) (Feller *et al.*, 1997b)

Bilayer-protein systems (MacKerell, 1995)

1) pack preequilibrated lipids around protein

translate and rotate lipid/protein as rigid units to obtain optimum packing
solvate etc.

2) take preequilibrated bilayer, remove lipids/solvent, insert protein

Minimization/Simulation Protocol

PME

Follow approach for DNA duplexes

SDS Micelle simulation(MacKerell, 1995)

1) Generation of initial model

Aggregation number in 50 mM NaCl is approximately 62
Distribute head groups based on C60 Buckminster Fullerene
build lipids along vectors from carbons to center of sphere
all-trans geometries
terminal methyls placed 3.0 Å from center
paraffinic radius of 16.7 Å
total radius of 22-23 Å

2) Relax vdW contacts (in vacuum)

100 step SD minimization with 5 kcal/mole harmonic constraints on all non-hydrogen atoms and sulfur atoms fixed
300 step ABNR minimization with 1 kcal/mole harmonic constraints on all non-hydrogen atoms and sulfur atoms fixed
7 ps MD simulation
gradual heating from 0 to 298 K over 2 ps
0.001 ps timestep
100 kcal/mole/Å harmonic constraints on sulfur atoms
SHAKE
5 ps equilibration (velocity scaling)

3) Micelle solvation

overlay with 57.8x57.8x57.8 Å cube of water
waters with oxygen within 2.8 Å of micelle non-hydrogen atom removed
100 step SD minimization with micelle fixed
1 ps thermalization of waters
hot start at 298 K
micelle fixed
0.002 ps timestep
waters within 12 Å of geometric center of micelle removed
experiment indicates micelle interior does not contain water

4) Addition of ions(Jung, 1989 #658)

Insure that a sodium ion is close to head groups (may also be used for DNA)

i) water hydrogens deleted

ii) each water oxygen within 8.5 Å of a sulfate sequentially replaced by a sodium

iii) total interaction energy with environment determined

iv) lowest energy position water replaced with a sodium

v) ii to iv repeated for each lipid molecule

vi) water hydrogens replaced

Water molecules relaxed

sodium and micelle fixed

100 step SD minimization

2 ps MD thermalization

Solvent relaxed

2 ps MD thermalization with micelle fixed

Relax entire system

6 ps NPT simulation (length based on obtaining 1 ATM pressure)

edglength decreased from 57.8 Å to 54.1 Å

Production simulation

NVT ensemble

velocity scaling every every 0.1 ps if the temperature was ± 5 K from 298 K

120 ps total simulation time

or get the previously equilibrated SDS micelle in solution (see **lipid/gen_micelle_1.inp**)

Error analysis of single simulations (Feller et al., 1997b; Loncharich *et al.*, 1992)

RMS fluctuations are not errors; they indicate how the system is moving (if you ran the simulation out to infinity, the “error” based on RMS fluctuations would not go to zero, it would converge to a finite value).

Proper error analysis is based on independent samples

1) Perform multiple MD simulations

different starting configuration or random number seed
average and standard deviation or standard error over averages from individual simulations

2) Break single MD simulation into separate, independent blocks

test independence via analysis of autocorrelation function of property of interest
correlation time should be ca. 4 times shorter than total block
average of averages from the individual blocks
standard error

$$S.E. = \frac{\sigma}{\sqrt{N}}$$

where σ is the standard deviation and N is the number of blocks.

Typically, perform 500 ps equilibration followed by a 1500 ps production simulation, where the 1500 ps are broken up into 5 300 ps blocks for analysis. Note that different properties will require different block sizes (too long a block is fine, too short and the blocks are not independent samples).

References

- Beglov, D. & Roux, B. (1994). Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050-9063.
- Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). The Nucleic Acid Database: A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751-759.
- Brooks, C. L. & Karplus, M. (1983). Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.* **79**, 6312.
- Dickerson, R. E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucl. Acids Res.* **26**, 1906-1926.
- Feller, S. E., Venable, R. M. & Pastor, R. W. (1997a). Computer Simulation of a DPPC Phospholipid Bilayer: Structural Changes as a Function of Molecular Surface Area. *Langmuir* **13**, 6555-6561.
- Feller, S. E., Yin, D., Pastor, R. W. & MacKerell, J., A.D. (1997b). Molecular Dynamics Simulation of Unsaturated Lipid Bilayers at Low Hydration: Parametrization and Comparison with Diffraction Studies. *Biophys. J.* **73**, 2269-2279.
- Lavery, R. & Sklenar, H. (1989). Defining the structure of irregular nucleic acids: convention and principles. *J. Biomol. Struct. Dynam.* **6**, 655-667.
- Loncharich, R. J., Brooks, B. R. & Pastor, R. W. (1992). Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers* **32**, 523.
- Lu, X.-J., Babcock, M. S. & Olson, W. K. (1999). Overview of Nucleic Acid Analysis Programs. *J. Biomol. Struct. Dyn.* **16**, 833-843.
- MacKerell, A. D., Jr. (1995). Molecular Dynamics Simulation Analysis of a Sodium Dodecyl Sulfate Micelle in Aqueous Solution: Decreased Fluidity of the Micelle Hydrocarbon Interior. *J. Phys. Chem.* **99**, 1846-1855.
- Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R. & Sklenar, H. (1989). Conformational and Helicoidal Analysis of 30 ps of Molecular Dynamics on the d(CGCGAATTCGCG) Double Helix: "Curves", Dials and Windows. *J. Biomol. Struct. Dyn.* **6**, 669-699.
- Stote, R. H., States, D. J. & Karplus, M. (1991). On the treatment of electrostatic interactions in biomolecular simulation. *J. Chim. Phys.* **88**, 2419-2433.
- Tobias, D. J., Tu, K. & Klein, M. L. (1997). Atomic-scale molecular dynamics simulations of lipid membranes. *Curr. Opin. Coll. & Inter. Sci.* **2**, 15-26.
- Venable, R. M., Zhang, Y., Hardy, B. J. & Pastor, R. W. (1993). Molecular Dynamics Simulations of a Lipid Bilayer and of Hexadecane: An Investigation of Membrane Fluidity. *Science* **262**, 223-226.