User notes on simulation code,
GROMACS **I**ncluding **M**onte Carlo **M**utation **M**oves for **M**ixtures of **Li**pids
GIMLi-1.0
(not official GROMACS!)

Purpose: The purpose of this modified version of Gromacs 3.3.1 is to perform an isomolar semi-grand canonical ensemble (constant $N,\Delta\mu,T$ – depending on settings, can be also constant p,γ or A) simulation of a mixture of 2 lipid types with different tail structures. Why would you want to do this? To achieve an equilibrated lateral distribution of the lipids in your mixture (potentially) much faster than you can by waiting for your lipids to diffuse around your system and sample different arrangements. This can be particularly interesting if there are different microenvironments in your system, determined by proximity to a membrane protein or other perturbation to the bilayer. See Bibliography below for some applications.

Credit: The modifications to the code (e.g., contents of file "mc.c") were written by Jason de Joannis, with minor edits by James Kindt and Fuchang Yin. We thank the Gromacs developer community for providing the MD code and for providing essential assistance and expertise, without which this would not have been possible. This work has been supported by NSF grant CHE-0616383. We request that (along with the references given in the md.log files for Gromacs and other methods) users of this modified version cite the following paper:[1]

J. de Joannis, Y. Jiang, F. Yin, and J. T. Kindt, "Equilibrium distributions of DPPC and DLPC in a mixed lipid bilayer: Atomistic semi-grand canonical ensemble simulations." *Journal of Physical Chemistry B*. **110**, 25875-25882 (2006).

Disclaimer: Use at your own risk. We make no guarantees about the method or its implementation. Our liability in the event that any user sustains any damages is limited by the GNU General Public License (GPL). Please send questions, suggestions, bug reports or bug fixes to James Kindt at jkindt[at]emory.edu; we will do our best to be accommodating. If you would like to hear about updates, bugs found, etc. then also please send us an e-mail and we will put you on a list. *THIS IS NOT OFFICIAL GROMACS; please do not bother the Gromacs developers with questions pertaining to the use of this code.*

Conditions for use: You are free to use and modify this code at no cost and to distribute it or its modifications within the limits set forth by the GPL. If you do use it, we'd like to hear from you!

Some more details:
For a simulation of a mixture of N lipids containing two types (A and B), the Gromacs topology file used contains a set of N type A lipids and a corresponding set of N type B

[1] Note that the 2006 de Joannis publication describes work using an earlier prototype of the MC/MD code, which only treated mixtures differing in tail length. In particular, for the current version, the discussion of "ghost tails" is moot.

lipids. At each time step, for each of the N lipids, one version (A or B) is treated according to its full potential for MD purposes while its counterpart is treated as a "ghost" with its non-bonded interaction forces turned off entirely. After every MD step (immediately before forces are generated for the next step), a lipid i is randomly selected. A new, trial conformation for the lipid's "ghost" counterpart is generated, preserving the positions of all sites that the two lipid types share in common while regrowing one or both tails from the node or nodes of structural divergence using the configurational bias Monte Carlo method of Siepmann and Frenkel. An acceptance probability is calculated based on the Rosenbluth weights of the original and final tail structures and the desired $\Delta\mu$. If the mutation move is successful, the current lipid type becomes the "ghost" and the trial conformation is adopted for its counterpart, which has its non-bonded interactions restored. In the event of a successful mutation, all re-grown sites are given new velocities with a Maxwell distribution of magnitudes and random orientations, and a full neighbor list update is performed before the forces are calculated for the next MD step.

Important information for setting up a simulation:

1) The code allows mutations between two different lipid types, which shall be called "A" and "B". "A" and "B" must have the same structure *except* for two unbranched chain segments, each of which must terminate with a methyl group (site named "LP3"). The partial charge at every site on both segments must be zero.[2] The tails may differ in number of sites, bond lengths (which must be constrained), angles, dihedrals, Lennard-Jones parameters, and 1-4 pair interactions.

2) The topology and input structural files must follow this order: [3]

i. all "A" lipids (N molecules)
ii. all "B" lipids (N molecules)
iii. any other molecules not involved in mutations
(e.g. protein, cholesterol, other lipids, solvent.)

Furthermore, the code will look for one atom type "HW" whose Lennard-Jones parameters will be used for all ghost atoms – there must be one molecule defined in the system containing a site of type "HW" with no Lennard-Jones interactions.

3) At any point during the simulation, N_A of the "A" lipids are fully active for MD and MC calculation, while the remainder (N-N_A = N_B) of the "A" lipids have intermolecular interactions turned off. Likewise N_B of the "B" lipids are active and the remainder are ghosts. Ghost lipids can be distinguished by a shift in their x coordinates of +100 nm.

---

[2] The code uses the configuration-bias chain growth method of Siepmann and Frenkel for mutation moves, and an unbranched chain topology is assumed. Only bonded and L-J energies are incorporated into trial move generation and acceptance probabilities, so partial charges will not be included correctly. The code uses "LP3" as a chain termination indicator in defining the regions of common structure between A and B.

[3] Lipid "A" is actually defined as the first type listed under `tc_grps`, not the first lipid listed in the input configuration and topology files. We recommend that you maintain a consistent order of lipids in all lists to avoid confusion. The choice of which lipid is "A" and which is "B" is unimportant; keep in mind, however, that if you switch "A" for "B", you must invert `mcactivityratio` to obtain consistent results.

Visualization of any output (e.g., confout.gro or traj.trr) will show the currently active lipids in the normal simulation box and the "ghosts" in a box shifted +100 nm in the *x* direction. For input structures, for each A/B pair of lipids (i.e. "A" lipid with index *i*, "B" lipid with index *N*+*i*), the real one should have regular input coodinates and its "ghost" counterpart should have coordinates shifted by +100 nm in the *x* direction. This is how the program will decide which lipids' interactions to turn on or off at the start of the simulation. *Exception*: *if you do not shift any lipids, the program will assume that all "A" lipids are active and all "B" lipids are not – i.e., the simulation will start with 100% lipid "A".*

4) Three additional parameters should be included in the .mdp file:

`mckmax:`

represents k_{max}, the number of random trial positions generated (selected from a Boltzmann distribution of bond and dihedral angles) for each segment grown. MC simulation time will scale roughly linearly with k_{max}, while acceptance probabilities will plateau for high k_{max}. We have found $k_{max} = 4$ or 8 can work well. Should not affect results, only simulation efficiency.

`mcactivityratio:`

represents $\alpha = a_A/a_B$, the ratio of thermodynamic activities (or fugacities) of lipids A and B, which is related to the difference in chemical potentials as $\alpha = a_A/a_B = \exp[(\mu_A-\mu_B)/(k_BT)]$. Chemical potentials are defined with respect to some standard state. In this case, the standard states of A and B correspond to ideal (non-interacting) gases at some common number density.[4] In the limit that "A" and "B" have negligible differences, α would correspond to the average mole ratio $\langle N_A/N_B \rangle$, and setting `mcactivityratio` equal to 1.0 would yield a 50:50 mixture. In the more common case, the two lipids have different sizes and therefore different free energies of interaction with the bilayer. (For instance, in DLPC/DPPC mixtures at 323 K, an activity ratio of $\alpha = a_{DLPC}/a_{DPPC} = 4000$ yielded a roughly 50:50 mixture – see examples provided.) To find a rough calibration of what activity ratios will lead to what compositions, it can be useful to perform an initial series of short simulations with α varied by powers of 10.[5]

`mcseed:`

random number seed used for MC. To continue a pseudo-random number sequence from a previous run, set `mcseed` equal to zero, and include in the working directory the file "rng" generated during the previous run. If mcseed is set to zero and file "rng" is not found, then the GSL default seed is used.

---

[4] To note a minor technicality, in this ideal gas reference state, not only are intermolecular forces exactly zero, but the non-bonded intramolecular interactions involving sites that are not common to both lipid types are also neglected. The actual concentration of the reference state need not be defined, as it will not affect the *difference* in chemical potentials.

[5] While the absolute value of α is of little or no physical relevance (since the reference states are unphysical), the variation of α with composition and environment can yield insight into mixing thermodynamics. First, in an ideal mixture that obeys Raoult's law, $\langle N_A/N_B \rangle$ is directly proportional to α. Deviation from this proportionality is indicative of a nonzero excess free energy of mixing. Secondly, two systems simulated at the same temperature, pressure, and α but different environments can be considered to be at equilibrium with each other with respect to exchange of A and B lipids.

5) Temperature coupling: Our practice has been to avoid thermostats (like Berendsen and Nosé-Hoover) that rely on a calculation of the total kinetic temperature of the system, since at any timestep half of the lipids are non-interacting ghosts and should therefore not influence the temperature scaling of the active lipids. Instead we have used stochastic dynamics to regulate temperature. In our experience, using "sd" leads to some differences in properties such as area per headgroup *even when MC is not turned on* relative to Berendsen thermostat for instance, so in comparing MC/MD simulations of mixtures with MD simulations of pure lipids, it is important to run the MD simulations with the same thermostat.
6) Pressure coupling: Pressure coupling must be turned on to allow for equilibration of a system whose components have different sizes; otherwise, the composition of the system will be determined largely by the fixed volume and not by the activity ratio. We generally use semi-isotropic Berendsen pressure coupling with zero surface tension for lipid bilayers. Note that an applied surface tension could result in a spurious shift in composition due to differences in equilibrium area per headgroup. The contribution of "ghosts" has been removed from the virial.
7) Pre-processing step is as normal in Gromacs 3.3.1 (using "mcgrompp" executable included with this package.)
8) Aside from the file `rng` that allows restart of the random number sequence, there is just one new output file `mc.log` . Information at the head of `mc.log` allows a check on whether the code is interpreting the structural differences between lipids correctly. An update on cumulative MC success rate and current composition is logged every 500 steps, along with details of the last MC step (whether successful or not). Note that Gromacs analysis tools do not distinguish between "active" and "ghost" lipids – be extremely careful when using g_energy, g_order, etc.
9) There is not yet a working parallel implementation.

Random "pearls of wisdom":

- expect a run time to be roughly doubled, relative to an MD simulation with the equivalent number of steps
- the most desirable success rate for mutations is not necessarily the highest – ideally, lipid tails should have time to relax via MD updates for at least ~100-200 ps in between successful mutation moves

Bibliography of publications describing applications of this method:

J. de Joannis, Y. Jiang, F. Yin, and J. T. Kindt, "Equilibrium distributions of DPPC and DLPC in a mixed lipid bilayer: Atomistic semi-grand canonical ensemble simulations." *Journal of Physical Chemistry B*. **110**, 25875-25882 (2006).

H. Wang, J. de Joannis, Y. Jiang, J. C. Gaulding, B. Albrecht, F. Yin, K. Khanna, and J. T. Kindt, "Bilayer edge and curvature effects on partitioning of lipids by tail length: Atomistic simulations." *Biophysical Journal* **95**, 2647-2657 (2008).

P. S. Coppock, J. T. Kindt, "Atomistic simulations of mixed-lipid bilayers in the gel and fluid phases." *Langmuir* **25**, 352-359 (2009).