Title

Training and Validation of a Liquid-Crystalline Phospholipid Bilayer Force Field

Journal

Journal of Chemical Theory and Computation, 12(12)

ISSN

1549-9618

Authors

McKiernan, Keri A
Wang, Lee-Ping
Pande, Vijay S

Publication Date

2016-12-13

DOI

10.1021/acs.jctc.6b00801

Peer reviewed

HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2022 November 11.

Published in final edited form as:

J Chem Theory Comput. 2016 December 13; 12(12): 5960–5967. doi:10.1021/acs.jctc.6b00801.

Training and Validation of a Liquid-Crystalline Phospholipid Bilayer Force Field

Keri A. McKiernan[†], **Lee-Ping Wang**[‡], **Vijay S. Pande**[*,†,§,‖,⊥]

[†]Department of Chemistry, Stanford University, Stanford, California 94305, United States

[§]Department of Computer Science, Stanford University, Stanford, California 94305, United States

[‖]Department of Structural Biology, Stanford University, Stanford, California 94305, United States

[⊥]Program in Biophysics, Stanford University, Stanford, California 94305, United States

[‡]Department of Chemistry, University of California—Davis, Davis, California 95618, United States

Abstract

We present a united-atom model (gb-fb15) for the molecular dynamics simulation of hydrated liquid-crystalline dipalmitoylphosphatidylcholine (DPPC) phospholipid bilayers. This model was constructed through the parameter-space minimization of a regularized least-squares objective function via the ForceBalance method. The objective function was computed using a training set of experimental bilayer area per lipid and deuterium order parameter. This model was validated by comparison to experimental volume per lipid, X-ray scattering form factor, thermal area expansivity, area compressibility modulus, and lipid lateral diffusion coefficient. These comparisons demonstrate that gb-fb15 is robust to temperature variation and an improvement over the original model for both the training and validation properties.

Graphical Abstract

[*]**Corresponding Author** pande@stanford.edu.

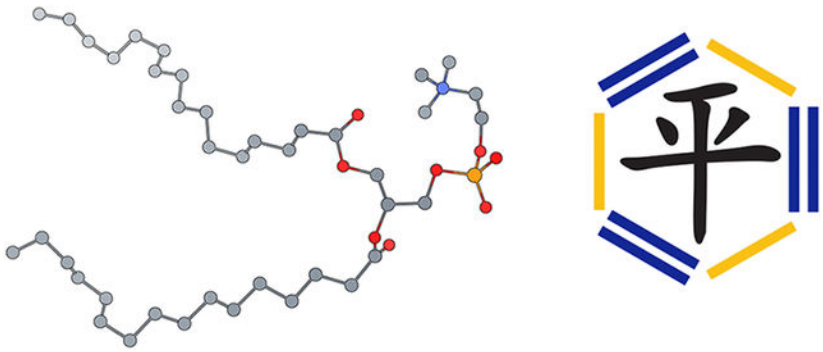
Author Contributions

KM and LPW wrote the paper. KM, LPW, and VSP edited the paper. VSP supervised the project.

Supporting Information

TThe Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00801. Numerical gb-fb15 force field parameters, as well as a figure describing the magnitude of the ForceBalance objective function as a function of iteration (ZIP)

The authors declare the following competing financial interest(s): VSP is a consultant and SAB member of Schrodinger, LLC and Globavir, sits on the Board of Directors of Omada Health, and is a General Partner at Andreessen Horowitz. The authors declare no competing financial interest. The gb-fb15 force field and related materials are available online at http://github.com/pandegroup/gb-fb15. The ForceBalance software is open-source and available online at http://github.com/leeping/forcebalance.



INTRODUCTION

Because of varying local concentrations of proteins, saturated fatty acids, unsaturated fatty acids, and sterols, biological membranes exhibit complex, composition-dependent phase behavior.[1-4] Even in heterogeneous model systems, the coexistence of multiple gel and liquid phases, as well as the segregation of different membrane components into domains, has been observed.[5-8] Bilayer phases can be characterized by the lipid acyl chain rotational disorder, S_{CD}, and lateral diffusion rate, D_l. Under physiological conditions, phosphocholine type lipid bilayers typically exist in a disorder liquid-crystalline phase, referred to as the L_d or L_α phase, in which S_{CD} is low (high lipid tail disorder) and D_l is fast (on the order of 1 $\mu m^2\ s^{-1}$).[9-14]

As the boundary of the cell and cellular organelles, membranes are involved in many intra- and intercellular sensory, signaling, transport, and regulatory processes.[15-19] A growing body of evidence suggests that lipids play a direct, functional role in many of these phenomena. Two areas of research related to membrane functional regulation are the theories of mechanosensitivity, and lipid-rafts. Mechanosensitivity describes a protein's ability to respond to the mechanical forces imposed by its local and global lipid environment.[20,21] The lipid-raft hypothesis posits that, in true biological membranes, cholesterol and sphingomyelin lipids associate to form rigid raft-like structures. These rafts are potentially involved in trafficking where they may surround and transport membrane components to different regions of the membrane, and also in signal transduction, where they may, for example, isolate receptors from enzymes involved in nontargeted signaling pathways.[22-24] Although of scientific interest, the time scales over which these processes occur are on the order of nano- to milliseconds. These time scales tend to be faster than what is currently accessible by experiment. It is thought that the method of Molecular Dynamics (MD) simulations can be usefUl for studying the specifics of biological processes, due to its atomistic spatial and temporal resolution.[25-27]

The validity of an MD simulation is predicated on the accuracy of its underlying model. Here the term "model" refers to the system's atomistic description, as well as the functional form and parameters of its potential energy function, or force field.[28] The accuracy of the model is evaluated by comparing structural, thermodynamic, and kinetic properties of the simulated system to reference data, which come from experimental measurements or other

theoretical models. This evaluation is valuable because it could be used to optimize the model or validate its ability to make predictions outside of the training data.

Atomistic descriptions of phospholipids fall into three primary categories: “all-atom” (AA), “united-atom” (UA), and “coarse grained” (CG). AA models represent every atom explicitly. Current, widely used versions of these models include Slipids,[29] Lipid14,[30] CHARMM36,[31] and GAFFlipids.[32] UA and CG models are structurally approximate, grouping atoms together for computational efficiency. UA models remove all nonpolar hydrogen atoms. Within this type of model are the Berger lipids,[33] GROMOS family (43A1-S3,[34] 53a6$_L$,[35] and 53a6 Kukol[36]), C36-UA,[37] as well as the model presented here (gb-fb15). CG models map the atoms of a particular structural motif into a smaller number of beads. Of this model variety, MARTINI[38] and the SDK models[39] are among the most widely used. The AA and UA models mentioned here have quantitatively comparable accuracy, having very low error with respect to the ability to reproduce a standard set of bilayer thermodynamic and kinetic properties. The CG models are qualitatively accurate but tend to have higher quantitative error.[40] Structural coarse-grained approximations decrease the number interactions of a system, and have a smoothing effect on the system’s free energy landscape. This allows for faster phase space sampling, but also can result in an underestimation of system entropy.[41,42] As is generally the case, the optimal model is going to be application and resource dependent. For detailed comparisons of most AA, UA, and CG lipid force fields in use today, see refs 43-47.

MD is a classical theory. The force field is determined through the fitting of empirical data and/or results of quantum mechanical (QM) calculations. It is hoped that by fitting this potential function to experiment, equilibrium ensemble properties and mean-field quantum effects will be accurately described. Many of the original force fields were developed via manual parameter search techniques. Since then, several tools for automatic parameter optimization have been developed. Some major packages used for this purpose include GAFF[48,49] and CGenFF[50-53] for small molecules, and ParamFit,[54] Wolf2 Pack,[55] GROW,[56] and our methodology, ForceBalance,[57,58] for general force field optimization.

In this study, we report the systematic training and validation of an united-atom phospholipid bilayer force field using experimental thermodynamic data. A pure bilayer consisting of dipalmitoylphosphatidylcholine (DPPC) lipids, hydrated by simple point charge (SPC) water[59] molecules, was chosen for our first model because it has been studied in the greatest detail in both theory and experiment. This allows for the most thorough optimization, characterization, and quality evaluation. A united-atom description was chosen for computational efficiency. It was found that our united-atom approximation is capable of capturing the relevant physical and chemical properties of the lipid bilayer, while improving sampling efficiency when compared to AA models. This new model will be referred to as gb-fb15.

METHODS

The Model.

The model presented here follows a standard functional form, given by a classical piece-wise summation over bonded and nonbonded interaction potentials.

$$V(r^N;\,\Lambda) = V_{\text{bonded}}(r^N;\,\Lambda) + V_{\text{nonbonded}}(r^N;\,\Lambda) \tag{1}$$

Note that the potential terms are given as a function of system configuration, r^N, and force field parameter values, λ, represented by the set Λ.

The bonded terms describe the interaction between covalently bonded atoms, and they consist of bond and bond-angle terms, as well as improper and proper torsional dihedral angle terms. The acyl chain torsional dihedral terms are uniquely treated using a Ryckaert–Bellemans potential.

$$\begin{aligned} V_{\text{bonded}}(r^N;\,\Lambda) = {} & V_{\text{bond}}(r^N;\,K_{b_n},\,b_{0_n}) + V_{\text{angle}}(r^N;\,K_{\theta_n},\,\theta_{0_n}) \\ & + V_{\text{dihedral}_{\text{im}}}(r^N;\,K_{\xi_n},\,\xi_{0_n}) \\ & + V_{\text{dihedral}_{\text{proper}_{\text{non-acyl}}}}(r^N;\,K_{\phi_n},\,\delta,\,m_n) \\ & + V_{\text{dihedral}_{\text{proper}_{\text{acyl}}}}(r^N;\,K_{\phi_n},\,\delta,\,m_n) \end{aligned} \tag{2}$$

$$\begin{aligned} = {} & \sum_{n=1}^{N_b} \frac{1}{4} K_{b_n}(b_n - b_{0_n})^2 + \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n}[\cos(\theta_n) - \cos(\theta_{0_n})]^2 \\ & + \sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n}(\xi_n - \xi_{0_n})^2 \\ & + \sum_{n=1}^{N_\phi} K_{\phi_n}[1 + \cos(\delta_n)\cos(m_n\phi_n)] \\ & + \sum_{n=0}^{6} C_n(\cos(\phi - \pi))^n \end{aligned} \tag{3}$$

The i_n variables (where $i = b, \theta, \xi, \phi$) describe the time-dependent simulation values.
The nonbonded terms consist of a Lennard-Jones potential and a Coulombic electrostatic potential.

$$V_{\text{nonbonded}}(r^N;\,\Lambda) = V_{\text{LJ}}(r^N;\,C_6,\,C_{12}) + V_{\text{Coulomb}}(r^N;\,q) \tag{4}$$

$$= \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \left(\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} \right) + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \tag{5}$$

For this study, the initial parameter set consisted of the GROMOS 53a6 (G53a6) force field, modified to include Berger lipid Lennard-Jones (LJ) and tail dihedral parameters. Here we will refer to this model combination as “g53a6-b”. GROMOS 53a6 was originally parametrized for the reproduction of small biomolecule densities, enthalpies of vaporization, and hydration free enthalpies.[60] This force field has been shown to accurately model several experimental structural properties for peptides, proteins, and DNA in explicit SPC water,[61] but it fails to describe the correct phase behavior for hydrated phosphatidylcholine bilayers.[36] The qualitatively correct phase behavior is recovered when GROMOS 53a6 is combined with the aforementioned elements from the Berger lipid parameter set.

Developed in 1997, Berger lipids remain among the most popular lipid bilayer force fields. It borrows its Lennard-Jones (LJ) parameters from the united-atom OPLS force field,[62] modifying the values for the hydrocarbon tail group atoms. These OPLS tail group LJ sigma and epsilon values were found by fitting to experimental density and enthalpy of vaporization measurements.[33] The tail group dihedral potential was taken from Ryckaert and Bellemans, who fit their potential using spectroscopic data of an *n*-butane monomer.[63]

For our parametrization process, we optimized the Lennard-Jones parameters for all lipid atom types (as illustrated in Figure 1). These parameters are described by the C_{12} and C_6 terms in eq 5. The bonded and electrostatic parameters were left unmodified.

Simulation Conditions.

We combined the force field parameters described above following Justin Lemkul’s GROMACS membrane-protein tutorial.[64] The force field topology, along with an equilibrated DPPC bilayer, were obtained from Peter Tieleman’s Web site.[65] The simulated system consists of 128 DPPC united-atom lipid molecules, arranged in two leaflets, hydrated by 3655 water molecules.

Model optimization is an iterative process wherein, at each step, simulations are run using an updated parameter set. These simulations were first energy minimized via a steepest descent algorithm over 10000 steps, and then equilibrated using a Berendsen barostat for 200 ps with a time step of 2 fs. An NPT ensemble was enforced during the production dynamics via a Nose–Hoover thermostat (τ_t = 0.5 ps, ref_t = 323, 333, 338, and 353 K) and a semiisotropically coupled Parrinello–Rahman barostat (τ_p = 2.0 ps, ref_p = 1.0, k_t = 4.5e–5 bar^{-1}). Electrostatics were calculated using a particle-mesh Ewald scheme of order 4 and Fourier spacing of 0.16 nm. The Ewald real space and van der Waals cutoffs were both set to 0.12 nm. The LINCS algorithm was used to fourth order to constrain all bonds to their equilibrium lengths. Note that lipid force fields are highly sensitive to simulation parameters, and should be used with the specifications applied during model development. A GROMACS parameter file has been supplied with the force field developed in this work, available in the GitHub repository described in the Notes. For each temperature point, ten parallel simulations with unique initial conditions were run for 3 ns each in a rectangular box with periodic boundary conditions using a leapfrog integrator. All simulations were performed using the GROMACS 4.6.5 software package.[66-69] Training properties were then calculated for each production simulation, and property averages were calculated using all trajectories belonging to a particular NPT macrostate (40 simulations per iteration, 10 per

temperature, resulting in a 30 ns time series for each thermodynamic property at each temperature point). Pre-equilibrating the production simulations and parallelizing over a range of unique initial conditions was found to drastically accelerate the convergence of the training properties. The final frame from each production simulation was used as the initial condition for the following iteration of the optimization process. The distribution of jobs was facilitated through use of the Work Queue library.[70]

Optimization.

The ForceBalance method addresses the challenge of finding an optimal approximate potential function via a supervised learning method. A set of training data is used to compute an objective function which is then iteratively minimized in the space of force field parameters. The ForceBalance objective function is of a least-squares form and is given by the following expression:

$$\chi^2(\Lambda,\ \alpha,\ \beta) = \sum_{i}^{|T|} \alpha_i (t_i - a_i(\Lambda))^2 + \sum_{i}^{|\Lambda|} \beta_i \lambda_i^2 \tag{6}$$

This objective function is a quantification of model quality, determined by the force field parameter set, $\lambda_i \in \Lambda$, through comparison of simulation properties, $a_i \in O$, to a training set, $t_i \in T$, consisting of empirical and/or ab initio data. Residuals are of a least-squares form and are guaranteed to be relatively small because parameters of a suitable, existing model are used as an initial guess. This ensures that the system's thermodynamic properties are stable and reasonably close to the experimental values. To ensure that each training property contributed as evenly as possible to the optimization process, the units of the training data were set so that all data was of the same order of magnitude, and the α weights were both set to 0.5.

This method implements an L_2 regularization scheme in order to discourage overfitting, given by the second summation in eq 6. The regularization term guarantees that the parameters only deviate from their initial values by an amount that is small compared to the initial parameter values. It is given by

$$\sum_{i}^{|\Lambda|} = \frac{(\lambda - \lambda_0)^2}{P_\lambda} \tag{7}$$

where λ_0 is the initial value of the parameter and P_λ is a hyperparameter representing the allowable deviations of the parameter. To ensure the objective function remains dimensionless, P_λ has the same physical units as λ. We chose the values of the hyperparameter P_λ by inspection of the initial parameter values. Because the initial values of the Lennard-Jones σ and ϵ values were all between 0.2 and 1.0 (in the GROMACS unit system), we chose P_σ = 1.0 nm and P_ϵ = 1.0 kJ/mol. In practice, the RMS percentage change of the parameters from their initial values was 10.9%, and the largest change for any parameter was for the LC2 σ parameter, which increased by 28.5%.

The objective function is minimized using the Levenberg–Marquardt algorithm[71,72] with an adaptive trust radius,[73,74] where parameter shifts are given as follows:

$$\lambda_{i+1} = \lambda_i + \frac{\nabla \chi^2(\Lambda, \alpha, \beta)}{\nabla^2 \chi^2(\Lambda, \alpha, \beta) + bI} \tag{8}$$

The objective function parameter-space gradients are calculated analytically, and the elements of the Hessian are approximated using the Gauss–Newton approximation. I is the identity matrix, and the b term is found by tracking the relative error between optimization iterations, and is used to interpolate between gradient descent (large b, far from a minimum) and the Gauss–Newton method (small b, close to a minimum, where a quadratic approximation is more appropriate).

For high-dimensional small-residual nonlinear least-squares optimization problems, gradient-based Newton-type methods, such as Levenberg–Marquardt, have been shown to exhibit rapid local convergence.[75] When the training properties are not explicitly dependent on the system's potential energy, their parameter-space gradients can be represented analytically.[58] This can be seen if one considers the ensemble average of an observable, A, which is not explicitly dependent on the system's potential energy, as seen in eq 9. Through differentiation, this derivative expression reduces to the analytic fluctuation formula shown by eq 10.

$$\langle A \rangle_\lambda = \frac{1}{Q(\lambda)} \int A(r, V) \exp[-\beta(E(r, V; \lambda)) + PV] \, \mathrm{d}r \, \mathrm{d}V \tag{9}$$

$$\begin{aligned} Q(\lambda) &= \int \exp[-\beta(E(r, V; \lambda)) + PV] \, \mathrm{d}r \, \mathrm{d}V \\ \frac{\mathrm{d}}{\mathrm{d}\lambda}\langle A \rangle_\lambda &= \frac{1}{Q(\lambda)} \int A(r, V) \exp[-\beta(E(r, V; \lambda)) + PV] \\ &\quad \times \left(-\beta \frac{\mathrm{d}E(r, V)}{\mathrm{d}\lambda}\right) \mathrm{d}r \, \mathrm{d}V - \frac{1}{Q(\lambda)^2} \frac{\mathrm{d}Q(\lambda)}{\mathrm{d}\lambda} \\ &\quad \int A(r, V) \exp[-\beta(E(r, V; \lambda)) + PV] \, \mathrm{d}r \, \mathrm{d}V \\ &= -\beta\left(\langle A \frac{\mathrm{d}E}{\mathrm{d}\lambda} \rangle_\lambda - \langle A \rangle_\lambda \left\langle \frac{\mathrm{d}E}{\mathrm{d}\lambda} \right\rangle_\lambda\right) \end{aligned} \tag{10}$$

where $Q(\lambda)$ is the partition function for the isothermal–isobaric ensemble, r is the system's spatial configuration, V is the volume, $\beta \equiv \frac{1}{k_B T}$, where k_B is Boltzmann's constant, T is the temperature, and $E(r, V; \lambda)$ is the potential energy. Analytic computation of these gradients is advantageous because numerical finite-difference approximations are resource intensive and subject to numerical error. Analytic gradients are still subject to sampling error, but as described above, pre-equilibration of the production simulations and analysis across a diverse range of initial conditions allowed for fast convergence of the training properties, resulting in sampling error of less than 5%.

For this optimization procedure, area per lipid and deuterium order parameter were contained in the training set and, hence, used to train the model. For a homogeneous lipid bilayer, area per lipid can be calculated from a simulation using the following equation:

$$\langle A_{\mathrm{L}}(t)\rangle = 2\frac{\langle d_x(t)\cdot d_y(t)\rangle}{n_L} \tag{11}$$

where $d_i(t)$ is the time series for the i component of the box volume, and n_L is the number of lipids in the simulation. The box dimension time series were obtained using the GROMACS routine g_energy. The deuterium order parameter, for each tail CH_2 carbon node (atomtype LP2), S_{CDi} ($i \in 1, \ldots, 14$), is calculated as follows:

$$S_{CD_i} = \frac{1}{2}\langle 3\cos^2(\theta_i(t)) - 1\rangle \tag{12}$$

where $\theta_i(t)$ is the time series for the angle determined by the C_i–D bond vector and the bilayer normal vector. This angle was computed by assuming a tetrahedral geometry for each united-atom tail carbon.

For each iteration, a set of simulations were run for each unique temperature point. From these simulations, the objective function was computed using the training set properties, and parameters were modified as described in eq 8. The next iteration was then initialized using the updated parameter set. This process was repeated until the objective function was minimized.

Validation.

Model validation involves the evaluation of properties contained in a “test set” from trajectories generated using the new parameter set. The test set consists of properties not used during the training process. Extensibility of model accuracy beyond the properties used for model training is the true test of model quality.

Following completion of the optimization process, for each temperature point, continuous 600 ns long validation simulations were run using the final parameter set. These trajectories were first equilibrated for 20 ns, and had simulation conditions identical to those specified for the optimization simulations (with the exception of simulation length and parameter set). Property uncertainties were calculated using the autocorrelational statistical inefficiency of the property time series.[76]

In this study, our validation set consisted of experimental bilayer isothermal area compressibility modulus, k_A, lateral lipid diffusion, D_l, volume per lipid, V_L, and X-ray structure factor, $|F(q)|$. k_A is related to the fluctuations in A_L via the following expression:

$$\kappa_{A_{\mathrm{L}}} = 2\frac{k_{\mathrm{B}}T}{n_L}\frac{\langle A_{\mathrm{L}}\rangle}{\sigma_{A_{\mathrm{L}}}} \tag{13}$$

Wwhere k_B, T, n_L, and σ_{AL} represent Boltzmann’s constant, temperature, number of lipids in the simulation, and the variance in $\langle A_L\rangle$, respectively.

Lateral lipid diffusion was calculated by first pulling 50 snapshots from the NPT validation trajectories, spaced 20 ns apart. These frames were then used as the initial conditions for 100 ps long double precision NVE simulations. The 2D Einstein relation was then used to find the lateral diffusion constant from the slope of the linear regime of the mean square displacement time series[77] of the headgroup phosphate atom, as in eq 14, where the lateral displacement is along the bilayer–water (xy) interface, and the phosphate xy coordinates at time t are represented by $r(t)$. The reported lateral diffusion values are averages over the 50 NVE trajectory results.

$$D_l = \lim_{t \to \infty} \frac{1}{4t} \langle \mid r(t) - r(0) \mid \rangle^2 \tag{14}$$

Volume per lipid was calculated according to

$$V_{\mathrm{L}} = \frac{\langle V_{box} \rangle - n_w V_w}{n_L} \tag{15}$$

where $\langle V_{box} \rangle$, n_w, and V_w are the average simulation box volume, the number of water molecules, and the temperature-dependent volume of one water molecule, respectively. V_w was found from separate simulations of pure SPC water with simulations run with specifications identical to those applied to the validation simulations.

The electron density profiles were found using the GROMACS routine g_density. The bulk water electron densities were then subtracted from these EDPs. The Fourier transform of these curves yields the X-ray structure factors.

$$\mid F(q) \mid \; = \; \mid \int_{-z/2}^{z/2} \mathrm{e}^{iqz} (\rho(z) - \rho_{\mathrm{bulk}}(z)) \, \mathrm{d}z \mid \tag{16}$$

where $\rho(z)$ is the electron density profile, and ρ_{bulk} represents the electron density of bulk water.

RESULTS AND DISCUSSION

Training Set Results.

For this study, the training set consisted of two experimentally measured properties: average area per lipid, $\langle A_{\mathrm{L}} \rangle$, and phospholipid acyl chain deuterium order parameter, $|S_{\mathrm{CD}}|$. These properties belong to the “training set”, because they were used for model fitting. These particular characteristics were chosen for model training because they are structurally descriptive in distinct ways, and experimental data for each are available across a wide temperature range. Simulations run using the optimized parameters were found to closely reproduce training set experimental properties. As illustrated by Figures 2 and 3, our model shows an improvement from the original parameter set, across all temperatures.

Average area per lipid describes the expected surface area of each lipid molecule, projected onto the membrane xy plane. Of all bilayer properties, this has been studied the most extensively by experiment.[14,78-80] Consequently, it is also used to fit the majority of

phospholipid force fields. However, due to thermal motion of the bilayer, it is difficult to measure this quantity with high precision in real systems.[79] For this reason, the reported experimental values have relatively high uncertainty.

The original model, g53a6-b, results in a bilayer that is too laterally dense, as demonstrated by low A_L values. For the optimized parameter set, the A_L values are closer to experiment. For the most thoroughly characterized temperature value, T = 323.15 K, the new model falls within the error of the experimental value. Improvement of this property is important for the accurate structural description of the bilayer.

The deuterium order parameter quantifies the expected rotational disorder of each lipid tail methylene group with respect to the normal of the membrane xy plane.[81] As carbon "nodes" (acyl groups) are closer to the bilayer center, they become characteristically disordered, and $|S_{CD}|$ tends to zero. Reproducibility of acyl chain deuterium order parameters from simulation is highly valued because its experimental values are independent from analysis modeling, as they are calculated directly from ^{2}H NMR quadrupole splitting and have low uncertainty.

Across all temperatures, the original model displays overly ordered tail rotational dynamics, as evidenced by higher $|S_{CD}|$ values. The acyl groups nearest to the headgroup were found to have the highest error, with the error residual decreasing for the acyl groups near the bilayer center. For all temperatures, the updated model yields lower error when compared to experiment. It can be seen that the errors for the new model remain the greatest for the acyl groups nearer to the headgroup, but the magnitude is reduced.

Validation Set Results.

The validation set results are also an overall improvement to the original model. These properties were not used to fit the model, and therefore lend unbiased support to model quality.

As can be seen in Figure 4, the X-ray structure factor for g53a6-b has better agreement with experiment when T = 323.15 K, while the new model has better agreement to experiment for T = 333.15 K. At lower temperature, the new models have q-axis zeros shifted to slightly higher q values. This means that, on average, the new model produces a bilayer thinner than that of the original model. Typically, as A_L increases, bilayer xy density decreases, and the membrane will compress in the z direction. Considering the degree of increase for A_L, the structure factor is very modestly affected. At the higher temperature, the new model has very good agreement with experiment, showing improvement over the original model.

For the remaining validation properties, experimental data was only available for T = 323.15 K. These results are given in Table 1. Although the area per lipid is within experimental error for T = 323.15 K, volume per lipid shows an improvement, but it is not as drastic. This highlights how the bilayer is likely too narrow in the z direction at this temperature, and it stands as a direction for future improvement. Isothermal area compressibility has decreased for the new model and now lies closer to experiment. This shows that not only is the average A_L more accurate, but the fluctuation behavior of this behavior has also improved. Lastly,

the lateral diffusion constant has also improved. For the available experimental data point, the result is very close. This is an important result because no kinetic information was used to fit the new model. Optimization using a set of thermodynamic properties has also led to an improvement in this critical kinetic property.

Parameter Behavior.

At the heart of the ForceBalance method is the parameter-space minimization of the objective function. The objective function, χ^2, is considered minimized once $\Delta\chi^2$ begins to oscillate within a use-specified window (0.01 in this work, achieved after 10 iterations, see SI Figure S1). Because the objective function is computed by summing over the square of training set property error residuals, the parameter space derivatives of these terms can provide insight into the progression of the optimization procedure.

For this model, the water–lipid interaction has been optimized, while the lipid–lipid and water–water interactions remain constant. The was accomplished through the modification of the DPPC σ and ϵ components of the water–lipid Lennard-Jones interaction. SI Table S1 tabulates the original and final parameter values, as well as the percent difference, for each atom type (as illustrated in Figure 1). In Figure 5, the derivatives of the training property errors with respect to the modifiable parameters are illustrated for both the final and initial models. It can be seen that the original model shows significantly higher derivative values. For the new model, the derivatives are, in all cases, nearly zero. This signifies that the parameters of gb-fb15 exist in a local minimum. This is corroborated by the improved performance of the new model in both the training and validation sets.

It can be seen that the initial derivatives for each training property residual share many similarities. For several of the parameters, when the derivative is high for one training property, it is also high for the other training property. This is potentially a reflection of coupling between the training properties. It is expected that an increase in A_L would result in an increase in S_{CD}. If the lipids molecules are less compact, the lipid tails will have more space in which to rotate. There are some notable exceptions to this trend. The primary differences involve the phosphate moiety of the headgroup (LOS, LOM, and LP), where it can be seen that the initial derivatives do not follow an obvious trend. For example, the A_L sigma derivative for the sigma LOS parameter is much higher than the corresponding derivative for S_{CD}. The reverse is true for the epsilon LOM parameter. This moiety is near the bottom of the headgroup, and therefore likely important for intramolecular headgroup–tailgroup interaction, and also intermolecular headgroup–headgroup interactions. For this reason, we speculate that the coupling between the training properties is least strong for these atom types, and that is why this disparity is seen.

CONCLUSION

A united-atom model for the NPT molecular dynamics simulation of SPC-hydrated DPPC bilayers, named gb-fb15, has been presented. We recommend the use of this model in simulations of both homo- and heterogeneous DPPC bilayer systems. This model was constructed using the ForceBalance method, which uses the Levenberg–Marquardt algorithm to minimize a regularized nonlinear least-squares objective function. Our application of this

method to the DPPC Lennard-Jones parameters has yielded a model demonstrating clear improvements over the original parameter set across a range of thermodynamic and kinetic properties contained in an extensive training and validation set. Additionally, results show that this new model displays an improved response to temperature variation. The quality of our resultant model demonstrates the effectiveness of this optimization technique for the improvement of phospholipid bilayer force fields.

Future directions aim to improve model utility and accuracy, while reducing computational cost. In particular, we aim to apply this optimization procedure to a wider variety of lipid types. We also hope to ensure maximum compatibility of these models with the recent release of more accurate water and protein models. Updated parameter sets for a variety of united-atom lipid representations will be particularly useful for those interested in the simulation of large membrane proteins, for whom structural approximation is especially valuable.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

KM acknowledges support from the Stanford Center for Molecular Analysis and Design (CMAD). We acknowledge support from NSF-MCB-0954714 and NIH grant R01 GM062868.

REFERENCES

(1). Veatch SL; Keller SL Biochim. Biophys. Acta, Mol. Cell Res 2005, 1746, 172–185.
2. Feigenson GW Nat. Chem. Biol 2006, 2, 560–563. [PubMed: 17051225]
3. Feigenson GW Biochim. Biophys. Acta, Biomembr 2009, 1788, 47–52.
4. Korlach J; Schwille P; Webb WW; Feigenson GW Proc. Natl. Acad. Sci. U. S. A 1999, 96, 8461–8466. [PubMed: 10411897]
5. Shimshick EJ; McConnell HM Biochemistry 1973, 12, 2351–2360. [PubMed: 4351059]
6. Shimshick EJ; McConnell HM Biochem. Biophys. Res. Commun 1973, 53, 446–451. [PubMed: 4352068]
7. Heberle FA; Feigenson GW Cold Spring Harbor Perspect. Biol 2011, 3, a004630.
8. Kiessling V; Wan C; Tamm LK Biochim. Biophys. Acta, Biomembr 2009, 1788, 64–71.
9. Luzzati V; Husson FJ Cell Biol. 1962, 12, 207–219.
10. Steim JM; Tourtellotte ME; Reinert JC; McElhaney RN; Rader RL Proc. Natl. Acad. Sci. U. S. A 1969, 63, 104–109. [PubMed: 5257955]
11. Small DM J. Colloid Interface Sci 1977, 58, 581–602.
12. Small D In Handbook of Lipid Research; Small D, Ed.; Plenum Press: New York, NY, 1986.
13. Pascher I; Lundmark M; Nyholm P-G; Sundell S Biochim. Biophys. Acta, Rev. Biomembr 1992, 1113, 339–373.
14. Nagle JF; Tristram-Nagle S Biochim. Biophys. Acta, Rev. Biomembr 2000, 1469, 159–195.
15. Holthuis JC; Levine TP Nat. Rev. Mol. Cell Biol 2005, 6, 209–220. [PubMed: 15738987]
16. Wang X Curr. Opin. Plant Biol 2004, 7, 329–336. [PubMed: 15134755]
17. Chen C; Bazan NG Prostaglandins Other Lipid Mediators 2005, 77, 65–76. [PubMed: 16099392]
18. Hannun YA; Obeid LM Nat. Rev. Mol. Cell Biol 2008, 9, 139–150. [PubMed: 18216770]
19. Van Meer G; Voelker DR; Feigenson GW Nat. Rev. Mol. Cell Biol 2008, 9, 112–124. [PubMed: 18216768]

20. Phillips R; Ursell T; Wiggins P; Sens P Nature 2009, 459, 379–385. [PubMed: 19458714]
21. Groves JT; Kuriyan J Nat. Struct. Mol. Biol 2010, 17, 659–665. [PubMed: 20495561]
22. Simons K; Ikonen E Nature 1997, 387, 569–572. [PubMed: 9177342]
23. Simons K; Toomre D Nat. Rev. Mol. Cell Biol 2000, 1, 31–39. [PubMed: 11413487]
24. Sengupta P; Baird B; Holowka D Semin. Cell Dev. Biol 2007, 18, 583–590. [PubMed: 17764993]
25. Karplus M; McCammon JA Nat. Struct. Biol 2002, 9, 646–652. [PubMed: 12198485]
26. Dror RO; Jensen MØ; Borhani DW; Shaw DE J. Gen. Physiol 2010, 135, 555–562. [PubMed: 20513757]
27. Dror RO; Dirks RM; Grossman J; Xu H; Shaw DE Annu. Rev. Biophys 2012, 41, 429–452. [PubMed: 22577825]
28. Mackerell AD J. Comput. Chem 2004, 25, 1584–1604. [PubMed: 15264253]
29. Jambeck JPM; Lyubartsev AP J. Phys. Chem. B 2012, 116, 3164–3179. [PubMed: 22352995]
30. Dickson CJ; Madej BD; Skjevik ÅA; Betz RM; Teigen K; Gould IR; Walker RC J. Chem. Theory Comput 2014, 10, 865–879. [PubMed: 24803855]
31. Klauda JB; Venable RM; Freites JA; O'Connor JW; Tobias DJ; Mondragon-Ramirez C; Vorobyov I; MacKerell AD Jr; Pastor RWJ J. Phys. Chem. B 2010, 114, 7830–7843. [PubMed: 20496934]
32. Dickson CJ; Rosso L; Betz RM; Walker RC; Gould IR Soft Matter 2012, 8, 9617–9627.
33. Berger O; Edholm O; Jahnig F Biophys. J 1997, 72, 2002–2013. [PubMed: 9129804]
34. Chiu S-W; Pandit SA; Scott H; Jakobsson E J. Phys. Chem. B 2009, 113, 2748–2763. [PubMed: 19708111]
35. Poger D; Van Gunsteren WF; Mark AE J. Comput. Chem 2010, 31, 1117–1125. [PubMed: 19827145]
36. Kukol A J. Chem. Theory Comput 2009, 5, 615–626. [PubMed: 26610227]
37. Lee S; Tran A; Allsopp M; Lim JB; Hénin J; Klauda JB J. Phys. Chem. B 2014, 118, 547–556. [PubMed: 24341749]
38. Marrink SJ; Risselada HJ; Yefimov S; Tieleman DP; de Vries AH J. Phys. Chem. B 2007, 111, 7812–7824. [PubMed: 17569554]
39. Shinoda W; DeVane R; Klein ML J. Phys. Chem. B 2010, 114, 6836–6849. [PubMed: 20438090]
40. Bennun SV; Hoopes MI; Xing C; Faller R Chem. Phys. Lipids 2009, 159, 59–66. [PubMed: 19477311]
41. Chaimovich A; Shell MS J. Chem. Phys 2011, 134, 094112. [PubMed: 21384955]
42. Guenza M Eur. Phys. J.: Spec. Top 2015, 224, 2177–2191.
43. Piggot TJ; Piñeiro A; Khalid S J. Chem. Theory Comput 2012, 8, 4593–4609. [PubMed: 26605617]
44. Paloncýová M; Fabre G; DeVane RH; Trouillas P; Berka K; Otyepka M J. Chem. Theory Comput 2014, 10, 4143–4151. [PubMed: 26588554]
45. Cascella M; Vanni S Chem. Model 2015, 12, 1–52.
46. Ollila OS; Pabst G Biochim. Biophys. Acta, Biomembr 2016, 1858, 2512–2528.
47. Lyubartsev AP; Rabinovich AL Biochim. Biophys. Acta, Biomembr 2016, 1858, 2483–2497.
48. Wang J; Wolf RM; Caldwell JW; Kollman PA; Case DA J. Comput. Chem 2004, 25, 1157–1174. [PubMed: 15116359]
49. Wang J; Wang W; Kollman PA; Case DA J. Mol. Graphics Modell 2006, 25, 247–260.
50. Vanommeslaeghe K; Hatcher E; Acharya C; Kundu S; Zhong S; Shim J; Darian E; Guvench O; Lopes P; Vorobyov I; Mackerell AD J. Comput. Chem 2009, 31, 671–690.
51. Yu W; He X; Vanommeslaeghe K; MacKerell AD J. Comput. Chem 2012, 33, 2451–2468. [PubMed: 22821581]
52. Vanommeslaeghe K; MacKerell AD Jr J. Chem. Inf. Model 2012, 52, 3144–3154. [PubMed: 23146088]
53. Vanommeslaeghe K; Raman EP; MacKerell AD Jr J. Chem. Inf. Model 2012, 52, 3155–3168. [PubMed: 23145473]
54. Betz RM; Walker RC J. Comput. Chem 2015, 36, 79–87. [PubMed: 25413259]

55. Kramer-Fuhrmann O; Neisius J; Gehlen N; Reith D; Kirschner KN J. Chem. Inf. Model 2013, 53, 802–808. [PubMed: 23452048]
56. Hülsmann M; Köddermann T; Vrabec J; Reith D Comput. Phys. Commun 2010, 181, 499–513.
57. Wang L-P; Chen J; Van Voorhis T J. Chem. Theory Comput 2013, 9, 452–460. [PubMed: 26589047]
58. Wang L-P; Head-Gordon T; Ponder JW; Ren P; Chodera JD; Eastman PK; Martinez TJ; Pande VS J. Phys. Chem. B 2013, 117, 9956–9972. [PubMed: 23750713]
59. Berendsen HJC; Postma JPM; van Gunsteren WF; Hermans J In Intermolecular Forces; Pullman B, Ed.; Springer: Dordrecht, Netherlands, 1981; pp 331–342.
60. Oostenbrink C; Villa A; Mark AE; Gunsteren WFV J. Comput. Chem 2004, 25, 1656–1676. [PubMed: 15264259]
61. Villa A; Fan H; Wassenaar T; Mark AE J. Phys. Chem. B 2007, 111, 6015–6025. [PubMed: 17489626]
62. Jorgensen WL; Tirado-Rives J J. Am. Chem. Soc 1988, 110, 1657–1666. [PubMed: 27557051]
63. Ryckaert JP; Bellemans A Faraday Discuss. Chem. Soc 1978, 66, 95–106.
64. Lemkul J $KALP_{15}$ in DPPC. http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/membrane_protein/, Accessed: 2014-04-25.
65. Tieleman P University of Calgary Biocomputing Group, http://wcm.ucalgary.ca/tieleman/downloads, Accessed: 2014-04-25.
66. Berendsen HJ; van der Spoel D; van Drunen R Comput. Phys. Commun 1995, 91, 43–56.
67. Van Der Spoel D; Lindahl E; Hess B; Groenhof G; Mark AE; Berendsen HJ J. Comput. Chem 2005, 26, 1701–1718. [PubMed: 16211538]
68. Lindahl E; Hess B; Van Der Spoel D J. Mol. Model 2001, 7, 306–317.
69. Hess B; Kutzner C; Van Der Spoel D; Lindahl E J. Chem. Theory Comput 2008, 4, 435–447. [PubMed: 26620784]
70. Thain D Work Queue: A Scalable Master/Worker Framework. http://ccl.cse.nd.edu/software/workqueue/, Accessed: 2014-04-25.
71. Marquardt DW J. Soc. Ind. Appl. Math 1963, 11, 431–441.
72. Moré JJ In Numerical Analysis; Watson GA, Ed.; Springer: Berlin, Heidelberg, 1978; pp 105–116.
73. Dennis JE Jr; Gay DM; Walsh RE ACM Trans. Math. Softw 1981, 7, 348–368.
74. Moré JJ; Sorensen DC SIAM J. Sci. Comput 1983, 4, 553–572.
75. Nocedal J; Wright S In Numerical Optimization; Mikosch TV, Ed.; Springer: New York, NY, 2006; pp 245–269.
76. Chodera JD; Swope WC; Pitera JW; Seok C; Dill KA J. Chem. Theory Comput 2007, 3, 26–41. [PubMed: 26627148]
77. Nadler W; Tavan P; Schulten K Eur. Biophys. J 1985, 12, 25–31. [PubMed: 4006877]
78. Kučerka N; Nagle JF; Sachs JN; Feller SE; Pencer J; Jackson A; Katsaras J Biophys. J 2008, 95, 2356–2367. [PubMed: 18502796]
79. Petrache HI; Dodd SW; Brown MF Biophys. J 2000, 79, 3172–3192. [PubMed: 11106622]
80. Kučerka N; Nieh M-P; Katsaras J Biochim. Biophys. Acta, Biomembr 2011, 1808, 2761–2771.
81. Feller SE; Venable RM; Pastor RW Langmuir 1997, 13, 6555–6561.
82. Kučerka N; Katsaras J; Nagle JF J. Membr. Biol 2010, 235, 43–50. [PubMed: 20407764]
83. Vaz WL; Clegg RM; Hallmann D Biochemistry 1985, 24, 781–786. [PubMed: 3994985]

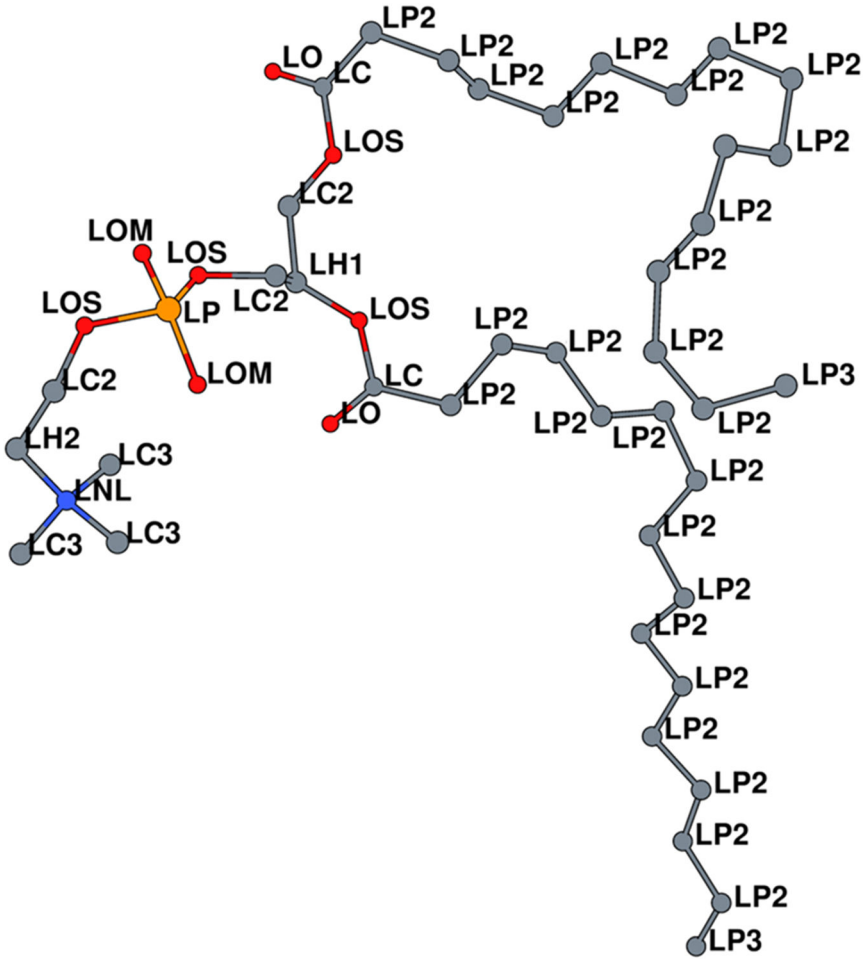


Figure 1.
Schematic of a united-atom (UA) DPPC lipid, with labeled atom types. This conformation was sampled from a simulation trajectory.

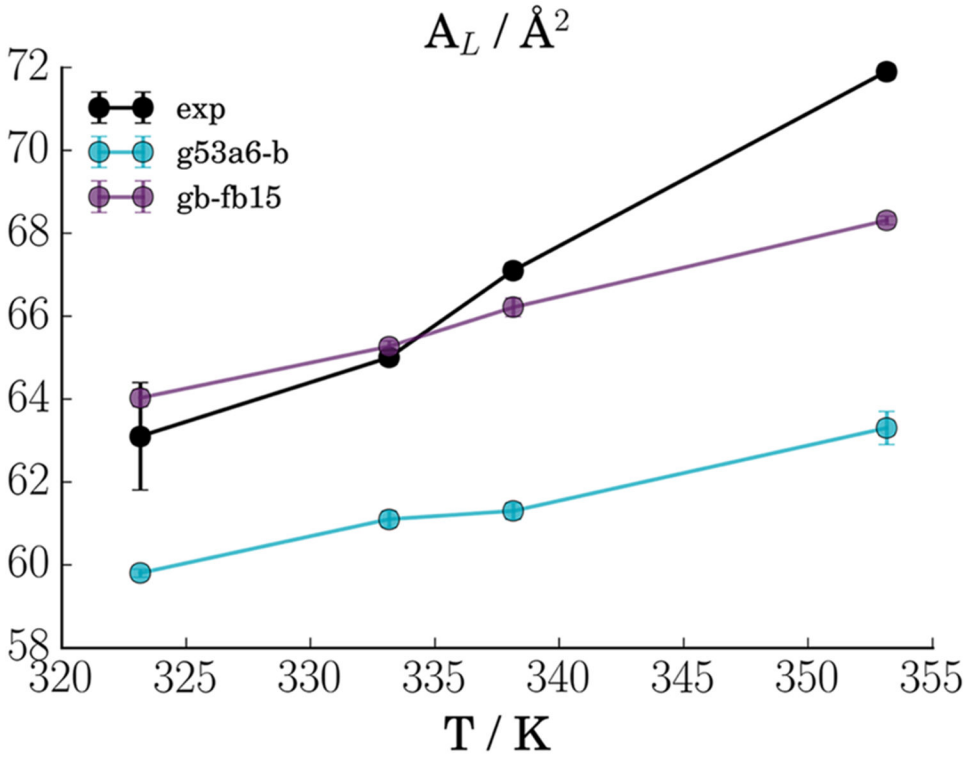


Figure 2.
DPPC area per lipid as a function of temperature compared to experiment[78-80] for the original and optimized models.

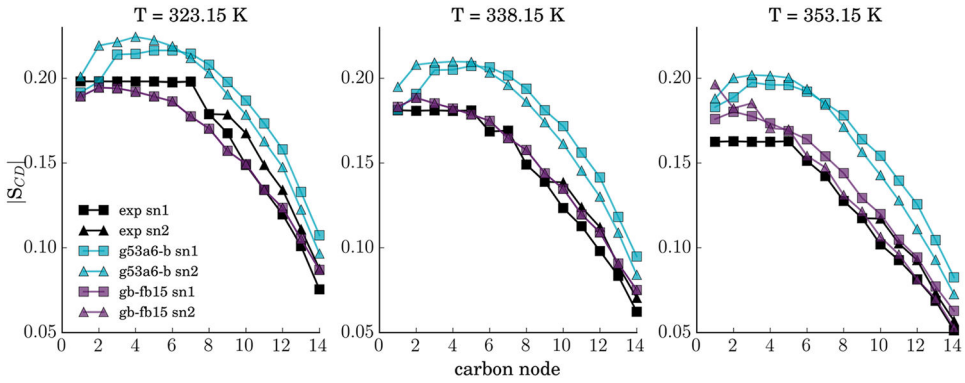


Figure 3.

DPPC deuterium order parameter as a function of temperature compared to experiment[79] for the original and optimized models.

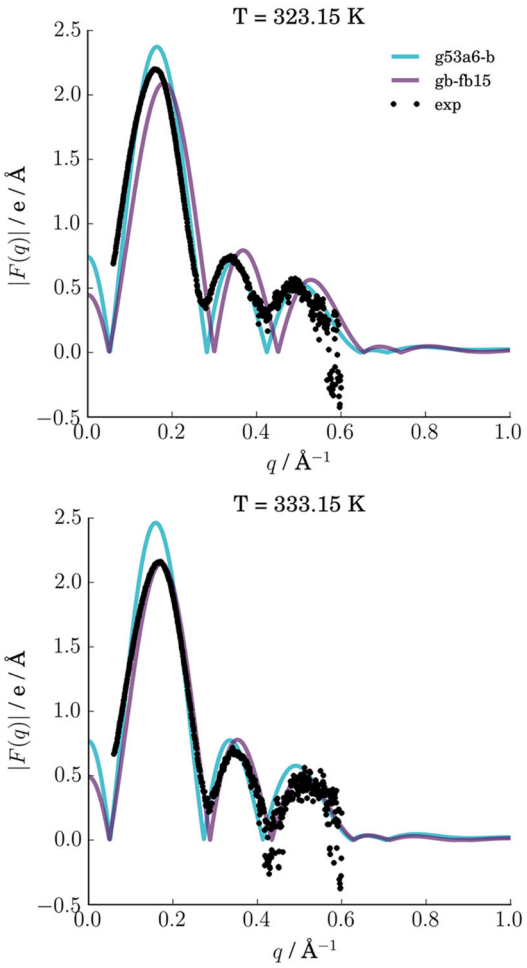


Figure 4.
DPPC X-ray structure factor at the temperatures for which there exists experimental data.[78,82]

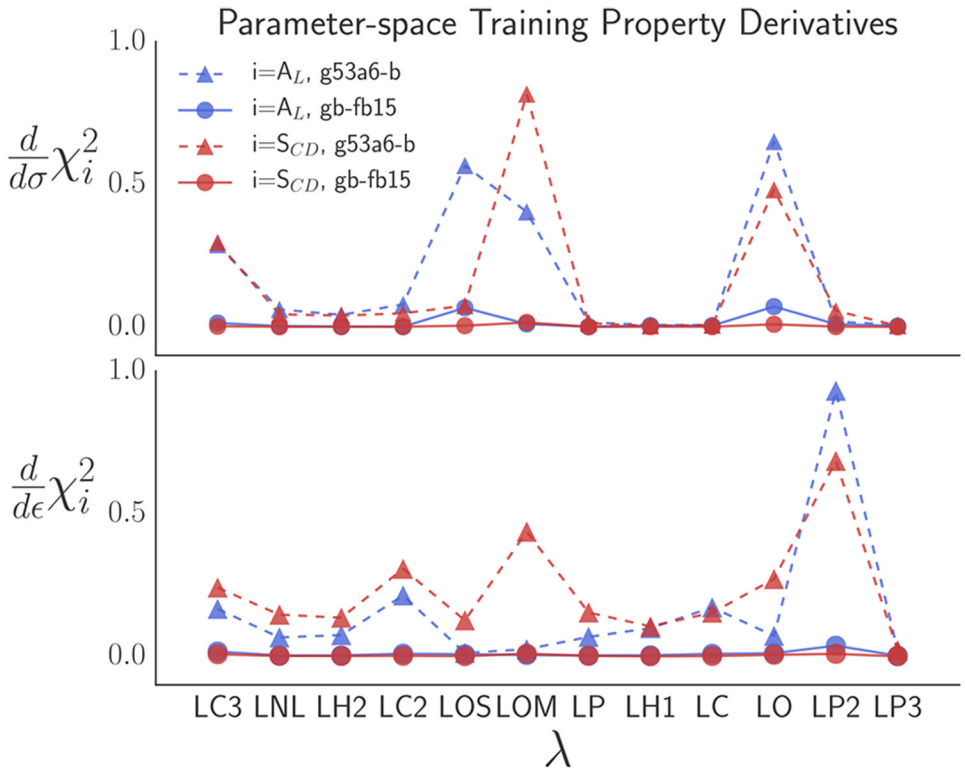


Figure 5.

Magnitude of the initial (dashed) and final (solid) parameter derivatives for each objective function training set residual. The derivatives have been normalized by the L_2 norm of the starting model derivative values.

Table 1.

Comparison of Validation Set Results to Experiment

Property	T (K)	g53a6-b	gb-fb15	exp
V_L (Å^3)	323.15	1171.7 (1)	1191.7 (3)	1232[14]
	333.15	1179.6 (1)	1199.0 (3)	
	338.15	1183.7 (1)	1203.4 (2)	
	353.15	1194.5 (1)	1214.3 (1)	
κ_{AL} (mN m^{-1})	323.15	290 (9)	190 (14)	231[14]
	333.15	294 (13)	172 (3)	
	338.15	299 (3)	170 (3)	
	353.15	315 (2)	218 (8)	
D_l (10^{-8} cm^2 s^{-1})	323.15	8.8 (5)	12.8 (5)	12.5[83]
	333.15	10.6 (8)	14.4 (5)	
	338.15	11.9 (7)	15.1 (5)	
	353.15	16.5 (8)	19.8 (8)	