PCCP

PAPER

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d0cp05041e

The automated optimisation of a coarse-grained force field using free energy data†

Javier Caceres-Delpiano,[a] Lee-Ping Wang*[b] and Jonathan W. Essex*[a]

Received 23rd September 2020,
Accepted 18th October 2021

DOI: 10.1039/d0cp05041e

rsc.li/pccp

Atomistic models provide a detailed representation of molecular systems, but are sometimes inadequate for simulations of large systems over long timescales. Coarse-grained models enable accelerated simulations by reducing the number of degrees of freedom, at the cost of reduced accuracy. New optimisation processes to parameterise these models could improve their quality and range of applicability. We present an automated approach for the optimisation of coarse-grained force fields, by reproducing free energy data derived from atomistic molecular simulations. To illustrate the approach, we implemented hydration free energy gradients as a new target for force field optimisation in ForceBalance and applied it successfully to optimise the un-charged side-chains and the protein backbone in the SIRAH protein coarse-grain force field. The optimised parameters closely reproduced hydration free energies of atomistic models and gave improved agreement with experiment.

Open Access Article. Published on 19 October 2021. Downloaded on 11/9/2021 2:18:16 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Introduction

Computational tools have become very important in revealing the driving forces in bio-molecular processes; in particular, molecular dynamics (MD) simulations provide a physically motivated picture based on Newton's equations of motion coupled with empirical model potentials (force fields).[1] Classical atomistic (AT) models provide a detailed representation of the system, but are inadequate for simulations of very large systems over long timescales.[2] Coarse-grained (CG) models currently represent one of the most important approximations for the construction and simulation of larger systems.[3,4] By subsuming groups of atoms into single interaction sites, much faster calculations can be realised. However, a disadvantage of CG models is the loss of accuracy associated with reducing the number of interacting particles. Moreover, coarse-graining typically smooths the energy landscape compared to classical atomistic models, diminishing the energy barriers between different states and reducing trapping in energy minima.[5] This can greatly affect calculated dynamic properties such as the rates of conformational change. Despite these drawbacks, CG models have become a widely used approximation, allowing us to extend spatial and temporal scales for the simulation of bigger and more complex systems. Given this, new approaches for the optimisation of CG models are highly desirable.

The accuracy of a force field depends in part on the empirical parameters in the model, which are usually determined by fitting simulation results to a training data set (*i.e.* the *targets*). For example, these targets can come from supermolecule calculations such as QM simulations or experimental information, but often such data are not available for the system of interest. These difficulties, in conjunction with its' iterative nature and complexity, mean that force field optimisation is something of a black art.[6] Different frameworks and approximations to optimise parameters have been proposed: (1) *ad hoc* methods where parameters are iteratively adjusted until a specific property can be reproduced or stable simulations achieved,[7–9] (2) machine learning methods that have been used in tandem with QM calculations,[10,11] and (3) force or energy matching to reproduce QM calculations or other simulation data.[6,12,13]

ForceBalance[14,15] is an automated parameter optimisation method and software package that enables reproducible development of force field parameters. It has been used for the optimisation of different types of force fields, such as a series of water models (iAMOEBA,[16] AMOEBA14,[17] TIP3P-FB, TIP4P-FB[15] and uAMOEBA[18]), a united-atom phospholipid bilayer model (gb-fb15),[19] and an all-atom protein force field (AMBER-FB15).[20] ForceBalance is able to incorporate multiple sources of experimental or simulated reference data. The objective function to be minimised in parameter space is a weighted sum of squared differences between the reference and calculated properties, with a regularisation term added that penalises large parameter deviations from their initial

[a] School of Chemistry, University of Southampton, Southapton, S017 1BJ, UK. E-mail: j.w.essex@soton.ac.uk
[b] Department of Chemistry, University of California, Davis, California 95616, USA. E-mail: leeping@ucdavis.edu

† Electronic supplementary information (ESI) available: Discussion for the choice of α values, parameter dependence and the methionine case are summarised. A full list of the new parameters is also included. See DOI: 10.1039/d0cp05041e

Open Access Article. Published on 19 October 2021. Downloaded on 11/9/2021 2:18:16 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

values to prevent overfitting. A harmonic penalty function, which corresponds to a Gaussian prior distribution, is usually used. ForceBalance uses a trust-radius Newton-Raphson optimiser that can efficiently optimise the objective function to within the statistical noise of the simulation after 5–10 iterations; other gradient-based and stochastic optimisation procedures may also be used in a modular fashion (*e.g.* L-BFGS, Simplex and Powell algorithms). The physical force field parameters are mapped to abstract optimisation variables of order one to improve the conditioning of the optimisation problem – this also enables one to adjust the regularisation strengths applied to different parameter types. The molecular mechanics property calculations are automated by interfaces to classical molecular dynamics software packages (*engines*) such as GROMACS,[21] TINKER[22] and OpenMM.[23] Properties previously used in ForceBalance range from energies, atomistic forces, and vibrational modes from *ab initio* calculations,[20] *ab initio* gas phase properties such as cluster interaction energies, temperature and pressure dependent bulk phase properties of liquids such as density, enthalpy of vaporisation, dielectric constant, thermal expansion coefficient, isothermal compressibility and isobaric heat capacity,[15,17] and lipid membrane properties such as area per lipid and deuterium order parameters.[19]

Hydration free energies (HFEs) are an important property for aqueous systems such as proteins. They help us to understand biological processes such as ligand recognition, protein–protein interactions, folding and conformational change. Moreover, hydration free energies have been used for the validation of molecular force fields, and they are an integral part of the calculation and estimation of solubilities, partition coefficients and solute–solvent interactions.[24–27] For these reasons, use of solvation free energies as a parameterisation target for coarse-grained models may improve their performance. Moreover, it has been recently stated that there is considerable interest in methods that can automatically generate a coarse-grained model which is representative in terms of local structure and free energy changes.[28]

Here we present a general approach to optimise coarse-grain force fields by reproducing free energy gradients derived from atomistic simulations. We exemplify the method by optimising the SIRAH CG protein force field using atomistic hydration free energy (HFE) data in the ForceBalance software. The gradient of the hydration free energy is optimised to match the result from an AT simulation, with the goal of improving the CG solvation free energies as a consequence. The approach of fitting atomistic HFE gradients has the advantage of reducing the computational cost of the parameter optimisation because it does not require full HFE calculations of the CG model at every optimisation step. The parameters of charged and uncharged amino acids were both optimized, but we rejected the charged amino acid parameters because they failed validation tests. A full HFE calculation is carried out after CG model optimisation to validate the approach by comparison to atomistic and experimental HFEs. The newly optimised SIRAH-OBAFE (Optimised Based on Atomistic Free Energies) force field, is briefly evaluated in terms of conventional MD simulations of proteins in solution. To facilitate the development of the new force field we have also optimised the WT4 water model in SIRAH using experimental properties such as density, enthalpy of vaporisation and dielectric constant.

Methods

Optimisation based on free energy gradients: overview

To calculate the free energy difference between two states, X and Y, it is useful to include a coupling parameter to connect both states.[29–32] This coupling parameter, α, changes from 0 to 1, and can be expressed as a linear function of the potential energy $U(\mathbf{r}^{\mathrm{N}};\alpha)$ by:

$$U\left(\mathbf{r}^{\mathrm{N}};\alpha\right)=\alpha U_0\left(\mathbf{r}^{\mathrm{N}}\right)+U_1\left(\mathbf{r}^{\mathrm{N}}\right)(1-\alpha) \qquad (1)$$

where $\mathbf{r}^{\mathrm{N}}$ corresponds to the system coordinates of N particles, $U_0(\mathbf{r}^{\mathrm{N}})$ corresponds to the potential energy of a "reference system" and $U_1(\mathbf{r}^{\mathrm{N}})$ corresponds to the potential energy of a system of interest. α connects the two states through a physical or non-physical pathway. Based on thermodynamic integration theory,[29] one can express the difference in free energy between two states by:

$$\Delta F=\int_0^1\left\langle\frac{\partial U\left(\mathbf{r}^{\mathrm{N}};\alpha\right)}{\partial\alpha}\right\rangle_{\alpha}\mathrm{d}\alpha=\int_0^1\langle\Delta U\rangle_{\alpha}\mathrm{d}\alpha \qquad (2)$$

where the change in the free energy ΔF, between a reference state and a target state, can be computed from the integral between values of 0 (un-perturbed) and 1 (perturbed) of the ensemble average of the derivative of the potential energy with respect to the coupling parameter α. In the case of the linear coupling of $U(\mathbf{r}^{\mathrm{N}};\ \alpha)$, corresponding to eqn (1), this is equivalent to the ensemble average of ΔU as a function of α, where ΔU is the internal energy change between the $\alpha = 0$ and $\alpha = 1$ states.

We have implemented a new mathematical expression for the optimisation of coarse-grained force field parameters based on free energy gradients from atomistic simulations. Starting with a set of simulations that evaluate $\langle\Delta U\rangle_\alpha$ for AT systems at selected values of α, we fit these values in our CG simulations by optimising the CG parameters, which indirectly improves the hydration free energies. The objective function that is minimized may be written as:

$$L(k)=\sum_{m=1}^{M}L_m(\boldsymbol{k})+w_{\mathrm{reg}}|\boldsymbol{k}|^2 \qquad (3)$$

Here $L_m(\boldsymbol{k})$, called the target terms, are the contributions of each molecule to the objective function; in this work the parameters for each molecule are optimized separately, thus there is only one term in the sum. $L_m(\boldsymbol{k})$ is given by a weighted sum of squared differences between the AT and CG free energy gradients:

$$L_m(k)=\frac{1}{d^2}\sum_{i=1}^{N_{\mathrm{a}}}\left|\langle\Delta U\rangle_{\alpha,\mathrm{CG}}(\boldsymbol{k})-\langle\Delta U\rangle_{\alpha,\mathrm{AT}}\right|^2 \qquad (4)$$

where $\boldsymbol{k}$ is the vector of dimensionless "mathematical" parameters being directly manipulated by the optimization

Open Access Article. Published on 19 October 2021. Downloaded on 11/9/2021 2:18:16 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

algorithm, $L(\boldsymbol{k})$ is the overall objective function, $L_m(\boldsymbol{k})$ is the contribution from molecule m, and w_{reg} is a regularization term, here set to 0.01 to ensure that large excursions in the parameters are properly penalized without being overly restrictive. The $\boldsymbol{k}$-vector is related to the physical force field parameters in the simulation $\boldsymbol{K}$ by a shifting and scaling as:

$$K_i = K_i(0) + t_i k_i \tag{5}$$

where K_i and $K_i(0)$ are the current and initial values of the force field parameter, and t_i is a scaling factor, also called the prior width, that carries the same dimension as K_i and represents the expected variation of the force field parameters over the course of the optimization.

To optimize the objective function efficiently, the first derivatives of the simulated quantities with respect to force field parameters are needed. The analytical derivative of $\langle\Delta U\rangle_\alpha$ with respect to the force field parameters can be obtained as:

$$\frac{\partial\langle\Delta U_\alpha\rangle}{\partial\lambda} = \left\langle\frac{\partial\Delta U}{\partial\lambda}\right\rangle_\alpha - \beta\left(\left\langle\Delta U\frac{\partial E_\alpha}{\partial\lambda}\right\rangle_\alpha - \langle\Delta U\rangle_\alpha\left\langle\frac{\partial E_\alpha}{\partial\lambda}\right\rangle_\alpha\right) \tag{6}$$

where λ corresponds to the force field parameter, $\langle\Delta U\rangle_\alpha$ is the ensemble average of the energy difference between $\alpha = 0.0$ and $\alpha = 1.0$, simulated at a defined α value, ΔU corresponds to the instantaneous energy difference for each snapshot between $\alpha = 0.0$ and $\alpha = 1.0$, and E_α is the potential energy of the system at α. Rather than optimising the free energies directly, we optimise against the ensemble average of the free energy gradients at specific α values, $\langle\Delta U\rangle_\alpha$. The derivative of the free energy gradients, $\langle\Delta U\rangle_\alpha$, with respect to the force field parameters λ is composed of ensemble averages of instantaneous ΔU values, and derivatives of ΔU and the potential energy with respect to the FF parameters, at each α point used, where both derivatives are obtained numerically by finite difference using snapshots from the corresponding trajectories.

Optimisation of a CG protein force field: uncharged side-chains and backbone

A workflow showing the steps followed in this work, and separated into four main stages, is presented in Fig. 1. Briefly, hydration free energies for atomistic systems are calculated by decoupling both van der Waals and charge parameters. Then, atomistic free energy gradients are collected as an average of ΔU values, at simulations with different α values, $\langle\Delta U\rangle_\alpha$. These data are used to optimise each specific CG side-chain (or the backbone) with its corresponding $\langle\Delta U\rangle_\alpha$ value. Then, parameters corresponding to the smallest objective function are collected. These parameters are then used to re-calculate new hydration free energies of the CG side-chains. Full details of the simulation setup and parameters for each of the stages shown in Fig. 1 are provided in the ESI.†

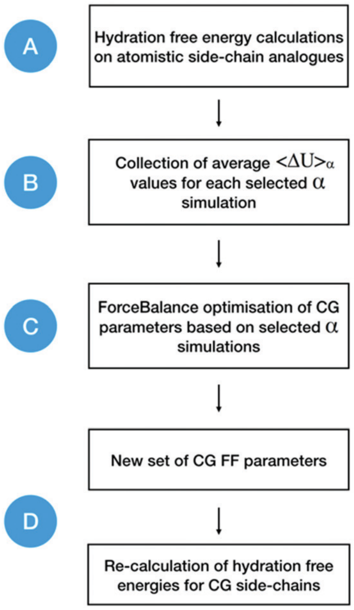
A
Hydration free energy calculations on atomistic side-chain analogues
B
Collection of average <ΔU>α values for each selected α simulation
C
ForceBalance optimisation of CG parameters based on selected α simulations
New set of CG FF parameters
D
Re-calculation of hydration free energies for CG side-chains

Fig. 1 General workflow for the CG force field optimisation. Free energy gradients are collected from atomistic simulations and used as optimisation targets in ForceBalance. New parameters are obtained and later used in the calculation of hydration free energies for CG beads (side-chains and backbone). Letters from A to D correspond to each of the main stages in the optimisation and validation process (see ESI†).

Hydration free energies of charged side-chains

The calculation of hydration free energies for charged systems is a more complex process compared to the classical use for uncharged systems. The standard raw hydration free energy ($\Delta G^{\ominus}_{\text{hyd}}$) for an ion is calculated as the sum of three processes: charging (ΔG_{chg}), cavitation (ΔG_{cav}) and a standard convention term ($\Delta G^{\ominus}_{\text{std}}$, which is equal to 7.95 kJ mol^{-1}, considering a water density of 997 kg m^{-3} at a pressure of 1 atm[33]), as:

$$\Delta G^{\ominus}_{\text{hyd}} = \Delta G_{\text{chg}} + \Delta G_{\text{cav}} + \Delta G^{\ominus}_{\text{std}} \tag{7}$$

The cavitation term corresponds to the creation of a molecule in solution through the scaling of intermolecular Lennard-Jones interactions, coupled to a parameter α.

The calculation of raw charging free energies ($\Delta G^{\text{raw}}_{\text{chg}}$) is especially sensitive to the chosen simulation methodology[33–35] where different corrections have been introduced to alleviate these effects (see ESI†). Following these corrections,[33–35] raw hydration free energies and these corrections (ΔG_{cor}, see ESI†) were used to calculate the methodology-independent free energy values for the charged side-chains, as:

$$\Delta G_{\text{chg}} = \Delta G^{\text{raw}}_{\text{chg}} + \Delta G_{\text{cor}} \tag{8}$$

These corrections, and their application, have been demonstrated before for monoatomic[34,35] and polyatomic ions.[33] They are usually named as type A, B, C and D corrections, which are related to approximations in the electrostatic interactions (A), approximations of the system size (finite) (B), deviations of the solvent generated electrostatic potential given the choice of an inappropriate summation scheme (C), and a wrong estimation of the dielectric constant for solvent model used (D), respectively. In the case of polyatomic ions (such as the charged side-chains used in this work), numerical solutions of the Poisson equation are needed to obtain an estimation of the charging free energy in an idealised system that obeys a macroscopic regime

Open Access Article. Published on 19 October 2021. Downloaded on 11/9/2021 2:18:16 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

(non-periodic with Coulombic electrostatic interactions) and based on the experimental solvent permittivity ($\Delta G_{\text{chg}}^{\text{NPBC}}$). Simulations of a periodic systems with a specific electrostatic scheme and based on the model solvent permittivity are also needed ($\Delta G_{\text{chg}}^{\text{PBC,LS}}$ for a periodic boundary condition system using a Lattice-Summation scheme). These two terms can be used for the calculation of A + B + D corrections from continuum electrostatic calculations.

A type C1 correction is required for lattice-summation (LS) and Barker-Watts reaction field (BM) schemes, and corrects the P-summation (atom-based cut-off) implied by these schemes to a proper M-summation (molecule-based cut-off). This correction is calculated analytically.

Finally, and summarising all the necessary methodology-dependent corrections, standard hydration free energies were calculated as:

$$\Delta G_{\text{hyd}}^{\ominus} = (\Delta G_{\text{chg}}^{\text{raw}} + \Delta G_{\text{cav}}) + \Delta G_{\text{A+B+D}} + \Delta G_{\text{C}_1} + \Delta G_{\text{std}}^{\ominus} \qquad (9)$$

Further information regarding the corrections used in the calculation of free energies of hydration of charged molecules is available in the ESI.†

Optimisation of charged side-chains

Optimisations for the charged side-chains were performed in a similar fashion as for uncharged side-chains and the protein backbone, where free energy gradients from atomistic simulation were used as optimisation target for the CG parameters (see eqn (3), and ESI† for simulation details). As this type of free energy calculation is methodology dependent, the inclusion of corrections is required. Assuming that the final free energies between the atomistic and coarse-grained models must be equal, the sum of their free energy gradients and the necessary correction gradients must be equal as well. Since the corrections are added ex post, the fitting data used is given as,

$$\frac{\partial\left\langle \Delta G_{\text{chg}}^{\text{raw}} \right\rangle_{\alpha,\text{AT}}}{\partial\lambda} + \frac{\partial\langle \Delta G_{\text{cor}}\rangle_{\alpha,\text{AT}}}{\partial\lambda} = \frac{\partial\left\langle \Delta G_{\text{chg}}^{\text{raw}} \right\rangle_{\alpha,\text{CG}}}{\partial\lambda} + \frac{\partial\langle \Delta G_{\text{cor}}\rangle_{\alpha,\text{CG}}}{\partial\lambda} \qquad (10)$$

and moving the property that we want to optimise to one side,

$$\frac{\partial\left\langle \Delta G_{\text{chg}}^{\text{raw}} \right\rangle_{\alpha,\text{CG}}}{\partial\lambda} = \frac{\partial\left\langle \Delta G_{\text{chg}}^{\text{raw}} \right\rangle_{\alpha,\text{AT}}}{\partial\lambda} + \frac{\partial\langle \Delta G_{\text{cor}}\rangle_{\alpha,\text{AT}}}{\partial\lambda} - \frac{\partial\langle \Delta G_{\text{cor}}\rangle_{\alpha,\text{CG}}}{\partial\lambda} \qquad (11)$$

where $\partial\langle\Delta G_{\text{chg}}^{\text{raw}}\rangle_{\alpha,\text{CG}}/\partial\lambda$ and $\partial\langle\Delta G_{\text{chg}}^{\text{raw}}\rangle_{\alpha,\text{AT}}/\partial\lambda$ correspond to the derivative of the raw charging hydration free energy gradients with respect to the force field parameters (at a specific α value), for a coarse-grained and atomistic system, respectively. $\partial\langle\Delta G_{\text{cor}}\rangle_{\alpha,\text{CG}}/\partial\lambda$ and $\partial\langle\Delta G_{\text{cor}}\rangle_{\alpha,\text{AT}}/\partial\lambda$ are the derivatives of the free energy corrections with respect to the force field parameters (at a specific α value), for a coarse-grained and atomistic system, respectively. The derivatives of the corrections were calculated using finite differences, based on a set of α values between 0.4 and 1.0, where the parameters were scaled accordingly (*i.e.* for $\alpha = 0.9$, parameters were scaled to a 90% of their original value). See ESI† for more details on the calculation of these corrections and the protocol used in optimisation runs. Full details of the simulation setup and parameters used for charged molecules are provided in the ESI.†

The SIRAH model

Our new parameterisation approach has been applied to the optimisation of the SIRAH force field,[36] a CG force field and a promising alternative to conventional atomistic protein force fields. Unlike MARTINI,[37] SIRAH does not use elastic networks to overcome the problem of secondary structure stability. The use of a higher resolution backbone representation produces hydrogen bond-like interactions, which stabilise the secondary structure. Moreover, SIRAH models long-range electrostatic interactions using the particle mesh Ewald method (PME) and a dielectric constant of unity. At the moment, the SIRAH force field contains parameters for DNA,[38] water,[39] proteins[36] and DMPC lipid,[40] and it has been used in the simulation of protein–DNA interactions,[41] hybrid AT/CG simulations[42] and in the implementation of a supra-CG water model for the simulation of virus-like particles.[43]

The SIRAH CG protein model uses a higher resolution backbone compared to previous CG models, where positions for nitrogen, α-carbon and oxygen are maintained. Each bead possesses its own partial charge, which helps to stabilise secondary structures through the formation of hydrogen bond-like interactions. Dihedral angles define the secondary structure for the system, forcing the existence of the two main conformations, α-helices and β-strands. Side-chains are modelled using one to five pseudo-atoms and partial charges are placed based on the number of hydrogen-bond acceptors and/or donors. van der Waals parameters were set using an *ad hoc* procedure, and van der Waals interactions are calculated based on the Lorentz–Berthelot combining rules, with the addition of some corrections.[36] The SIRAH water model (WT4) is represented by four linked beads in a tetrahedral geometry, each with a specific partial charge. Each CG water molecule represents approximately 11 atomistic water molecules based on the mass of CG beads (50 a.u.).[39] A new, updated version of the SIRAH protein model was recently released, named as SIRAH 2.0, where corrections were made to bonded and non-bonded interactions of amino-acids, showing decreased RMSD values up-to 0.1 nm, for different protein systems, compared to the previous SIRAH 1.0 version.[44]

Optimisation of the WT4 model

For the WT4 model optimisation, three condensed-phase properties were optimised: density, enthalpy of vaporisation and dielectric constant. Experimental values (taken from ref. 15) for these properties were used as targets, at 298.15 K and 1 atm. The trust-radius Newton-Raphson algorithm was used to minimise the objective function (see ESI† for more details). For this work, the optimisation was regularised using a Gaussian prior that is centred on the original SIRAH parameter. This is done to prevent the optimisation from changing the parameters too much and to avoid over-fitting, adding a penalty

Open Access Article. Published on 19 October 2021. Downloaded on 11/9/2021 2:18:16 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

that is applied to the objective function. Conceptually speaking, addition of a penalty function is equivalent to imposing a prior probability distribution on the parameters. Only non-bonded parameters were optimised, including van der Waals Sigma (σ) and epsilon (ε) values, and partial charges.

100 optimisation cycles were run, with the following simulation protocol: the system was minimised for 5000 steps using a steepest descent algorithm followed by an NPT equilibration time of 5 ns. Production runs were performed for 15 ns. A leap-frog algorithm was used to integrate Newton's equations of motion with a time-step of 20 fs. Electrostatic interactions are calculated using the Particle Mesh Ewald method[45] with a direct cut-off of 1.2 nm and a grid spacing of 0.2 nm. A 1.2 nm cut-off was used for van der Waals interactions. The v-rescale thermostat[46] and the Parrinello–Rahman barostat[47] were used to maintain the temperature at 298.15 K and the pressure at 1 atm, respectively. The simulation protocol was based on the original publication of the SIRAH 1.0 protein force field.[36] All simulations were run with GROMACS v. 2018.2.[48] Statistical fluctuations in the thermodynamic properties dominated the objective function after 30 iterations, and the set of parameters with the lowest objective function was chosen as the best solution. Single point calculations were run three times, with the best parameter set, in order to estimate standard errors.

Protein simulations

To briefly evaluate the optimised force field, a series of proteins with sizes ranging from 585 to 69 residues were simulated (most were proteins tested in the original SIRAH 1.0 publication[36]).

Coarse-grained molecular dynamics simulations were performed using the SIRAH 1.0/WT4, SIRAH 2.0/WT4 and the ForceBalance reparameterised SIRAH-OBAFE/WT4-FB force fields, for all the previously mentioned protein systems. Energy minimisation was carried out for 10 000 iterations of the steepest descent algorithm. This was followed by an *NPT* equilibration dynamics procedure of 20 ns with positional restraints of 1000 kJ mol^{-1} nm^{-2} applied to all the protein beads. Production runs were performed for 3 μs for each system with an integration time-step of 20 fs. Electrostatic interactions were calculated using the Particle Mesh Ewald procedure[45] with a direct cut-off of 1.2 nm and a grid spacing of 0.2 nm. Non-bonded interactions were modelled using the Lennard-Jones potential with a cut-off of 1.2 nm. All simulations were run at 1 bar with the Parrinello–Rahman barostat[47] and at 298.15 K with the v-rescale thermostat.[46] Systems were neutralised by adding Na$^+$ and Cl$^-$ ions up to a concentration of 150 mM. Root mean square fluctuations (RMSF) and root mean square deviations (RMSD) time series were calculated with GROMACS v.2018.2.[48]

Results and discussion

One of the main points that encouraged the development and improvement of these CG models, and also an important limitation of the SIRAH force field, is the inaccuracy of the hydration free energies of amino acid side-chains, which could limit its predictive power in protein simulations. Calculations of SIRAH 1.0 decoupling hydration free energies yield completely different results compared to all-atom OPLS-AA results, with mean unsigned errors against experiment (MUE) of 5.03 kcal mol^{-1} *vs.* 1.04 kcal mol^{-1}, for SIRAH 1.0 and all-atom systems, respectively, calculated against experimental values (see below).

Table 1 Comparison of WT4 and WT4-FB models against experimental water properties at 298 K and 1 atm[a]

Property[a]	Expt.	WT4	WT4-FB (this work)[b]
ρ (kg m^{-3})	997.045	996.6 ± 0.3	995.4 ± 1.5
ΔH_{vap} (kJ mol^{-1})	43.989	39.8 ± 0.2	43.7 ± 0.2
ε_r	78.409	123.7 ± 14.2	74.2 ± 12.3
α (10^{-4} K^{-1})	2.572	11.6 ± 2.4	11.8 ± 2.7

[a] The calculated properties correspond to density (ρ), enthalpy of vaporisation (ΔH_{vap}), dielectric constant (ε_r) and expansion coefficient (α), and the experimental data were obtained from ref. 15. [b] Full set of parameters for the WT4-FB model provided in Table S2 (ESI). Error are reported as standard errors based on 3 simulations.

Optimisation of the WT4 water model

We start our ForceBalance calculation with the optimisation of the WT4 water model, where only non-bonded parameters were optimised (charges, Sigma and epsilon values). Three condensed-phase properties for liquid water were used as reference data: density, enthalpy of vaporisation and dielectric constant at 298 K and 1 atm. The original WT4 model is able to reproduce experimental thermodynamic properties such as the water density at 298 K, but it is less satisfactory in the prediction of other properties (*i.e.* dielectric constant, expansion coefficient, surface tension, *etc.*).[39] In contrast, the new WT4 model (now called WT4-FB) overcomes the previous issue with the dielectric constant in the original model by accurately reproducing experimental values for the three properties together (Table 1). Calculations of the thermal expansion coefficient yield similar results to those of the original model (11.8 × 10^{-4} K^{-1} *vs.* 11.6 × 10^{-4} K^{-1}).[39] Thus, optimising WT4 with ForceBalance does not necessarily improve all properties; the level of accuracy obtainable is dependent on the granularity of the CG representation and the choice of force field functional form.

Optimisation of the SIRAH protein force field: uncharged side-chains and backbone

Our new approach for CG FF optimisation is based on using derivatives of the free energy gradients *i.e.* $\langle \Delta U \rangle_\alpha$, at different values of the coupling parameter α, with respect to the force field parameters. We choose to work with free energy gradients due to their linear relationship with the easily computed "vertical energy gap", $\langle \Delta U \rangle_\alpha$. In practice, the thermally averaged CG $\langle \Delta U \rangle_\alpha$, is fitted to atomistic $\langle \Delta U \rangle_\alpha$, where one or more selected values of the coupling parameter α are used to carry out the simulations. HFEs were computed separately from the optimisation process. 10 sets of CG parameters were optimised

Open Access Article. Published on 19 October 2021. Downloaded on 11/9/2021 2:18:16 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

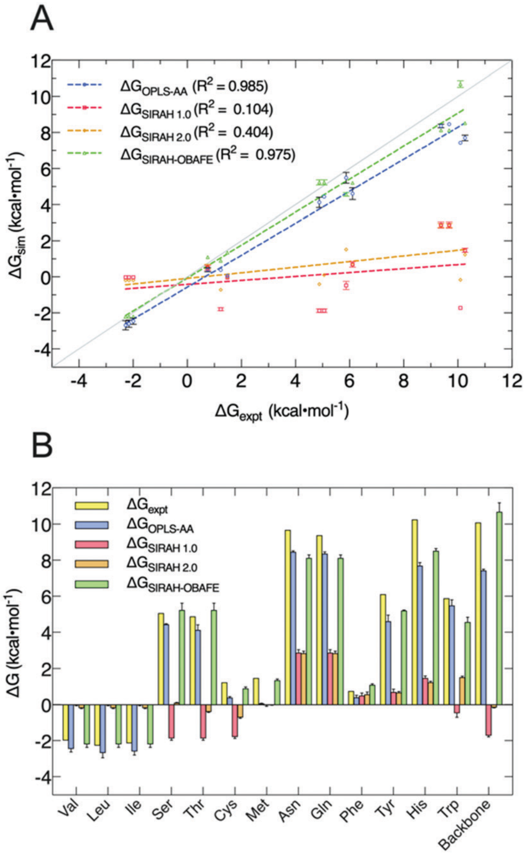

Fig. 2 Comparison of decoupling HFEs from the new set of optimised parameters (SIRAH-OBAFE) against atomistic simulations (AA), the original SIRAH 1.0 force field, the latest version SIRAH 2.0 and experimental data. (A) Linear regression of predicted ΔG values for AA (blue), SIRAH 1.0 (red), SIRAH 2.0 (orange) and SIRAH-OBAFE (green) force fields, against experimental data. Each point represents a specific side-chain. The grey line represents a perfect fit ($y = x$), and R^2 values are given in the inset legends. (B) Bar plot comparison of predicted ΔG values for AA (OPLS and AMBER-14SB) (blue), SIRAH 1.0 (red), SIRAH 2.0 (orange) and SIRAH-OBAFE (green) against experimental data (yellow; y axis) for all the neutral side-chains. Error estimates were calculated as standard errors based on three repeat simulations. For some cases, red bars appear to be missing as they are too small to be seen on the scale of the plot.

representing 13 uncharged side-chains because of the shared mapping scheme and bead types for some groups of side chains; *e.g.* Asn/Gln share the same mapping, as do Ser/Thr and Val/Leu/Ile, and the backbone. Fig. 2 and Table S3 (ESI†) summarises the performance of our new set of parameters for uncharged side-chains and the backbone, now called SIRAH-OBAFE, together with the new WT4-FB force field, against HFEs from atomistic force fields (OPLS-AA[49] for side-chains and AMBER-14SB[50] for the backbone), the original SIRAH 1.0 force field,[36] the updated SIRAH 2.0 force field,[44] and experimental data.[15] Those atomistic force fields with optimum published reference data were chosen to be part of the parametrisation process; the CG force field should be agnostic to the all atom data from which it is parameterised.

As can be seen, the original set of parameters in SIRAH 1.0 do not perform well for the prediction of decoupling HFEs, with an R^2 of 0.104 against experimental values (Fig. 2A). A similar case is observed for the latest version SIRAH 2.0, with an R^2 of 0.404 (Fig. 2A). SIRAH-OBAFE is able to greatly improve the agreement with experimental HFEs to be as good as atomistic force fields, with an R^2 of 0.985 and 0.975, respectively. SIRAH-OBAFE reproduces the correct sign of several neutral side-chains where the previous SIRAH 1.0 model predicted the wrong sign, such as Ser, Thr, Cys and Trp (Fig. 2B). Significant improvements have been made to the HFEs of hydrophobic residues such as Val, Leu, and Ile; these share the same representation in SIRAH, using just one bead. The original SIRAH 1.0 and the updated SIRAH 2.0 models predict −0.02 ± 0.01 and −0.18 ± 0.01 kcal mol^{-1} for the HFE, respectively (Table S3, ESI†), whereas SIRAH-OBAFE achieves a value of −2.26 ± 0.03 kcal mol^{-1} (Table S3, ESI†); the latter value is much closer to OPLS-AA simulations and experiment which provide HFEs of (−2.45, −2.69, −2.59) and (−1.99, −2.28, −2.15) kcal mol^{-1}, for Val, Leu, and Ile, respectively (Table S3, ESI†). In the case of methionine, SIRAH-OBAFE produced even more accurate HFE values than the OPLS-AA model that provided the HFE gradients to which the CG model was fitted; we think this result is fortuitous and the differences are within the residual errors of the CG model *vs.* the AT reference (see ESI† related to the methionine case and Fig. S2 and S3, ESI†).

Optimisation of the SIRAH protein force field: charged side-chains. A different approach, compared to the optimisation of uncharged side-chains and backbone, was followed for the charged side-chains. We started with ForceBalance optimisation procedures, where the gradients of the raw charging free energies plus the gradient of the methodology-dependent corrections were used (see methods). Calculated hydration free energies are reported in Table 2. Most of the ForceBalance optimisation results yield good agreement with experimental and AT hydration free energies (Table 2, denoted as HFE-fitted), but the parameters were driven to unphysical values, even when restraints to charges up to values of +1*e* or −1*e* were used (see Fig. S4 and S5, ESI,† for optimised charge and Lennard Jones parameters, respectively). Given this, we conclude that our optimisation procedure works, but given the existence of few parameters to represent charged side-chains in SIRAH, and despite the use of regularization, over-fitting might be an unavoidable consequence in this case. Moreover, coarse-graining is an important simplification of the physics, where the option to fully reproduce complex properties, such as the free energy of charged entities, might not be possible. We have therefore decided to use the original SIRAH 1.0 parameters for charged side-chains, in combination with the hydration free-energy optimised parameters for backbone and uncharged side-chains, for the test of the SIRAH-OBAFE force field on protein systems. To address the lack of physicality of the optimised charged side-chain parameters, the level of granularity of the coarse-grain representation will need to be revisited, and in particular that of the water model. This will necessitate a complete reparameterisation of the entire force field.

Protein simulations

To test the performance of SIRAH-OBAFE in protein simulations, Cα RMSD analyses (with respect to the crystal structure) were

Open Access Article. Published on 19 October 2021. Downloaded on 11/9/2021 2:18:16 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 2 Hydration free energies of charged side-chains using the GROMOS 54A8, SIRAH 1.0, and SIRAH 2.0 force fields. HFE-fitted values are also included with the sole intention of comparison and discussion

Force field	Expt.[a]	$\Delta G^{\text{raw}}_{\text{chg}} + \Delta G_{\text{cav}}$[b]	$\Delta G_{\text{A+B+D}}$	ΔG_{C_1}	$\Delta G^{\ominus}_{\text{std}}$	$\Delta G^{\ominus}_{\text{hyd}}$
ARG						
54A8	−276.5	−137.9 ± 0.4	−58.6	−67.8	7.9	−256.5
SIRAH 1.0 (original)		−149.0 ± 0.6	−57.5	−7.2	7.9	−205.8
SIRAH 2.0		−149.9 ± 0.6	−57.5	−7.2	7.9	−206.6
HFE-fitted		−223.7 ± 0.5	−54.8	−7.2	7.9	−277.8
LYS						
54A8	−289.5	−180.1 ± 0.6	−58.8	−67.8	7.9	−298.8
SIRAH 1.0 (original)		−134.5 ± 0.6	−54.7	−7.4	7.9	−188.7
SIRAH 2.0		−130.9 ± 0.5	−54.7	−7.4	7.9	−185.2
HFE-fitted		−178.6 ± 0.6	−57.5	−7.4	7.9	−235.8
GLU						
54A8	−315.4	−349.9 ± 0.5	−58.9	67.8	7.9	−332.9
SIRAH 1.0 (original)		−156.4 ± 0.4	−58.8	7.5	7.9	−199.3
SIRAH 2.0		−153.6 ± 0.5	−58.8	7.5	7.9	−196.5
HFE-fitted		−252.8 ± 0.6	−58.3	7.5	7.9	−295.6
ASP						
54A8	−321.2	−349.4 ± 0.5	−58.9	67.8	7.9	−332.5
SIRAH 1.0 (original)		−156.4 ± 0.4	−58.8	7.5	7.9	−199.2
SIRAH 2.0		−153.6 ± 0.5	−58.8	7.5	7.9	−196.4
HFE-fitted		−252.8 ± 0.6	−58.3	7.5	7.9	−295.6

[a] Values are in the units of kJ mol^{-1}. Experimental values were obtained from ref. 33. [b] Error bars modelled as standard errors across three repeat simulations.

performed on 6 protein systems of different sizes. Simulations using the optimised SIRAH-OBAFE with the optimised WT4-FB were run for 3 μs. While the computed RMSDs, calculated with respect to the initial crystal structure, are generally larger compared to atomistic simulations, all the simulations that used the optimised SIRAH-OBAFE model show improvements in protein stability with lower RMSD values throughout the whole trajectory with respect to the original SIRAH 1.0 force field (Fig. 3).

Even though the overall behaviour of the optimised SIRAH-OBAFE FF does not yield identical results compared to

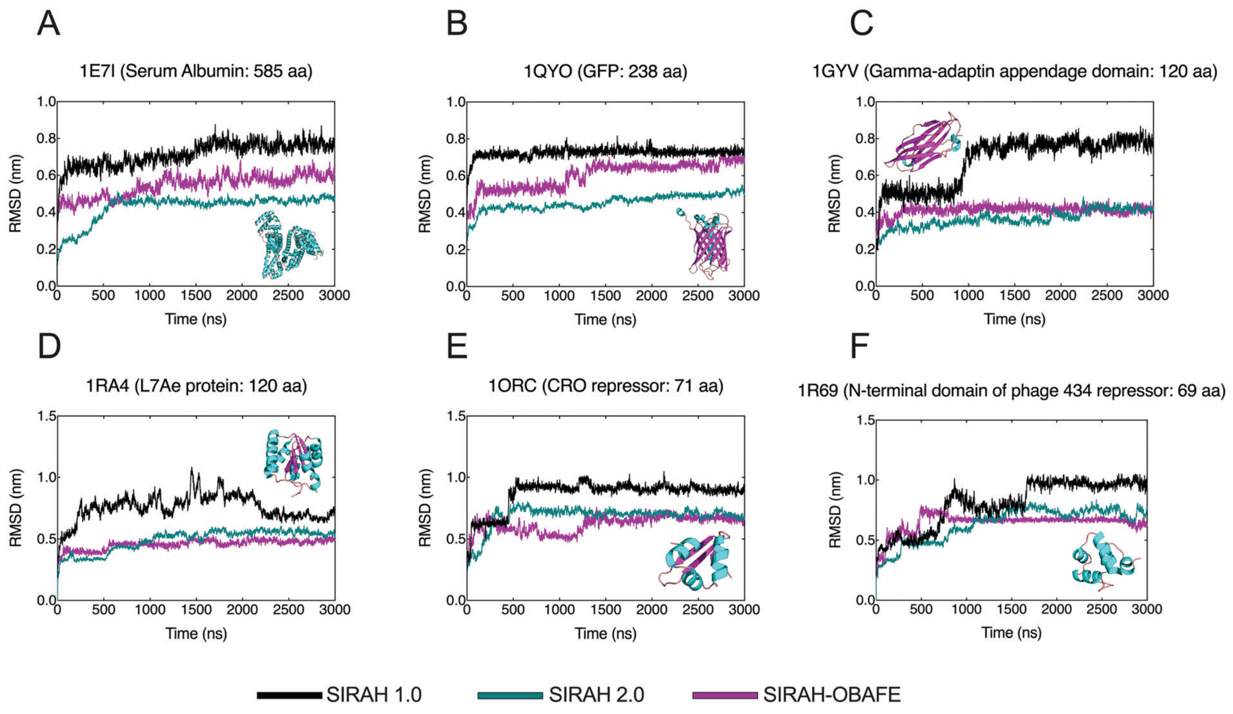

Fig. 3 RMSD time series comparison between the original SIRAH 1.0 FF (black), the updated SIRAH 2.0 FF (cyan) and the optimised SIRAH-OBAFE FF (purple). RMSD trajectory analysis is shown as a time series comparison with respect to the Cα carbons of the CG representation to the crystal structure for (A) Serum albumin, (B) GFP protein, (C) Gamma-adaptin domain, (D) L7Ae protein, (E) CRO repressor and (F) the N-terminal domain of phage 434 repressor. PDB codes are shown in the figure titles and legend colours are shown at the bottom of the figure. Protein structures, corresponding to each of the simulated cases, are shown inside each plot. All simulations and analysis were run in GROMACS v.2018.2.

Open Access Article. Published on 19 October 2021. Downloaded on 11/9/2021 2:18:16 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

atomistic RMSDs, it shows an important improvement compared to the original SIRAH FF. As a simple comparison, atomistic simulation of systems with PDB codes 1QYO, 1RA4 and 1R69 (chosen from the original SIRAH 1.0 publication[36]) were run for 200 ns, showing average RMSD values of 0.270 nm, 0.06 nm and 0.148 nm, respectively, with respect to the reference crystal structure. The original SIRAH 1.0 force field shows averaged RMSD values of 0.723 nm, 0.755 nm and 0.804 nm, while our optimised SIRAH-OBAFE force field shows averaged values (over the entire simulation) of 0.453 nm, 0.491 nm and 0.635 nm, for the same three cases, 1QYO, 1RA4 and 1R69, respectively. In the case of the updated SIRAH 2.0 force field, the overall behaviour of the RMSD timeseries is similar to the optimised SIRAH-OBAFE (Fig. 3), except for two larger systems, with average values (over the entire simulation) of 0.543 nm (SIRAH 2.0) *vs.* 0.429 nm (SIRAH-OBAFE) for 1E7I, and 0.601 nm (SIRAH 2.0) *vs.* 0.453 nm (SIRAH-OBAFE) for 1QYO (Fig. 3A and B). Even though the new RMSD values are not close to the atomistic RMSD and we would not necessarily expect them to be, there is as an improvement in the stability of protein systems based on our new optimisation approach. In more detail, calculating RMSD values against the last frame of the trajectories can give an insight into whether the large RMSDs are due to large fluctuations, or a change in conformation to another stable conformer. This analysis was performed using the updated SIRAH 2.0 and the optimised SIRAH-OBAFE force fields. Fig. S6 (ESI†) summarise the results. In the case of the 1E7I system, a big change of conformation is seen at around 1 μs, which stabilise afterwards. In the other systems, it seems that the higher RMSD values observed in Fig. 3 are due to conformational drift across the simulation, with similar behaviours for both the updated SIRAH 2.0 and the optimised SIRAH-OBAFE force fields. RMSD fluctuations within a particular protein conformation are of the order of 0.2 nm.

Conclusions

In this work we propose a new and promising approach for parametrising coarse-grain force fields by optimising the CG force field parameters against free energy gradients derived from atomistic simulations. Our implementation of this method into ForceBalance enables full automation of the complex optimisation procedure and the incorporation of flexible choices of target data. It has been stated that there is considerable interest in methods that can automatically generate a coarse-grained model, and that are representative in terms of local structure and free energy changes. Our method paves the way to new optimisation procedures that rely on the use of free energy data as a target.

Non-bonded interaction parameters of un-charged side-chains and the backbone of the SIRAH coarse grain protein force field were optimised against hydration free energy gradients of atomistic simulation models, and compared against experimental hydration free energies, yielding a new parameter set called SIRAH-OBAFE (Table S2, ESI† for parameter values).

The predicted hydration free energies show a much improved agreement with experiment, compared to the previous version of the SIRAH force field, with increased R^2 values of 0.97 for the new SIRAH-OBAFE parameter set, against values of 0.1 and 0.4 for the SIRAH 1.0 and SIRAH 2.0 sets, respectively. Attempts were made to optimise charged side-chains, using free energy gradients, with the necessary correction gradients. While force field parameters able to give improved estimates of the hydration free energies were derived, given the difficulty in this process to avoid an over-fitted model, even with regularization, and the lack of sufficient parameters to improve the hydration free energies in a physically meaningful way, the original charged parameters of the SIRAH force field were retained. The structural stability of proteins has been improved with the use of the new SIRAH-OBAFE force field. RMSD values were reduced by an average of ∼0.25 nm across the protein system tested (Fig. 3), compared to the original SIRAH 1.0 force field, which was used as the starting point in the optimisation procedures.

We believe that the simplification of the physics observed in coarse-grained force fields, such as the SIRAH model, presents a challenge for the reproduction of multiple experimental properties. Limitations in the optimisation methodology are arguably the main cause for this, mainly given by the size of the parameter set that is available to optimise the property of interest; there is insufficient granularity to capture the physics involved in the calculation of hydration free energies for charged and neutral species in a balanced way. The few parameters available in CG models will likely limit the applicability of our proposed optimisation method. To better understand the implications, future studies could be related to the use of a more complex CG protein force field (near atomistic resolution) in the optimisation process, and different scenarios in terms of protein simulations, such as calculating protein potentials of mean force (PMF) for conformational changes and the folding of small peptides. Moreover, a procedure to simultaneously and automatically include PMF data in the ForceBalance parameterisation might bring improvements. Significant and further validation is needed.

The development of new strategies and approaches for force field optimisation is of great interest. In this matter, our new method opens a door for the improvement of contemporary, or new, CG force fields, and it greatly increase the applicability of the CG models in different research areas, such as the study of protein conformational changes, which needs a correct description of protein–protein and protein-solute interactions. Furthermore, the parameterisation approach opens a new route to developing CG force fields for other classes of biomolecules such as carbohydrates, nucleic acids, lipids and metabolites, where experimental data may not be as readily available.

Conflicts of interest

There are no conflicts to declare.

Open Access Article. Published on 19 October 2021. Downloaded on 11/9/2021 2:18:16 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Acknowledgements

Calculations in this work made use of the Iridis4 supercomputers at the University of Southampton. LPW acknowledges funding support from NIH R01 AI130684-01A1. JCD gratefully acknowledges funding support from CONICYT-BECAS CHILE.

Notes and references

1 D. J. Huggins, P. C. Biggin, M. A. Dämgen, J. W. Essex, S. A. Harris, R. H. Henchman, S. Khalid, A. Kuzmanic, C. A. Laughton, J. Michel, A. J. Mulholland, E. Rosta, M. S. P. Sansom and M. W. van der Kamp, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2019, **9**(3), e1393.

2 A. Hospital, J. R. Goñi, M. Orozco and J. L. Gelpí, *Adv. Appl. Bioinf. Chem.*, 2015, **8**, 37–47.

3 H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole and S. J. Marrink, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 225–248.

4 S. C. L. Kamerlin, S. Vicatos, A. Dryga and A. Warshel, *Annu. Rev. Phys. Chem.*, 2011, **62**, 41–64.

5 S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid and A. Kolinski, *Chem. Rev.*, 2016, **116**, 7898–7936.

6 T. A. Halgren, *Curr. Opin. Struct. Biol.*, 1995, **5**, 205–210.

7 M. S. Miller, W. K. Lay, S. Li, W. C. Hacker, J. An, J. Ren and A. H. Elcock, *J. Chem. Theory Comput.*, 2017, **13**, 1812–1826.

8 D.-W. Li and R. Brüschweiler, *J. Chem. Theory Comput.*, 2011, **7**, 1773–1782.

9 M. Orsi and J. W. Essex, *PLoS One*, 2011, **6**, e28637.

10 Y. Li, H. Li, F. C. Pickard, B. Narayanan, F. G. Sen, M. K. Y. Chan, S. K. R. S. Sankaranarayanan, B. R. Brooks and B. Roux, *J. Chem. Theory Comput.*, 2017, **13**, 4492–4503.

11 V. Botu, R. Batra, J. Chapman and R. Ramprasad, *J. Phys. Chem. C*, 2016, **121**, 511–522.

12 R. M. Betz and R. C. Walker, *J. Comput. Chem.*, 2014, **36**, 79–87.

13 B. Waldher, J. Kuta, S. Chen, N. Henson and A. E. Clark, *J. Comput. Chem.*, 2010, **31**, 2307–2316.

14 L.-P. Wang, J. Chen and T. Van Voorhis, *J. Chem. Theory Comput.*, 2012, **9**, 452–460.

15 L.-P. Wang, T. J. Martínez and V. S. Pande, *J. Phys. Chem. Lett.*, 2014, **5**, 1885–1891.

16 L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martínez and V. S. Pande, *J. Phys. Chem. B*, 2013, **117**, 9956–9972.

17 M. L. Laury, L.-P. Wang, V. S. Pande, T. Head-Gordon and J. W. Ponder, *J. Phys. Chem. B*, 2015, **119**, 9423–9437.

18 R. Qi, L.-P. Wang, Q. Wang, V. S. Pande and P. Ren, *J. Chem. Phys.*, 2015, **143**, 014504.

19 K. A. McKiernan, L.-P. Wang and V. S. Pande, *J. Chem. Theory Comput.*, 2016, **12**, 5960–5967.

20 L.-P. Wang, K. A. McKiernan, J. Gomes, K. A. Beauchamp, T. Head-Gordon, J. E. Rice, W. C. Swope, T. J. Martínez and V. S. Pande, *J. Phys. Chem. B*, 2017, **121**, 4023–4039.

21 S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845–854.

22 Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder and P. Ren, *J. Chem. Theory Comput.*, 2013, **9**, 4046–4063.

23 P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts and V. S. Pande, *J. Chem. Theory Comput.*, 2013, **9**, 461–469.

24 M. R. Shirts, J. W. Pitera, W. C. Swope and V. S. Pande, *J. Chem. Phys.*, 2003, **119**, 5740–5761.

25 R. T. Bradshaw and J. W. Essex, *J. Chem. Theory Comput.*, 2016, **12**, 3871–3883.

26 S. A. Martins, S. F. Sousa, M. J. Ramos and P. A. Fernandes, *J. Chem. Theory Comput.*, 2014, **10**, 3570–3577.

27 J. Chang, A. M. Lenhoff and S. I. Sandler, *J. Phys. Chem. B*, 2007, **111**, 2098–2106.

28 T. D. Potter, J. Tasche and M. R. Wilson, *Phys. Chem. Chem. Phys.*, 2019, **21**, 1912–1927.

29 C. Chipot and A. Pohorille, *Springer Ser. Chem. Phys.*, 2007, **86**, 159–184.

30 C. D. Christ, A. E. Mark and W. F. van Gunsteren, *J. Comput. Chem.*, 2010, **31**, 1569–1582.

31 A. Pohorille, C. Jarzynski and C. Chipot, *J. Phys. Chem. B*, 2010, **114**(32), 10235–10253.

32 J. Michel and J. W. Essex, *J. Comput. Aided Mol. Des.*, 2010, **24**, 639–658.

33 M. M. Reif, P. H. Hünenberger and C. Oostenbrink, *J. Chem. Theory Comput.*, 2012, **8**, 3705–3723.

34 M. M. Reif and P. H. Hünenberger, *J. Chem. Phys.*, 2011, **134**, 144104.

35 M. M. Reif and P. H. Hünenberger, *J. Chem. Phys.*, 2011, **134**, 144103.

36 L. Darré, M. R. Machado, A. F. Brandner, H. C. González, S. Ferreira and S. Pantano, *J. Chem. Theory Comput.*, 2015, **11**, 723–739.

37 L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S.-J. Marrink, *J. Chem. Theory Comput.*, 2008, **4**, 819–834.

38 P. D. Dans, L. Darré, M. R. Machado, A. Zeida, A. F. Brandner and S. Pantano, *Advances in Bioinformatics and Computational Biology*, Springer International Publishing, Cham, 2013, vol. 8213, pp. 71–81.

39 L. Darré, M. R. Machado, P. D. Dans, F. E. Herrera and S. Pantano, *J. Chem. Theory Comput.*, 2010, **6**, 3793–3807.

40 E. E. Barrera, E. N. Frigini, R. D. Porasso and S. Pantano, *J. Mol. Model.*, 2017, **23**, 259.

41 A. Brandner, A. Schüller, F. Melo and S. Pantano, *Biochem. Biophys. Res. Commun.*, 2018, **498**(2), 319–326.

42 M. R. Machado, P. D. Dans and S. Pantano, *Phys. Chem. Chem. Phys.*, 2011, **13**, 18134–18144.

43 M. A. R. Machado, H. C. González and S. Pantano, *J. Chem. Theory Comput.*, 2017, **13**, 5106–5116.

44 M. R. Machado, E. E. Barrera, F. Klein, M. Sóñora, S. Silva and S. Pantano, *J. Chem. Theory Comput.*, 2019, **15**, 2719–2733.

45 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

46 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

47 M. Parrinello and A. Rahman, *Phys. Rev. Lett.*, 1980, **45**, 1196–1199.

48 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.

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

50 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

Open Access Article. Published on 19 October 2021. Downloaded on 11/9/2021 2:18:16 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.