



# Empirical optimization of molecular simulation force fields by Bayesian inference

Jürgen Köfner<sup>1,a</sup> and Gerhard Hummer<sup>1,2,b</sup>

<sup>1</sup> Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Max-von-Laue Str. 3, 60438 Frankfurt am Main, Germany

<sup>2</sup> Institute for Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

Received 28 July 2021 / Accepted 28 October 2021 / Published online 17 December 2021  
© The Author(s) 2021

**Abstract.** The demands on the accuracy of force fields for classical molecular dynamics simulations are steadily growing as larger and more complex systems are studied over longer times. One way to meet these growing demands is to hand over the learning of force fields and their parameters to machines in a systematic (semi)automatic manner. Doing so, we can take full advantage of exascale computing, the increasing availability of experimental data, and advances in quantum mechanical computations and the calculation of experimental observables from molecular ensembles. Here, we discuss and illustrate the challenges one faces in this endeavor and explore a way forward by adapting the Bayesian inference of ensembles (BioEn) method [Hummer and Köfner, *J. Chem. Phys.* (2015)] for force field parameterization. In the Bayesian inference of force fields (BioFF) method developed here, the optimization problem is regularized by a simplified prior on the force field parameters and an entropic prior acting on the ensemble. The latter compensates for the unavoidable over-simplifications in the parameter prior. We determine optimal force field parameters using an iterative predictor–corrector approach, in which we run simulations, determine the reference ensemble using the weighted histogram analysis method (WHAM), and update the force field according to the BioFF posterior. We illustrate this approach for a simple polymer model, using the distance between two labeled sites as the experimental observable. By systematically resolving force field issues, instead of just reweighting a structural ensemble, the BioFF corrections extend to observables not included in ensemble reweighting. We envision future force field optimization as a formalized, systematic, and (semi)automatic machine-learning effort that incorporates a wide range of data from experiment and high-level quantum chemical calculations, and takes advantage of exascale computing resources.

## 1 Introduction

The predictive power of molecular dynamics (MD) simulations relies on the accuracy and the computational efficiency of the applied force fields describing the molecular interactions. Widely used semi-empirical force fields aim to strike a balance between accuracy and efficiency [1,2]. The latter restricts their computational complexity. Force fields are parameterized in part by quantum mechanical calculations and in part by fitting calculated observables to experimental data. For decades, molecular force fields have been adapted and re-parameterized to fit an ever-growing resource of theoretical and experimental information [3]. These cumulative efforts have dramatically increased the predictive power of MD simulations [4].

The advent of exascale computing makes it possible – and ultimately necessary – to hand over the task of force field optimization to machines in a formal-

ized, systematic, and (semi)automatic manner [5–11]. However, the complexity of molecular force fields, their high-dimensional parameter spaces, and the interdependence of the parameters pose difficult challenges for automated force field optimization. Whereas this problem is well recognized for classical mechanical energy functions, a quite similar problem arises in quantum mechanical modeling. The latter also involves multiple model choices (say, which level of description or which basis set to use) or extensive parameterizations (e.g., in density functional theory).

Intrinsically, force field calibration is a problem of statistical inference, where we update our current understanding of molecular interactions, e.g., from quantum mechanical calculations for molecular fragments, with uncertain information from various sources [9,12]. Quantum mechanical calculations, experimental measurements, MD simulations, and the calculations of experimental observables from molecular models are all subject to systematic or sampling errors. In a Bayesian formulation, a prior distribution on the force field parameters encodes our current state of knowledge which is then updated with new experimental data

<sup>a</sup> e-mail: [juergen.koefinger@biophys.mpg.de](mailto:juergen.koefinger@biophys.mpg.de) (corresponding author)

<sup>b</sup> e-mail: [gerhard.hummer@biophys.mpg.de](mailto:gerhard.hummer@biophys.mpg.de)

through the likelihood. Unfortunately, it is difficult to define such priors. The complex and often strong interdependence of force field parameters is hard to quantify. Consequently, without good priors, even small changes in the parameters can lead to highly non-physical conformations in Boltzmann sampling.

We illustrate these challenges and explore the use of maximum entropy [13–16] and Bayesian [17–19] ensemble refinement methods to compensate for the lack of good priors on the force field parameters. In ensemble refinement, we use experimental data to refine the statistical weights of an ensemble of structures. Our starting point is the Bayesian inference of ensembles (BioEn) method [18]. In BioEn, an entropic prior acts on the statistical weights of the structures in an ensemble such that refinement changes the weights only minimally to better fit the experimental data. Properties of the system that are independent of the data remain untouched.

Here, we extend BioEn by expressing the statistical weights as functions on selected force field parameters and by adding a prior on these parameters. Unavoidable deficiencies in the parameter prior are compensated by the entropic prior on the ensemble of structures. In this Bayesian inference of force fields (BioFF) method, we iteratively optimize the resulting BioFF posterior by incorporating the results of MD simulations with original and newly optimized force field parameters to ensure convergence. To illustrate the BioFF method, we infer the value of the single force field parameter of a simple polymer model using a single label distance as experimental data.

The article is organized as follows. In Sect. 2 we briefly review ensemble refinement in the context of force field parameterization and introduce the BioFF method. We present method details in Sect. 3 and a simple proof of principle for a two-dimensional polymer in Sect. 4. We discuss the results in a broader context in Sect. 5 and present a tentative outlook on the future of force field optimization and conclusions in Sect. 6.

## 2 Theory

We briefly introduce ensemble refinement methods and review their use for force field optimization before presenting the BioFF method. We do so step-by-step: We first present the BioFF posterior in its general form as a probability density function of the force field parameters, which we then adapt for discrete ensembles. For optimization, we reweight the discrete weights as function of the force field parameters. The key to the accuracy and efficiency of BioFF by reweighting is to pool all MD simulations generated for different force fields and calculate new reference weights using the binless weighted histogram analysis method (WHAM) [20–22]. Equivalently, one can use the multistate Bennett acceptance ratio (MBAR) method [23, 24]. Binless WHAM has to be executed only once for each new simulation added to the pool. Note that pooling all simulations for

different force field parameters using binless WHAM also distinguishes BioFF from previous methods pursuing similar goals [25–27]. Pooling and priors stabilize the optimization procedure and minimize the risk of sampling non-physical conformations. We sketch the BioFF-by-reweighting procedure at the end of the section.

### 2.1 Background and preliminaries

In ensemble refinement, we adapt the statistical weights of the structures in an ensemble to obtain better agreement with experimentally measured ensemble averages [15, 18, 28]. By properly accounting for uncertainties and balancing the information encoded in the structural ensemble and in the experimental data, the resulting ensemble is a better representation of the true ensemble underlying the experiment. The structural ensemble is usually generated in free or restrained MD simulations. The refinement of such simulation ensembles by reweighting ensures convergence with ensemble size [18].

Many of the current approaches to ensemble refinement that appear to be different on the surface are actually quite similar and will in principle provide nearly the same optimal ensembles if we encode the same prior knowledge [13, 17–19, 29–32]. In the following, we use the formalism of the BioEn method [18] to introduce ensemble refinement. We then adapt BioEn for force field parameterization.

Consider a potential energy function  $U(\mathbf{x}|\mathbf{c})$  for conformations  $\mathbf{x}$  that depends on a vector of  $m$  force field parameters  $c_i$ ,  $\mathbf{c} = (c_1, \dots, c_m)$ . Structures are distributed according to the normalized Boltzmann distribution

$$p(\mathbf{x}|\mathbf{c}) = \frac{\exp[-\beta U(\mathbf{x}|\mathbf{c})]}{\int d\mathbf{x}' \exp[-\beta U(\mathbf{x}'|\mathbf{c})]} \quad (1)$$

$\beta = 1/k_B T$  is the inverse temperature and  $k_B$  is Boltzmann's constant.

For uncorrelated Gaussian errors, the BioEn posterior is given by the functional [18]

$$\mathcal{P}[p(\mathbf{x})] \propto e^{-\theta S_{\text{KL}}[p(\mathbf{x})] - \chi^2[p(\mathbf{x})]/2} \quad (2)$$

where the Kullback–Leibler divergence [33] is given by

$$S_{\text{KL}}[p(\mathbf{x})] = \int d\mathbf{x} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{p(\mathbf{x}|\mathbf{c}_0)} \quad (3)$$

and where

$$\chi^2[p(\mathbf{x})] = \sum_{k=1}^M \frac{[\int d\mathbf{x} p(\mathbf{x}) y_k(\mathbf{x}) - Y_k]^2}{\sigma_k^2} \quad (4)$$

is the chi-squared defined as the sum of the squared errors in units of the estimated standard errors.  $p(\mathbf{x}|\mathbf{c}_0)$  is the reference ensemble or  $\mathbf{c}_0$ -ensemble.  $Y_k$  are the

$M$  experimental data points,  $y_k(\mathbf{x})$  are the corresponding calculated observable values for structure  $\mathbf{x}$ , and  $\sigma_k$  is the combined theoretical and experimental error.  $\theta$  expresses our confidence in the reference ensemble [13, 18].

By maximizing the BioEn posterior given by Eq. (2) with respect to  $p(\mathbf{x})$ , we obtain the BioEn optimal ensemble

$$p(\mathbf{x}) \propto p(\mathbf{x}|\mathbf{c}_0) \exp\left(\sum_{j=1}^M y_j(\mathbf{x}) f_j\right) \tag{5}$$

determined self-consistently by the  $M$  generalized forces

$$f_k = -\frac{\langle y_k \rangle - Y_k}{\theta \sigma_k^2}. \tag{6}$$

Angular brackets indicate the average over all structures in the optimal ensemble, i.e.,  $\langle y_k \rangle = \int d\mathbf{x} p(\mathbf{x}) y_k(\mathbf{x})$  for a normalized  $p(\mathbf{x})$ . Thus, the mean force to restore the reference distribution is exactly balanced by the mean force to fit the data [18]. The generalized forces correspond to Lagrange multipliers in the corresponding expressions for Eq. (5) in the maximum entropy formulations of ensemble refinement [14, 28, 31, 34, 35]. The generalized forces parameterizing the optimal ensemble can be efficiently determined by gradient-based optimization methods [36, 37]. The optimal ensemble corresponds to a Boltzmann ensemble  $\propto \exp[-\beta U(\mathbf{x}|\mathbf{c}_0) - \beta \Delta U(\mathbf{x})]$ , where the energy is given by the sum of the potential energy of the reference ensemble  $\beta U(\mathbf{x}|\mathbf{c}_0)$  and a correction term

$$\beta \Delta U(\mathbf{x}) = -\sum_{k=1}^M y_k(\mathbf{x}) f_k \tag{7}$$

which is linear in the calculated observables  $y_k^\alpha = y_k(\mathbf{x}_\alpha)$  [14, 17, 18].

For well sampled ensembles, the bias energy given by Eq. (7) can be used to update force field parameters. However, in general, it is not guaranteed that we can adapt the force field parameters to fit  $\beta \Delta U(\mathbf{x})$ . This correction may not be adequately representable given the functional form of the force field. In the special case of  $^3J$  couplings, Cesari et al. exploited the functional similarity of the Karplus equation to dihedral potentials to calibrate RNA force fields [31]. More recently, Cesari et al. generalized this approach and replaced the observables  $y_k(\mathbf{x})$  in the expression for the optimal ensemble, Eq. (5), by a potential energy correction function and treated the generalized forces as fitting parameters [38]. A regularized error function keeps the calculated observables close to the measured values and the fitting parameters small. Cesari et al. tune the latter using a parameter to balance experiment and MD simulations similar to the BioEn confidence parameter.

In any systematic approach to force field refinement [5, 27, 31, 38, 39], we have to regularize the problem and

balance the information from experiment and MD simulations for the given uncertainties. BioEn has both features already built in. In the following, we adapt BioEn for systematic force field refinement.

### 2.2 The BioFF posterior

For force field refinement, we restrict the function space of  $p(\mathbf{x})$  in the BioEn posterior, Eq. (2), to probability distributions parameterized by  $\mathbf{c}$ , i.e., to  $p(\mathbf{x}|\mathbf{c}) \propto \exp[-\beta U(\mathbf{x}|\mathbf{c})]$ . Note that  $\mathbf{c}$  denotes the parameters we want to optimize and that  $p(\mathbf{x}|\mathbf{c})$  will usually depend on additional force field parameters. Doing so, the BioEn posterior becomes a function of  $\mathbf{c}$ . Introducing an additional prior  $p_0(\mathbf{c})$  on the force field parameters, we obtain the BioFF negative log-posterior as

$$\begin{aligned} \mathcal{L}(\mathbf{c}) = & -\ln p_0(\mathbf{c}) \\ & + \theta \int d\mathbf{x} p(\mathbf{x}|\mathbf{c}) \ln \frac{p(\mathbf{x}|\mathbf{c})}{p(\mathbf{x}|\mathbf{c}_0)} \\ & + \sum_{i=1}^M \frac{[\int d\mathbf{x} p(\mathbf{x}|\mathbf{c}) y_i(\mathbf{x}) - Y_i]^2}{2\sigma_i^2} \end{aligned} \tag{8}$$

The BioFF optimal solution satisfies

$$\begin{aligned} 0 = \mathcal{L}'(\mathbf{c}) = & -\frac{p'_0(\mathbf{c})}{p_0(\mathbf{c})} \\ & + \theta \int d\mathbf{x} p'(\mathbf{x}|\mathbf{c}) \left[ \ln \frac{p(\mathbf{x}|\mathbf{c})}{p(\mathbf{x}|\mathbf{c}_0)} + 1 \right] \\ & + \sum_{i=1}^M \frac{\int d\mathbf{x} p'(\mathbf{x}|\mathbf{c}) y_i(\mathbf{x}) [\int d\mathbf{x} p(\mathbf{x}|\mathbf{c}) y_i(\mathbf{x}) - Y_i]}{\sigma_i^2} \end{aligned} \tag{9}$$

where prime ( $'$ ) denotes a gradient with respect to  $\mathbf{c}$ . We show next how to optimize the BioFF posterior for a given ensemble of structures, which is at the core of the BioFF-by-reweighting procedure presented below.

### 2.3 BioFF optimization by reweighting

Let us assume we performed a reference simulation using the force field parameters  $\mathbf{c}_0$ . For unbiased MD simulations, each of the  $N$  structures in this reference ensemble has a statistical weight  $w_\alpha^{(0)} = 1/N$ . The optimal weights  $w_\alpha(\mathbf{c})$  of structures  $\mathbf{x}_\alpha$  minimize the BioFF posterior

$$\begin{aligned} \mathcal{L}(\mathbf{c}) = & -\ln p_0(\mathbf{c}) + \theta \sum_{\alpha=1}^N w_\alpha(\mathbf{c}) \ln \frac{w_\alpha(\mathbf{c})}{w_\alpha^{(0)}} \\ & + \sum_{i=1}^M \frac{[\sum_{\alpha=1}^N w_\alpha(\mathbf{c}) y_i(\mathbf{x}_\alpha) - Y_i]^2}{2\sigma_i^2} \end{aligned} \tag{10}$$

and thus satisfy

$$0 = \mathcal{L}'(\mathbf{c}) = -\frac{p'_0(\mathbf{c})}{p_0(\mathbf{c})} + \theta \sum_{\alpha=1}^N w'_\alpha(\mathbf{c}) \left[ \ln \frac{w_\alpha(\mathbf{c})}{w_\alpha^{(0)}} + 1 \right] + \sum_{i=1}^M \frac{\sum_{\alpha=1}^N w'_\alpha(\mathbf{c}) y_i(\mathbf{x}_\alpha) \left[ \sum_{\alpha=1}^N w_\alpha(\mathbf{c}) y_i(\mathbf{x}_\alpha) - Y_i \right]}{\sigma_i^2} \tag{11}$$

Note that  $w_\alpha(\mathbf{c})$  and  $w_\alpha^{(0)}$  are properly normalized. The  $\mathbf{c}$ -dependence of the weights  $w_\alpha(\mathbf{c})$  can be expressed by reweighting of the  $\mathbf{c}_0$ -ensemble to the  $\mathbf{c}$ -ensemble via

$$w_\alpha(\mathbf{c}) = w_\alpha^{(0)} \frac{p(\mathbf{x}_\alpha|\mathbf{c})}{p(\mathbf{x}_\alpha|\mathbf{c}_0)} \left[ \sum_{\gamma=1}^N w_\gamma^{(0)} \frac{p(\mathbf{x}_\gamma|\mathbf{c})}{p(\mathbf{x}_\gamma|\mathbf{c}_0)} \right]^{-1} \tag{12}$$

which becomes

$$w_\alpha(\mathbf{c}) = \frac{w_\alpha^{(0)} e^{-\beta[U(\mathbf{x}_\alpha|\mathbf{c})-U(\mathbf{x}_\alpha|\mathbf{c}_0)]}}{\sum_{\gamma=1}^N w_\gamma^{(0)} e^{-\beta[U(\mathbf{x}_\gamma|\mathbf{c})-U(\mathbf{x}_\gamma|\mathbf{c}_0)]}}. \tag{13}$$

The gradient of these weights, which is needed to evaluate the gradient of the negative log-posterior, Eq. (11), is given by

$$w'_\alpha(\mathbf{c}) = w_\alpha(\mathbf{c}) \left[ \sum_{\gamma=1}^N w_\gamma(\mathbf{c}) \beta U'(\mathbf{x}_\gamma|\mathbf{c}) - \beta U'(\mathbf{x}_\alpha|\mathbf{c}) \right]. \tag{14}$$

The calculation of the Hessian is straightforward. The Hessian is needed for optimization using the (L-)BFGS algorithm, for example. The evaluation of Eq. (13) for the weights  $w_\alpha(\mathbf{c})$  is computationally cheap such that numerical optimization can be performed efficiently.

### 2.4 Data for multiple different systems

We note that Eqs. (10) and (11) can readily be generalized to incorporate  $M_s$  data from multiple different systems  $s = 1, \dots, S$  sampled with  $N_s$  conformations each, as will be the case in most force field optimizations. The negative log-likelihood then becomes

$$\mathcal{L}(\mathbf{c}) = -\ln p_0(\mathbf{c}) + \theta \sum_{s=1}^S \left( \sum_{\alpha=1}^{N_s} w_{\alpha s}(\mathbf{c}) \ln \frac{w_{\alpha s}(\mathbf{c})}{w_{\alpha s}^{(0)}} + \sum_{i=1}^{M_s} \frac{\left[ \sum_{\alpha=1}^{N_s} w_{\alpha s}(\mathbf{c}) y_{is}(\mathbf{x}_{\alpha s}) - Y_{is} \right]^2}{2\sigma_{is}^2} \right) \tag{15}$$

For each system  $s$ , the weights satisfy Eqs. (13) and (14) with  $w_\alpha, \mathbf{x}_\alpha$ , and  $N$  replaced by  $w_{\alpha s}, \mathbf{x}_{\alpha s}$ , and  $N_s$ , respectively. If systems differ also in the thermodynamic state, e.g., temperature, then Eq. (13) and following have to be adapted accordingly.

### 2.5 Data for individual conformations from quantum mechanical calculations

In many practical cases, one also has data  $Y_j(\mathbf{x}_j)$  ( $j = M + 1, \dots, M + M_*$ ) for individual conformations  $\mathbf{x}_j$ , such as forces or potential energies from quantum mechanical calculations. Note that  $x_j$  can refer to the same conformation for different values of  $j$ . We can include such data in BioFF by adding a term  $\chi_*^2/2$  to the negative log-posterior,

$$\chi_*^2 = \sum_{j=M+1}^{M+M_*} \frac{[y_j(\mathbf{x}_j|\mathbf{c}) - Y_j(\mathbf{x}_j)]^2}{\sigma_j^2} \tag{16}$$

where  $y_j(\mathbf{x}_j|\mathbf{c})$  is the corresponding calculation for the force field with parameters  $\mathbf{c}$ , and  $\sigma_j^2$  is the combined squared uncertainty of data and calculation.  $\chi_*^2/2$  can be added to  $\mathcal{L}(\mathbf{c})$  in Eq. (10). Its gradient with respect to  $\mathbf{c}$  is

$$\left( \frac{\chi_*^2}{2} \right)' = \sum_{j=M+1}^{M+M_*} \frac{y'_j(\mathbf{x}_j|\mathbf{c}) [y_j(\mathbf{x}_j|\mathbf{c}) - Y_j(\mathbf{x}_j)]}{\sigma_j^2} \tag{17}$$

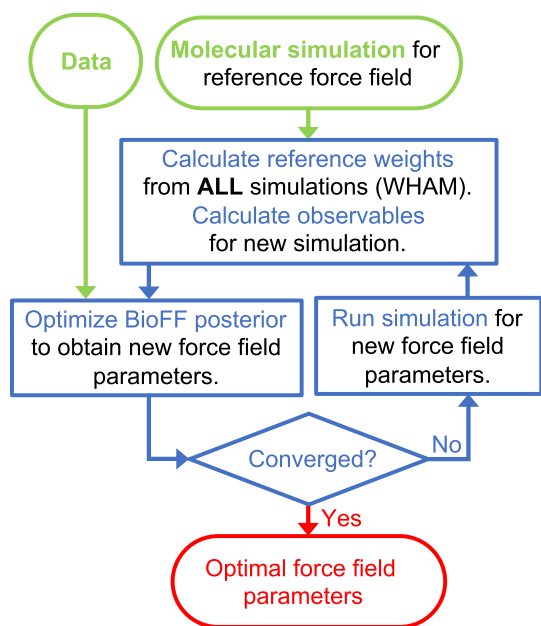
This expression can be added to the gradient  $\mathcal{L}'(\mathbf{c})$  in Eq. (11) to account for data reporting on the properties of individual conformations.

### 2.6 The BioFF-by-reweighting method

In the following, we sketch an iteration procedure to determine the BioFF optimal parameters by reweighting (Fig. 1).

At iteration 0, we start with a single simulation using force field parameters  $\mathbf{c}_{i=0}$  and calculate all  $M$  observables  $y_k(\mathbf{x}_\alpha)$  for all  $N_0$  structures in this initial estimate for the reference ensemble. The reference weights are given by  $w_\alpha^{(0)} = 1/N_0$ . We minimize Eq. (10) numerically to obtain new parameters  $\mathbf{c}_{i=1}$ .

1. We run an additional simulation for the newly optimized force field parameters  $\mathbf{c}_i$  and generate  $N_i$  structures in the  $\mathbf{c}_i$ -ensemble.
2. We calculate  $N = \sum_{j=0}^i N_j$  new reference weights  $w_\alpha^{(0)}$  from all MD simulations at parameters  $\mathbf{c}_0, \dots, \mathbf{c}_i$  using binless WHAM [21, 22, 24]. We calculate all observables  $y_k(\mathbf{x}_\alpha)$  for the new simulation ensemble.
3. We minimize Eq. (10) numerically to obtain new parameters  $\mathbf{c}_{i+1}$ .
4. We check for convergence of the force field parameters. If not converged, we continue with step 1 and the new force field parameters  $\mathbf{c}_{i+1}$ . If converged,



**Fig. 1** Flowchart of the BioFF-by-reweighting optimization procedure. The input (green) are experimental or theoretical data and a structural ensemble generated in a molecular simulation using the reference force field. The calculation of the reference weights and observables, optimization of the BioFF posterior, and running of an additional simulation for the newly optimized parameters are iterated until convergence (blue). The result are optimal force field parameter values (red)

$\mathbf{c}_i \approx \mathbf{c}_{i+1}$  are the BioFF optimal parameters and we are finished.

In the presented procedure, we start with an unbiased simulation using the reference force field. One can, however, also use any unbiased or biased simulations and different force fields than the reference force field and apply binless WHAM to calculate the reference weights  $w_\alpha^{(0)}$  in the  $\mathbf{c}_0$ -ensemble.

### 3 Methods

#### 3.1 BioFF implementation

The ultimate goal of full automatization of force field optimization requires careful consideration of steps that would otherwise be handled by experienced practitioners. In particular, when reweighting using Eq. (13) we have to ensure that the reference ensemble covers the reweighted ensemble sufficiently. If the population of structures in the overlap region between reference ensemble and reweighted ensembles is too low, then the reweighed ensemble will suffer from artifacts. As a measure of the coverage between the reference and reweighted ensemble we use Kish’s effective sample size

[40, 41] given by

$$N' = \frac{\left(\sum_{\alpha=1}^N w_\alpha\right)^2}{\sum_{\alpha=1}^N w_\alpha^2} \tag{18}$$

In contrast to Rangan et al. [41], we use the effective sample size  $N'$  itself and not the relative size  $N'/N$  as a measure. Our iterative optimization procedure ensures good coverage of the optimal ensemble even far from the reference ensemble, which grows with each iteration. Consequently, an absolute measure for the quality of sampling is needed because, relatively, the coverage may get smaller with each iteration.

To be able to use common optimization libraries, we include a threshold on the effective sample size  $N'$  given in Eq. (18) in the objective function. To account for the reweighting limit in the objective function used for optimization, we introduce

$$f(N', N^*) = \exp\left[-k \frac{N' - N^*}{N^*}\right] + 1 \tag{19}$$

with a suitable chosen  $N^* > 0$  and  $k > 0$ . This function approaches one for  $N' \gg N^*$ . For  $N'$  approaching  $N^*$ , it diverges exponentially. Defining the objective function for optimization as

$$\mathcal{O}(\mathbf{c}) = \mathcal{L}(\mathbf{c})f(N', N^*) \tag{20}$$

we thus ensure that we do not extrapolate further than supported by the data. Iterating optimization and simulation at the new optimal parameter sets, we make sure that  $\mathcal{O}(\mathbf{c})$  and  $\mathcal{L}(\mathbf{c})$  converge to the same optimum. As a check we could rerun the final optimization for  $\mathcal{O}(\mathbf{c}) = \mathcal{L}(\mathbf{c})$ .

#### 3.2 Polymer model

We use a simple polymer model to illustrate the principles of the BioFF method. The two-dimensional polymer consists of a string of  $n$  beads, with positions  $\mathbf{r}_i$  and  $\mathbf{r}_{i+1}$  of neighboring beads separated by a unit of length. That is,  $|\mathbf{v}_i| = 1$ , where the  $n - 1$  bond vectors are given by  $\mathbf{v}_i = \mathbf{r}_{i+1} - \mathbf{r}_i$ .  $\Delta\phi_i$  is the angle between the two bond vectors  $\mathbf{v}_i$  and  $\mathbf{v}_{i+1}$  such that  $\mathbf{v}_i^\top = (\sin(\phi_i), \cos(\phi_i))$  where  $\phi_i = \sum_{j=1}^i \Delta\phi_j$  and ‘ $\top$ ’ indicates the transpose of the column vector  $\mathbf{v}_i$ .

We introduce a potential energy acting on  $\Delta\phi_i$

$$\beta U(\Delta\phi_i | \mu_i, \kappa_i) = -\ln p(\Delta\phi_i | \mu_i, \kappa_i) \tag{21}$$

such that the Boltzmann distribution corresponds to the von Mises distribution given by

$$p(\Delta\phi_i | \mu_i, \kappa_i) = \frac{e^{\kappa_i \cos(\Delta\phi_i - \mu_i)}}{2\pi I_0(\kappa_i)} \tag{22}$$

where  $I_0(\kappa_i)$  is the modified Bessel function of order zero. The parameters  $\kappa_i > 0$  determine the stiffness of the polymer at positions  $i$ .  $\mu_i$  introduce directional biases.

The probability of a polymer with  $n$  beads is given by

$$p(\Delta\phi|\boldsymbol{\mu}, \boldsymbol{\kappa}) = \prod_{i=1}^{n-1} p(\Delta\phi_i|\mu_i, \kappa_i) = \frac{e^{\sum_{i=1}^{n-1} \kappa_i \cos(\Delta\phi_i - \mu_i)}}{(2\pi)^{n-1} \prod_{i=1}^{n-1} I_0(\kappa_i)} \quad (23)$$

where we introduced  $\Delta\phi = (\Delta\phi_1, \dots, \Delta\phi_{n-1})$ ,  $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_{n-1})$ , and  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{n-1})$ . Without distance-dependent bead interactions, we can efficiently generate uncorrelated polymer conformations by drawing  $n-1$  angles from the von Mises distributions given by Eq. (22). We can include distance-dependent bead interactions and sample conformations using Markov chain Monte Carlo simulations and trial moves in the angles  $\Delta\phi_i$ .

Despite its simplicity, the  $2(n-1)$  parameters of the polymer model facilitate rich structural diversity. Stiff, ordered parts have large  $\kappa_i$ -values. Disordered regions, linkers, and short, flexible hinges, have low  $\kappa_i$ -values. Turns can be introduced as short stretches with large  $\kappa_i$ -values and non-zero  $\mu_i$ -values of the same sign.

The computational efficiency of generating polymer conformations and the simplicity of their visualization in two dimensions make the polymer model an attractive system for method development. We use this polymer model here to sketch the BioFF optimization procedure.

### 3.3 Calculation details

As a simple example, we use the polymer model introduced above with  $n = 100$  beads and  $\kappa_i = \kappa$  and  $\mu_i = 0$  for  $i = 1, \dots, n-1$ . We use the distance between bead 1 and  $n/2 = 50$  as the experimental observable. For validation, we additionally monitor the end-to-end distance given by the distance between bead 1 and  $n$ .

We use the same model to generate synthetic experimental data for a chosen value  $\kappa_{\text{exp}} = 10$  of the single force field parameter by drawing  $N = 10000$  random conformations according to Eq. (23). We calculate the experimental value of the label distance as an average over this ensemble.

For the initial reference simulation, we set the value of our force field parameter, corresponding to  $\mathbf{c}_0$ , to  $\kappa_0 = 20$ . We draw  $N_0 = 4000$  conformations, which form the initial reference ensemble and for which we calculate the label distances. In contrast to the sampling in MD simulations, the structures in the ensemble are strictly uncorrelated. Correlations would have to be taken into account in the WHAM evaluation.

To detect the limit of the reweighting, we set  $N^* = 300$  and  $k = 10$  in Eq. (19). As we will see below, this choice is rather conservative in the example con-

sidered here. Lower values of  $N^*$  or higher values of  $k$  will decrease the number of iterations needed for convergence.

To detect convergence, we demand that for at least two consecutive iteration steps the relative difference between the previous and the actual value of  $\kappa$  is smaller than 0.05, i.e.,  $|\kappa_i - \kappa_{i-1}|/\kappa_i < 0.05$ .

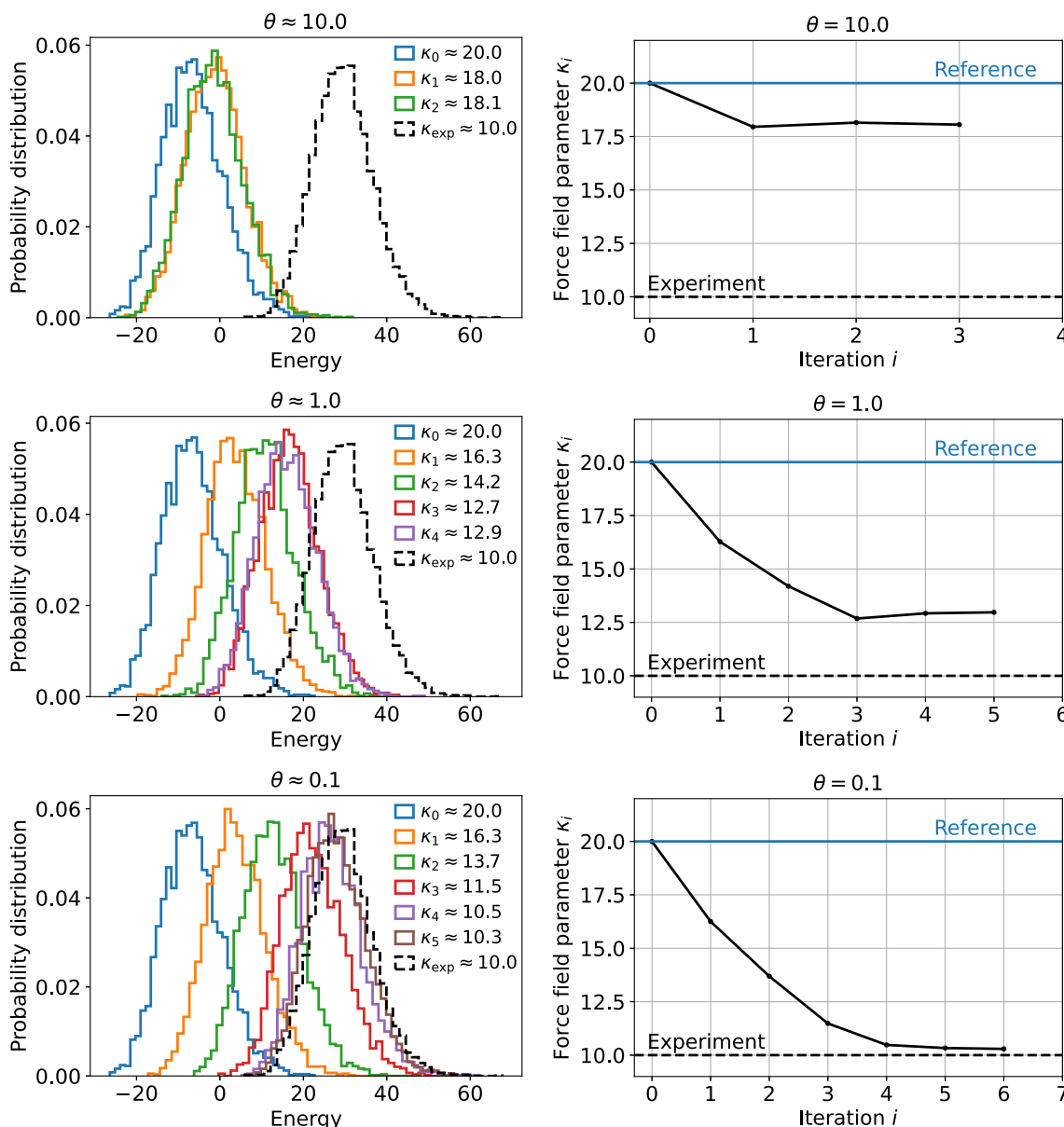
For selected values of the confidence parameter  $\theta = 10, 1$ , and 0.1, we optimize Eq. (20) using the Nelder-Mead method [42] for the reference simulation and  $w_\alpha^{(0)} = 1/N$ , which returns a new value  $\kappa_1$ . For this parameter value, we run a simulation, i.e., we draw  $N_1 = 4000$  new structures, and calculate the label distances for all conformations. We then perform binless WHAM [21, 22, 24] to obtain new values of the reference weights  $w_\alpha^{(0)}$  for the combined  $N = N_0 + N_1$  structures. With this new set of simulations, we again optimize the objective function given by Eq. (20). We iterate this procedure until we detect convergence. Note that we do not use a prior  $p_0(\kappa)$  in this example.

## 4 Results

We obtain convergence for all values of  $\theta$  (Fig. 2). For the smallest value of the confidence parameter  $\theta = 0.1$ , the BioFF optimal value ( $\kappa \approx 10.3$ ) comes close to the value underlying the synthetic experimental data ( $\kappa_{\text{exp}} = 10$ ).

For decreasing values of the confidence parameter  $\theta$ , BioFF optimal distance distributions converge to the experimental distribution (Figs. 3 and 4). Reflecting the fact that BioEn is less restrained in its optimization than BioFF, we find that the calculated average label distance of the BioEn optimal solution is consistently closer to the experimental value than the BioFF solution for the same  $\theta$  value (Fig. 3). However, by correcting for the underlying force field error, BioFF also improves the statistics of the end-to-end distance, which we do not use in the refinement (Fig. 4). By contrast, BioEn gives only small improvements for the end-to-end distance even at large values of  $\theta$ . Indeed, the BioEn-optimal average end-to-end distance is approximately the same for all  $\theta$ -values examined here. By contrast, the BioFF end-to-end distance distribution and their average converge towards the reference with decreasing  $\theta$  even though they are not used in the refinement.

Note that the difference between the BioEn optimal weights and the BioFF optimal weights increases with decreasing values of  $\theta$ . This difference is quantified by the Kullback–Leibler divergence  $S_{\text{KL}}$  between optimal weights and the reference weights (see figure legends in Fig. 3). For  $\theta = 0.1$ ,  $S_{\text{KL}} \approx 0.4$  for the BioEn optimal weights and  $S_{\text{KL}} \approx 14.7$  for the BioFF optimal weights. The reason for this large difference between the  $S_{\text{KL}}$  values is that the label distance distributions of the reference ensemble and the experimental ensemble have a large overlap and thus little BioEn reweighting is needed to obtain agreement. However, the potential energy distributions overlap only very little (Fig. 2),



**Fig. 2** BioFF convergence. (Left) Probability distribution functions of potential energies in the simulations at all steps  $i$  of the iterative optimization. (Right) Convergence of the force field parameter  $\kappa_i$  at iteration steps  $i$ . The prior confidence in the reference ensemble decreases top to bottom ( $\theta = 10, 1, 0.1$ ). The dashed horizontal lines in the panels on the right indicate the value of  $\kappa_{exp}$  used to create the synthetic experimental values of the label distance. Note that  $\kappa_0$  is the reference value and that each  $\kappa_{i>0}$  maximizes the BioFF posterior for the set of  $i$  simulations for parameters  $\kappa_0, \dots, \kappa_{i-1}$

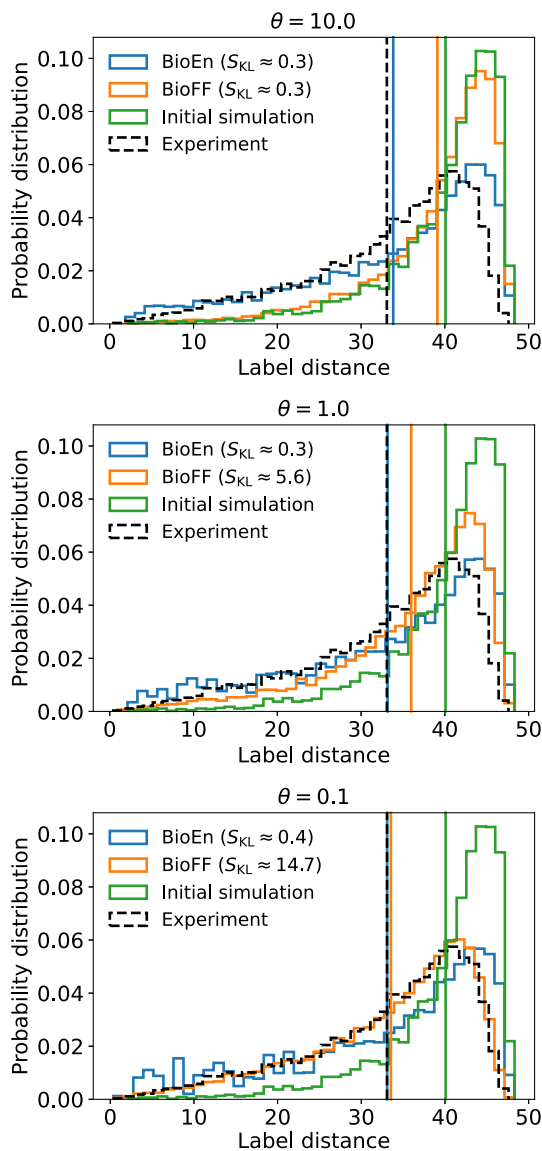
such that the BioFF optimal weights are quite different from the reference weights. By combining the ensembles obtained for different parameters  $\mathbf{c}_i$  with binless WHAM, issues with poor overlap are effectively avoided in the BioFF optimization.

### 5 Discussion

BioFF combines a simplified prior on the force field parameters and an entropic prior on the statistical weights of the sampled conformations. For good param-

eter priors, the entropic prior can be neglected. A good prior properly accounts for the interdependence of the parameters and their uncertainties. For poor parameter priors, the entropic prior takes care that we do not move too far from the reference ensemble. Indeed, the entropic prior prefers parameter values that keep the Kullback–Leibler divergence to the reference ensemble small.

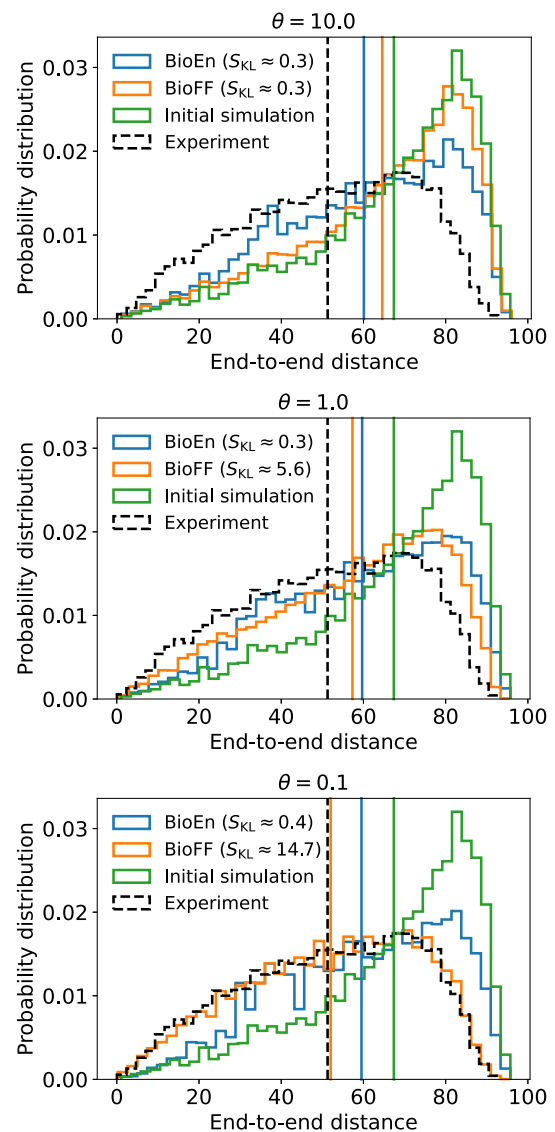
The BioFF method can be viewed as an umbrella of methods distinguished by the prior on the force field parameters and the value of the confidence parameter  $\theta$ . If we set  $\theta = 0$  then the prior on the force field parameters is the only a-priori regularization of the param-



**Fig. 3** Distribution of the label distance used for refinement and its mean value (vertical lines) for the initial reference simulation ( $\kappa_0 = 20$ , green), the synthetic experimental data ( $\kappa_{\text{exp}} = 10$ , black dashed lines), BioEn (blue), and BioFF (orange). The confidence in the reference ensemble decreases top to bottom ( $\theta = 10, 1, 0.1$ )

ter optimization problem. In this case, the prior has to have an appropriate functional form with accurately known uncertainties to properly account for the complex interdependence of parameters. If we use an uninformative prior on the force field parameters and  $\theta > 0$  then the a-priori regularization enters through the functional dependence of the weights on these parameters, Eq. (13), and the Kullback-Leibler divergence. If additionally the potential energy contains terms of the same functional form as the observables, Eq. (7), then the BioEn optimal ensemble and BioFF optimal ensemble are the same [31].

The BioEn optimal ensemble poses a limit for the BioFF optimal ensemble. For the same value of  $\theta$ , BioEn



**Fig. 4** Distribution of the end-to-end distance and mean values (vertical lines) for the initial reference simulation ( $\kappa_0 = 20$ , green), the synthetic experimental data ( $\kappa_{\text{exp}} = 10$ , black dashed lines), BioEn (blue), and BioFF (orange). The end-to-end distance was not used for refinement. The confidence in the reference ensemble decreases top to bottom ( $\theta = 10, 1, 0.1$ )

will essentially always provide a better fit to the experimental data used in refinement while staying closer to the reference ensemble than BioFF. In BioEn, the Kullback–Leibler divergence restrains the optimal statistical weights. In BioFF, the functional dependence of the statistical weights on the force field parameters and the force field parameter prior provide additional restraints. As a consequence, the less restrained BioEn posterior in Eq. (2) evaluated for the BioFF optimal weights is lower than or, at best, equal to that for the BioEn optimal weights.

Typically, one will select certain force field parameters for optimization based on prior knowledge. Experience with the force field and physical insight on



which structural features are probed by the experimental observables guide this process. For example, if an atomistic force field provides an ensemble of intrinsically disordered proteins that is too compact then likely candidates are dihedral angles and the water-protein interaction [43–45].

In a more systematic approach, force field parameters can be identified as candidates for optimization by a sensitivity analysis. Given a set of ensembles corresponding to  $U(\mathbf{x}|\mathbf{c})$ , we can use binless WHAM to estimate the sensitivity of the estimated observable averages  $\langle y_k \rangle_{\mathbf{c}}$  as a function of  $\mathbf{c}$  to the different force field parameters  $c_j$  by

$$J_{kj} = \Delta c_j \frac{(\langle y_k \rangle_{\mathbf{c}} - Y_k)}{\sigma_k^2} \frac{\partial \langle y_k \rangle_{\mathbf{c}}}{\partial c_j} \quad (24)$$

Here,  $\Delta c_j$  is the uncertainty in coefficient  $c_j$  as encoded, e.g., in a Gaussian prior. To linear order, the dimensionless  $|J_{kj}|$  reports on the improvement achievable for observable  $k$  by varying parameter  $j$ . Indeed, we have  $\Delta c_j \partial \chi^2 / \partial c_j = 2 \sum_k J_{kj}$ , i.e., the sum of the  $J_{kj}$  is twice the change in  $\chi^2$  associated with a small change  $\Delta c_j$ . In general, we select parameters that sensitively impact  $\chi^2$ . It remains to be seen to what extent the complexity of the force fields, the large number of their parameters, their poorly quantified couplings, and the importance of solvent-mediated interactions and the required Boltzmann averaging will require fine tuning of the optimization approach and human intervention.

BioFF optimization enables us to learn about the interdependence of parameters. For the BioFF optimal solutions, we can perform a sensitivity analysis by evaluating the Hessian of the log-posterior at the optimum. Performing this analysis for different sets of parameters, we are able to quantify how informative the experimental data are with respect to individual parameters. Importantly, the off-diagonal elements provide information about the interdependence of the parameters. This information is conditioned on the molecular systems we are simulating and the observables that have been measured in experiments and used for parameter refinement.

The reweighting approach in BioFF is more efficient for soft degrees of freedom but it can also handle stiffer degrees of freedom. An example for a soft degree is the dispersion energy parameter of the Lennard-Jones interaction. Adapting this parameter, mainly the depth of the minimum changes. In general, conformations will be part of both the reference and the reweighted ensemble with significant albeit different weights. By contrast, the diameter of a Lennard-Jones particle is a stiff degree of freedom due to the diverging repulsive part of the interaction. A small decrease of the diameter tends to introduce new conformations not previously sampled in the reference simulation. Consequently, reweighting is less efficient in predicting ensembles for decreased diameters. In BioFF, the iteration of parameter optimization for a given ensemble and running new MD simulations for optimized parameters will converge eventually. Con-

vergence can be sped up by softening stiff degrees of freedom, e.g., by replacing the diverging repulsive interaction by a soft-core potential of finite height, and using binless WHAM to obtain the proper reference weights.

## 6 Conclusions and outlook

*Optimized force fields from diverse data and systems.* Force field parameterization is a complex process with broad implications for the MD simulation community. In light of the many possible pitfalls, the optimization has to be performed with great care. Seemingly small corrections of certain force field parameters driven by data for one system may cause unexpected, negative consequences in MD simulations of other systems. In practice, it is therefore important to include a wide range of systems and data in the optimization process and to test the results on systems and data not used in the optimization process.

Covering diverse systems and data requires careful accounting for the respective errors. However, uncertainty quantification is a recognized challenge [9, 46, 47], as we are dealing with errors in experiments, quantum mechanical calculations, statistical sampling, and calculation of observables. The respective errors and error models will have to be carefully and regularly re-assessed to minimize the risk that data with poorly modeled errors skew the optimization process.

In this context, it is also worth noting that coarse-graining may be considered a force field learning and optimization problem [48, 49]. In this vein, Chen et al. used experimental inputs to learn coarse-grained models by constructing a Markov state model and expressing the likelihood as a function of the model parameters [48].

*Force field optimization as a community effort.* The requirement to cover a wide range of systems and data makes force field improvement a community effort. Data and computing resources of individual researchers are necessarily limited. To include as wide a range of inputs and tests as possible, data should be pooled, and optimization and testing tasks should be distributed and ultimately collected. Indeed, ultimate validation happens in the field by scientists applying these force field and checking for consistency.

Practitioners choose force fields depending on the molecular system at hand, their research question, their resources, and their prior experience with molecular force fields. They can compensate for known deficiencies by applying biases during MD simulations or by adapting their analysis, for example. However, for structurally and chemically heterogeneous systems, large system sizes, and long simulation times, it becomes more difficult to find a force field that satisfies all demands. Therefore, the development of computationally efficient force fields that accurately account for a wide range of data for diverse systems becomes critical.

*Systematic and automated force field optimization.* We can only meet the steadily growing demands on the accuracy of molecular force fields by applying systematic and comprehensive approaches to force field parameterization. We envision that in a systematic and (semi)automatic process, force field parameters are validated and updated and their uncertainties are quantified. Experimental data and their uncertainties would be collected in an open data base for wide ranges of methods, systems, and experimental conditions. Importantly, the results of high-level quantum chemical calculations can be readily included in the process, both at the level of priors (e.g., by defining tolerable variations of certain parameters) and of data (by entering into  $\chi^2$ ; see Sect. 2.5). For each of the systems, molecular models are built. Uncertainties in the modeling are accounted for by building multiple models for the same system and weighing them according to prior knowledge in the subsequent refinement. One then runs (un)biased MD simulations for these systems, combines them with previous runs, and quantifies the sampling errors. From these ensembles, one calculates the experimental observables and quantifies the uncertainties in these calculations. By performing Bayesian inference of the parameters, e.g., using BioFF or similar methods, the different sources of errors can be taken into account. Validation of the new parameters involves systems, observables, and data not included in the optimization.

Following this rough outline, resources could be set aside at a super-computing center to regularly perform force field parameterization, validation, and uncertainty quantification. Users can submit experimental data, molecular models, quantum mechanical calculations, and trajectories. With this steadily growing wealth in data, models, and MD simulations one could refine force fields with different emphasis. Specialized force fields can be parameterized for protein folding, liquid-liquid phase separations, protein-ligand binding, and so on. For structurally and chemically heterogeneous systems, as we encounter in (sub)cellular structures, we need to parameterize general-purpose force fields. The comparison of general-purpose force fields with specialized force fields also enables us to identify the limitations stemming from the chosen functional form of the force fields, i.e., model adequacy [9].

*BioFF produces experimentally refined simulation models.* The BioFF approach also addresses a fundamental problem in MD simulations, namely whether to accept or reject the simulation model as a faithful description of the system of interest. Quantitative comparisons of MD simulations to experiment often end in a frustrating experience, in which for a given system some of the observables agree well with experiment, yet others not so well or not at all. One is then left with the decision whether the simulation model should be discarded as a whole (on the basis of partial but significant disagreements) or accepted for further analysis. This decision is a serious one since one of the goals of molecular simulation studies is to gain insight beyond

what can be deduced readily from experiment alone. By providing a means to incorporate the experimental information into the model, frameworks like BioFF allow one to refine the original simulation model “on the fly” to account for the measurements. By providing an experimentally refined model, BioFF, like BioEn, offers a possible route out of the dilemma of accepting or rejecting the original model.

The outcome of BioFF is not only an empirically optimized force field but also an ensemble of molecular conformations that can then be analyzed further. As such, BioFF can be viewed as an enhanced sampling method. Seen from this perspective, the procedure addresses the dual challenges of obtaining better simulation models and of gaining insight into molecular systems on the basis of MD simulations. These simulations capture “all” available experimental information and remain solidly based on physics, which enters, e.g., through quantum mechanical calculations used in the parameterization.

*Improving the functional form of force fields.* BioFF, like many other approaches [5, 27, 31, 38, 39], tries to work around one main challenge in force field refinement: the lack of good force field priors. Ideally, priors not only act on the parameters but also on the functional form of the force fields. Going forward with any of these methods, we should be able to better understand force fields, the interdependence of their parameters and their uncertainties, and the limitations due to their functional forms. This effort should also lead to better priors and methods to develop force fields along the way.

Python 3 source code using Numpy [50], Scipy [51], Matplotlib [52], and Numba [53] and Jupyter notebooks [54] to perform BioFF for the simple example case presented here can be found at <https://github.com/bio-phys/BioFF>. This code uses the open-source BioEn optimization software (<https://github.com/bio-phys/BioEn>) and an open-source binless WHAM implementation (<https://github.com/bio-phys/binless-wham>).

**Acknowledgements** We thank Prof. Kresten Lindorff-Larsen for helpful discussions. This work was supported by the Max Planck Society.

## Author contributions

All authors contributed equally.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Data availability statement** This manuscript has no associated data or the data will not be deposited. [Author’s comment: All data shown and used can be readily generated using the software provided by us.]

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. P. Dauber-Osguthorpe, A.T. Hagler, J. Comput. Aided Mol. Des. **33**(2), 133 (2019)
2. R.B. Best, in *Biomolecular Simulations: Methods and Protocols*, ed. by M. Bonomi, C. Camilloni pp. 3–19(2019) (Springer New York, New York, NY )
3. X. Zhu, P.E.M. Lopes, A.D. MacKerell, WIREs Comput. Mol. Sci. **2**(1), 167 (2012)
4. T. Schlick, S. Portillo-Ledesma, Nat. Comput. Sci. **1**(5), 321 (2021)
5. L.P. Wang, T.J. Martinez, V.S. Pande, J. Phys. Chem. Lett. **5**(11), 1885 (2014)
6. J. Behler, J. Chem. Phys. **145**(17), 170901 (2016)
7. T. Mueller, A. Hernandez, C. Wang, J. Chem. Phys. **152**(5), 050902 (2020)
8. T. Fröhlking, M. Bernetti, N. Calonaci, G. Bussi, J. Chem. Phys. **152**(23), 230902 (2020)
9. F. Cailliez, P. Pernot, F. Rizzi, R. Jones, O. Knio, G. Arampatzis, P. Koumoutsakos, in *Uncertainty Quantification in Multiscale Materials Modeling* (2020), pp. 169–227 Elsevier
10. D. van der Spoel, Curr. Opin. Struct. Biol. **67**, 18 (2021)
11. O.T. Unke, S. Chmiela, H.E. Sauceda, M. Gastegger, I. Poltavsky, K.T. Schütt, A. Tkatchenko, K.R. Müller, Chem. Rev. **121**(16), 10142 (2021)
12. F. Cailliez, P. Pernot, J. Chem. Phys. **134**(5), 054124 (2011)
13. B. Rózycki, Y.C. Kim, G. Hummer, Structure **19**(1), 109 (2011)
14. J.W. Pitera, J.D. Chodera, J. Chem. Theory Comput. **8**(10), 3445 (2012)
15. W. Boomsma, J. Ferkinghoff-Borg, K. Lindorff-Larsen, PLoS Comput. Biol. **10**(2), e1003406 (2014)
16. A.P. Latham, B. Zhang, J. Chem. Theory Comput. **16**(1), 773 (2020)
17. K.A. Beauchamp, V.S. Pande, R. Das, Biophys. J. **106**(6), 1381 (2014)
18. G. Hummer, J. Köfinger, J. Chem. Phys. **143**(24), 243150 (2015)
19. M. Bonomi, C. Camilloni, A. Cavalli, M. Vendruscolo, Sci. Adv. **2**(1), 1 (2016)
20. A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett. **63**(12), 1195 (1989)
21. M. Souaille, B. Roux, Comput. Phys. Commun. **135**(1), 40 (2001)
22. E. Rosta, M. Nowotny, W. Yang, G. Hummer, J. Am. Chem. Soc. **133**(23), 8934 (2011)
23. C.H. Bennett, J. Comp. Phys. **22**(2), 245 (1976)
24. M.R. Shirts, J.D. Chodera, J. Chem. Phys. **129**(12), 124105 (2008)
25. A.B. Norgaard, J. Ferkinghoff-Borg, K. Lindorff-Larsen, Biophys. J. **94**(1), 182 (2008)
26. D.W. Li, R. Brüschweiler, J Chem Theory Comput **7**(6), 1773 (2011)
27. L.P. Wang, K.A. McKiernan, J. Gomes, K.A. Beauchamp, T. Head-Gordon, J.E. Rice, W.C. Swope, T.J. Martínez, V.S. Pande, J. Phys. Chem. B **121**(16), 4023 (2017)
28. S. Bottaro, K. Lindorff-Larsen, Science **361**(6400), 355 (2018)
29. R.B. Best, M. Vendruscolo, J. Am. Chem. Soc. **126**(26), 8090 (2004)
30. J. Köfinger, B. Rózycki, G. Hummer, in *Biomolecular Simulations*, vol., ed. by M. Bonomi, C. Camilloni **2019**, 341–352 (2022). (**Series Title: Methods in Molecular Biology**)
31. A. Cesari, A. Gil-Ley, G. Bussi, J. Chem. Theory Comput. **12**(12), 6192 (2016)
32. A. Cesari, S. Reißer, G. Bussi, Computation **6**(1), 15 (2018)
33. S. Kullback, R.A. Leibler, Ann. Math. Stat. **22**(1), 79 (1951)
34. L.R. Mead, N. Papanicolaou, J. Math. Phys. **25**(8), 2404 (1984)
35. S. Bottaro, G. Bussi, S.D. Kennedy, D.H. Turner, K. Lindorff-Larsen, Sci. Adv. **4**, 5 (2018)
36. J. Köfinger, L. Stelzl, K. Reuter, C. Allande, K. Reichel, G. Hummer, J. Chem. Theory Comput. **15**(5), 3390 (2019)
37. S. Bottaro, T. Bengtson, K. Lindorff-Larsen, in *Methods in Molecular Biology*, ed. by Z. Gáspári vol. 2112, (2020), pp. 219–240
38. A. Cesari, S. Bottaro, K. Lindorff-Larsen, P. Banáš, J. Šponer, G. Bussi, J. Chem. Theory Comput. **15**(6), 3425 (2019)
39. G. Tesei, T.K. Schulze, R. Crehuet, K. Lindorff-Larsen, Proc. Natl. Acad. Sci. U.S.A. **118**(44), e2111696118 (2021)
40. L. Kish, *Survey sampling* (A Wiley Interscience Publication (Wiley, New York, 1995))
41. R. Rangan, M. Bonomi, G.T. Heller, A. Cesari, G. Bussi, M. Vendruscolo, J. Chem. Theory Comput. **14**(12), 6632 (2018)
42. J.A. Nelder, R. Mead, Comput. J. **7**(4), 308 (1965)
43. R.B. Best, W. Zheng, J. Mittal, J. Chem. Theory Comput. **10**(11), 5113 (2014)
44. S. Piana, A.G. Donchev, P. Robustelli, D.E. Shaw, J. Phys. Chem. B **119**(16), 5113 (2015)
45. S. Piana, P. Robustelli, D. Tan, S. Chen, D.E. Shaw, J. Chem. Theory Comput. **16**(4), 2494 (2020)
46. K.K. Irikura, R.D. Johnson, R.N. Kacker, Metrologia **41**(6), 369 (2004)
47. A. Chernatynskiy, S.R. Phillipot, R. LeSar, Annu. Rev. Mater. Res. **43**(1), 157 (2013)
48. J. Chen, J. Chen, G. Pinamonti, C. Clementi, J. Chem. Theory Comput. **14**(7), 3849 (2018)

49. J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N.E. Charon, G. de Fabritiis, F. Noé, C. Clementi, A.C.S. Cent, Sci. **5**(5), 755 (2019)
50. C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T.E. Oliphant, Nature **585**(7825), 357 (2020)
51. SciPy 1.0 Contributors, P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, I. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, Nat. Methods **17**(3), 261 (2020)
52. J.D. Hunter, Comput. Sci. Eng. **9**(3), 90 (2007)
53. S.K. Lam, A. Pitrou, S. Seibert, In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC - LLVM '15* (ACM Press, Austin, Texas, 2015), pp. 1–6
54. K. Thomas, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, J.D. Team, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas, 20th International Conference on Electronic Publishing* (Göttingen, Germany, 2016), ELPUB, pp. 87–90