Abstract

We describe a statistical approach to the validation and improvement of molecular dynamics (MD) simulations of macromolecules. We emphasize the use of MD simulations to calculate thermodynamic quantities which may be compared to experimental measurements, and the use of a common set of energetic parameters across multiple distinct molecules. We briefly review relevant results from the theory of stochastic processes and discuss the monitoring of convergence to equilibrium, the obtaining of confidence intervals for summary statistics corresponding to measured quantities, and an approach to validation and improvement of simulations based on out-of-sample prediction. We apply these methods to replica exchange molecular dynamics (REMD) simulations of a set of eight helical peptides under the AMBER potential using implicit solvent. We evaluate the ability of these simulations to quantitatively reproduce experimental helicity measurements obtained by circular dichroism (CD). We also perform a sensitivity analysis to identify key parameters of the potential, and introduce Bayesian updating of these parameters. We demonstrate the effect of parameter updating applied to the internal dielectric constant parameter on the out-of-sample prediction accuracy as measured by cross-validation.

Running title: Predictive Molecular Dynamics

Keywords: Molecular dynamics, replica exchange, parallel tempering, helical peptides, Bayesian statistics
1 Introduction

Computer simulation, especially molecular dynamics simulation, has become an important and widely used tool in the study of biomolecular systems. With the growing availability of high-speed desktop computers and cluster computing, simulations once requiring access to specialized supercomputers are now within the range of many individual laboratories. Nevertheless, simulation of macromolecules such as proteins and nucleic acids remains a computationally expensive, complicated process with many options and parameters that may significantly affect the results. An important step in the development of standardized simulation approaches for such problems is the study of predictive power - the ability of the simulation to reproducibly generate some externally validatable quantity such as a future experimental measurement. In fields such as physics and chemistry, as well as in macroscopic areas of engineering and astronomy, simulations are regularly used in lieu of physical experiment, due to their ability to accurately and consistently predict physical quantities. Currently macromolecular simulations are primarily used for exploratory and visualization purposes, rather than quantitative prediction. However as computational resources grow and algorithms and theory improve, we can strive to develop truly accurate macromolecular computer experiments.

In order to do so, we must meet several challenges. First, molecular dynamics simulations of macromolecules typically generate configurations on a pico- or femto-second timescale, due to limiting frequencies of bond vibration. Because we cannot experimentally observe the motions of protein atoms at such timescales, the quantities measured in experimental settings can only ever be time-averaged quantities. With the exception of single-molecule experiments, most experimental studies of macromolecules are also ensemble-averaged. Thus we must be concerned primarily with the time- and ensemble-averaged behavior of our simulation model, which may be validated against real experimental observations. Because it cannot be compared against reality, more detailed information generated by a simulation such as specific atomic trajectories or kinetic pathways should be viewed with some skepticism. This distinguishes molecular dynamics simulations of macromolecules from common use of classical mechanics simulations in macroscopic engineering applications, where larger timescales allow for trajectories themselves to be predictively validated against observation.

Thus to compare simulation with experiment requires the computation of ergodic average quantities under our theoretical model (molecular mechanics potential). As is well known, doing so requires adequate exploration of conformational space during the simulation, a difficult problem. However, modern simulation algorithms have made it possible to achieve adequate sampling for small systems. We also require methods for determining when adequate sampling has been achieved. Finally, because sample path averages are only approximate due to finite simulation lengths, we must quantify the remaining uncertainty in computed quantities to properly compare them with experimental values. Some standard statistical methods for addressing these issues are described in Section 2.2.

A grand challenge of macromolecular simulation is the simulation of protein folding. Adequate ensemble-averaging for large proteins remains beyond current computational resources; here we study helical peptide folding, which has been widely studied a model system for protein folding both experimentally (with the large body of subsequent literature summarized elsewhere) and computationally. For short helical peptides, using modern simulation algorithms and a cluster of computer processors, we are able to adequately address the sampling issue. We apply our approach to study eight helical peptides from the experimental literature, and compare results obtained from simulation to experimental data.
A common concern is whether existing forcefields are adequate to simulate protein folding. We approach this question from a predictive perspective: all molecular mechanics potentials are “wrong”, but we can ask if they are “good enough” to accurately predict specified experimental quantities of interest, just as we judge any other theoretical model. Due to the difficulties of comparing simulations against experiment described above, it has previously been very difficult to separate the question of forcefield accuracy from that of adequate configurational sampling. Here we systematically and quantitatively address the latter, enabling us to focus on the former. In particular we ask the question: given reproducible, quantitative predictions of ensemble quantities (here, equilibrium helicities), how well does the forcefield and parameter values used predict experimental quantities (here, CD measurements)?

In taking a predictive perspective, we emphasize the need for a single set of energy parameters which successfully predict experimental quantities of multiple different molecular systems. Recent work has evaluated simulation versus experiment for helicity and thermal melting of single peptides\textsuperscript{11,15}. However, the forcefields and parameter settings chosen for simulation studies often vary significantly across studies, making generalizability difficult to assess: in Section 3.2 we show that the ranges of parameter values used in the literature provide widely different equilibrium values. In addition to choice of parameters, often the forcefield may be modified to improve reproduction of experimental values, an issue we address from a formal statistical perspective in Sections 2.3 and 2.4. A single set of parameters (or well defined criteria for choosing) is critical for prediction of a new molecular system by simulation.

We emphasize that helicity is a coarse-grained measure of the equilibrium ensemble and thus provides only a first-step in evaluating the simulation accuracy; in this manner our approach is meant to be demonstrative rather than exhaustive. However, even by looking only at helicity we obtain important results about reproducibility, parameter sensitivity, and experiment predictive accuracy using a common set of parameters for simulations of multiple distinct peptide systems.

Section 2.1 describes the simulation algorithm (replica-exchange) used in our studies. Section 2.2 describes the statistical tools used to determine when the simulation has converged and to measure the accuracy of quantities computed from the simulation trajectories. Section 3 reports the results obtained from applying our approach to evaluate predictive accuracy for the eight peptides in Table 1. Section 2.3 explores the critical issue of sensitivity of simulation quantities to the parameters of the simulation potential, and demonstrates the use of Bayesian statistics to estimate improved parameter values based on available experimental data.

2 Theory and Methods

We have run extensive molecular dynamics simulations for eight naturally occurring and designed helical peptides which have previously been studied experimentally by circular dichroism (CD) and shown to have measurable helicity (mean \( \theta_{222} \) ellipticity) in solution. Table 1 shows the peptides studied along with their original experimental characterization; these peptides were selected from a database of helical peptides\textsuperscript{8} to obtain a range of helicities among native peptides at physiological pH.

[Table 1 about here.]
2.1 Replica Exchange Molecular Dynamics

All simulations were performed using replica-exchange molecular dynamics (REMD)\textsuperscript{17}, an application of the parallel tempering method\textsuperscript{18} to molecular dynamics (MD) simulation. REMD runs isothermal molecular dynamics simulations in parallel at a ladder of temperatures and attempts to swap chains between temperatures intermittently. Each replica was run under the AMBER94 force field\textsuperscript{19} with a generalized Born model of implicit solvent\textsuperscript{20,21} and a time-step of 2fs. SHAKE\textsuperscript{22} was used (tolerance $5 \times 10^{-5}$ Å) to constrain hydrogen atoms, and a weakly-coupled heat bath with coupling constant of

$$\lambda = 1 + \frac{\Delta t}{2\tau_T}\left(\frac{T_N}{T} - 1\right),$$

is used to maintain constant temperature\textsuperscript{23}, where $T_N$ is the fixed reference temperature and $\tau_T = 1.0$ controls the strength of the coupling. The specific force field, solvent, and heat bath parameters used are given in Table 2, and are an attempt to replicate as closely as possible a protocol which has previously been successful in simulating helical peptide folding\textsuperscript{11,24}. A key question in the wider use of simulation techniques is whether such parameter sets which are successful in one instance are generalizable to other systems; in Section 3 we explore this issue by evaluating the use of these parameters to predict experimentally measured helicities for the eight distinct peptides given in Table 1. A more detailed exploration of the effect of varying these parameter choices is described in Section 2.3.

Our REMD protocol utilizes 30 distinct MD simulations run in parallel at temperatures ranging from the target temperature $T_0$ (273°K, 276°K, 277°K, or 298°K) to $T_{29} = 624°K$ for each peptide simulation. Temperatures are spaced exponentially with $T_i = [T_0 \exp\{ki\}]$, where $k = \ln(624/T_0)/29$ and $i = 0, \ldots, 29$. During the REMD simulation, each replica is run at the assigned temperature for cycles of 1000 MD steps (2ps), after which the translational and rotational motion of the center of mass is removed and 300 temperature-swapping moves attempted, as per a previous protocol\textsuperscript{24}. Let $x_T = (p, q)_T$ denote the coordinates (positions and momenta) of the replicate at temperature $T$. At each temperature-swap, two replicas $x_A$ and $x_B$ are chosen at random and a swap of their respective temperatures $T_A$ and $T_B$ is proposed, with acceptance probability given by the Metropolis criteria:

$$P_{\text{accept}} = \min\left\{1, \frac{\pi(x_B, T_A)\pi(x_A, T_B)}{\pi(x_A, T_A)\pi(x_B, T_B)} \right\}$$

where $\pi(x, T) \propto \exp[-E(x)/k_BT]$ is the Boltzmann distribution over configurations at temperature $T$, and $E$ is the total energy $E(x) = U(x) + \Delta G_{\text{solv}} + \frac{1}{2} \sum_i \|p_i\|^2/m_i$ with $U$ the potential function given by (4) and $\Delta G_{\text{solv}}$ the implicit solvent free energy term given by (5). When a swap is accepted the two replicas exchange temperatures, otherwise they remain at their respective temperatures. Associated velocities are rescaled to reflect the temperature swap before the next cycle of MD steps. This process of 1000 MD steps followed by 300 attempted temperature swaps is repeated until the convergence criteria described in Section 2.2 is reached.

The above REMD protocol is used to conform as closely as possible to existing uses of REMD in protein simulation in the literature. The Metropolis criteria is used to guarantee invariance of (and therefore convergence to) the Boltzmann ensemble; however, recent theoretical analysis\textsuperscript{25} shows that corrections are needed to guarantee the proper invariant measure.
2.2 Statistical Analysis of Simulation Output

As described above, molecular dynamics simulations of molecules differ somewhat from the use of classical mechanics simulations in macroscopic engineering applications, since detailed comparison of dynamical trajectories to experimental data is typically impossible. In fact such trajectories are highly sensitive to starting conditions, parameterizations of the energy model, and other simulation details. Instead, it is the long-run time-averaged behavior of the simulation which we can expect to produce observable macroscopic (thermodynamic) physical quantities if the simulation model is adequate. To evaluate simulations against experimental data then, we must be able to accurately compute the long-run time-averaged behavior implied by our theoretical model, specified by the molecular forcefield or potential. To do so, we rely on two important results. First, an ergodic theorem saying that if the dynamics of our simulation are ergodic (able to reach any region of the configuration space from any other region), then the time-averaged behavior of the simulation will converge to the configuration space integral representing the ensemble-averaged behavior for any (integrable) quantity of interest. Writing the quantity of interest as a function \( h(x) \) of configurations \( x \) in configuration space \( X \), we have:

\[
\lim_{t \to \infty} \frac{1}{t} \int_{0}^{t} h(x(s)) \, ds = Z^{-1} \int h(x)e^{-\frac{1}{kT}E(x)} \, dx \overset{def}{=} \langle h \rangle_T
\]

(1)

for the canonical (constant N,V,T) ensemble, where \( \langle h \rangle \) denotes the expectation or ensemble average of \( h(x) \) under the stationary Boltzmann distribution. Here \( h(x) \) is any quantity we wish to compute from a given configuration, and may be used to compute means (e.g. internal energy or helicity), variance-covariance matrices (for essential dynamics), indicator functions (for free energies), and so on. A major advantage of simulation-based methods is the ability to calculate a variety of such quantities from a single simulation. The right-hand integral yields the ensemble averaged quantity under the theoretical model (forcefield); it is this quantity that can be compared with real world experiments which are themselves averaged over both time and molecules in solution.

Because we can not run simulations infinitely long, we can only ever compute an approximation to the left-hand side of (1). So to use this result in practice, we need to know two things: how long must the simulation be run such that this approximation is “pretty good” (the convergence in (1) is approximately achieved), and how good is “pretty good” (error bounds on the computed quantities). Not only must the simulations have reached equilibrium, but they must have run in equilibrium long enough to produce accurate approximations of the time/ensemble averaged quantities of interest. From this perspective, MD simulation is simply a tool for computing the integral (1), and often alternative numerical integration methods such as Monte Carlo sampling or replica-exchange dynamics may be more efficient than standard MD at this task. However these methods often disrupt the kinetics of the process; interestingly, recently developed simulation methods which do not not guarantee proper ensemble sampling may be useful in taking ensemble samples generated by methods such as MC or REMD and reconstructing the kinetics\(^\text{27}\). The second important result is statistical, and provides guidance on these questions. It says that for well-behaved functions \( h(x) \) (those having finite variance under Boltzmann distribution \( \pi(x) \)), and under stronger assumptions on the simulation dynamics (a sufficient condition is geometric ergodicity\(^\text{28}\)) which are difficult to verify in practice, the time-average of \( h \) computed from a simulation of \( N \) steps converges to the true value \( \langle h \rangle \) as \( N \to \infty \). Moreover, this sample path average obeys a central limit theorem, converging in
distribution to a normal random variable centered at the true value $\langle h \rangle$:

$$\hat{h} = \frac{1}{N} \sum_{t} h(x^{(t)}) \xrightarrow{d} N(\langle h \rangle, \sigma_{h}^{2})$$

where $\sigma_{h}^{2} = \sigma_{h}^{2} + 2 \int_{t} \rho(t)dt$ (2)

where $\sigma_{h}^{2} = \langle h^2 \rangle - \langle h \rangle^2$ is the variance of $h(x)$ under the Boltzmann distribution $\pi(x)$, and $\rho(t) = \langle (h(x_{i0}) - h(x_{i0+t}))^2 \rangle / \sigma_{h}^{2}$ is the autocorrelation function for fluctuations in $h$ of configurations at time separation $t$ when the process is in equilibrium. Note that the Metropolis step in REMD creates a stochastic process, so we state results in those terms; central limit theorems for deterministic ergodic dynamical systems exist but are somewhat more delicate. In any case, determinism of MD for molecules in solution is artificial and often replaced with stochastic (Langevin or Brownian) dynamics anyway; in the stochastic case ergodicity of the system is not an assumption but can be shown directly.

This theoretical result has important implications. It provides the distribution of errors obtained when we use the time average from a finite length simulation to approximate the theoretical ensemble average. This allows us to quantify uncertainty and produce error bars based on (100-$\alpha$)% confidence intervals, which is critical for comparing the simulation output with experimental data. This in turn allows us to determine simulation time needed in order to approximate quantities to a predetermined level of accuracy. Failure to run a simulation long enough to adequately estimate quantities of interest is a common pitfall of molecular dynamics simulation.

In addition, knowledge that the errors are approximately normally distributed allows us to treat the simulation model as a (rather complicated) statistical model, and perform likelihood-based statistical inference on the simulation parameters, as described in Section 2.3.

### 2.2.1 Interval predictions

A critical aspect of comparing simulation output with experiment is to account for the inherent variability of both the simulation output and the experimental measurement. As described above, variability in the simulation output can be characterized by a central limit theorem: the quantity $\hat{h}$ approaches $\langle h \rangle$ in the limit of large $N$, with error $\langle h \rangle - \hat{h}$ begin normally distributed with variance $\sigma_{h}^{2}$ given by (2). This result allows us to construct normal-based confidence intervals for $h$ of the form $\langle h \rangle \pm 2\sigma_{h}$. The variance of $\hat{h}$ therefore determines how long we need to run a given simulation to obtain a predetermined level of accuracy. Since $\sigma_{h}$ depends on the function $h$ of interest, some quantities can converge significantly faster than others, a fact observed empirically; however apparent convergence of some quantities while others have not converged can also be misleading. Theoretical guarantees on how long a simulation must be run are extremely difficult to come by, although recent progress has been made in this area for parallel tempering algorithms.

To determine this interval we require an estimate $\hat{\sigma}_{h}$ for $\sigma_{h}$. Direct estimates of the summed autocorrelation (2) are inconsistent, but several other estimation methods exist. A common and relatively straightforward technique, which we use here, is the batch estimate, obtained by dividing the simulation of length $N$ into $a = N/M$ regions or batches of size $M$. Each batch is used to independently estimate $\langle h \rangle$:

$$\hat{h}_{i} = \frac{1}{M} \sum_{k=iM}^{(i+1)M} h(x_{k})$$

$$\hat{h} = \frac{1}{a} \sum_{i} \hat{h}_{i}$$

and $M$ is chosen large enough to ensure the autocorrelation $\rho_{\hat{h}_{i},\hat{h}_{i+1}} \approx 0$. The batch estimates are then...
approximately independent samples whose empirical variance

\[ \hat{\sigma}_h^2 = \frac{1}{a-1} \sum_i (\hat{h}_i - \bar{h})^2 \]

yields a simple estimate of the variance \( \sigma_h^2 \). The quantile plot in Figure 4a indicates approximate normality is a reasonable assumption for our converged simulations.

### 2.2.2 Monitoring convergence

The energy surface of proteins and polypeptides is characterized by large energy barriers and multiple local minima, making adequate exploration of configuration space a major challenge of protein simulation. While theoretical guarantees are very difficult to obtain for complex simulations, and observing the output of a simulation can never guarantee convergence, convergence diagnostics can be constructed to identify lack of convergence from simulation output. Our preferred approach is the use of multiple parallel simulations starting from diverse initial conditions in order to monitor the convergence by comparison of sample path quantities across distinct simulations. We use the multiple-chain approach\(^{33}\) to assess convergence of our simulations by running multiple independent REMD simulations for each peptide in parallel starting from a diverse set of initial configurations, with each individual REMD simulation run according to the protocol of Section 2. Let \( M \) denote the number of simulations and \( x^{(i)}_j \) for \( j = 1, \ldots, M \) the configuration of the \( j^{\text{th}} \) simulation at time step \( i \). Convergence of an observable quantity \( h(x) \) is by calculating

\[
B_N = \frac{1}{M} \sum_{j=1}^{M} (\hat{h}_j - \bar{h})^2 \quad \text{and} \quad W_N = \frac{1}{NM} \sum_{j=1}^{M} \sum_{i=1}^{N} (h(x^{(i)}_j) - \hat{h}_j)^2,
\]

with \( \bar{h} = \frac{1}{N} \sum_{i=1}^{N} h^{(i)}_j \) and \( \hat{h} = \frac{1}{M} \sum_{j=1}^{M} \bar{h}_j \). \( B_N \) represents the between-chain variability and \( W_N \) represents the within-chain variability. When multiple starting configurations are chosen to be widely dispersed throughout configuration space, early in the simulation the chains will be sampling distinct regions of phase space and the between-chain variance will be significantly higher than the within-chain. As the simulations converge to sampling from the same equilibrium Boltzmann distribution, these two quantities will converge. Comparison is based on techniques from the analysis of variance to determine whether significant differences remain, and convergence is monitored using the Gelman-Rubin shrink factor

\[
\sqrt{R_N} = \sqrt{\frac{N-1}{N} + \frac{M+1}{M} B_N \frac{v_N}{W_N v_N - 2}}
\]

where \( v_N = 2(\hat{\sigma}_N^2 + \frac{B_N}{M})^2 / W_N \). The quantity \( \sqrt{R_N} \) estimates the reduction in variance of the estimator \( \hat{h} \) if the simulation were to be run infinitely long, and converges to one as all of the parallel simulations converge to equilibrium.

Once the chains have equilibrated, samples from all \( M = 4 \) independent simulations can be combined to obtain a pooled estimate of \( \langle h \rangle \), with individual chain estimates combined inversely proportional their respective variances:

\[
\hat{h} = \frac{\sum_j \hat{h}_j \hat{\sigma}_{h_j}^{-2}}{\sum_j \hat{\sigma}_{h_j}^{-2}} \quad \text{and} \quad \hat{\sigma}_h^2 = \left( \sum_j \hat{\sigma}_{h_j}^{-2} \right)^{-1}
\]

Thus the effective trajectory length of the combined estimate is \( M \bar{N} \) where \( \bar{N} \) is the average production phase length; the only price paid for using multiple simulations compared to a single simulation is the
replication of the equilibration phase. In our opinion, the advantage of being able to run in parallel and to obtain convergence diagnostics by inter-run comparisons far outweighs this cost in most situations. Note that combining the results of multiple simulations who have not been determined to have individually converged to the same stationary distribution, as is sometimes done in MD simulation, has no theoretical justification and can be badly misleading.

Numerous other convergence diagnostics have been developed in the statistics and operations research literature\textsuperscript{34}, including further developments of the approach used here\textsuperscript{35,36}. Note that no diagnostic based on simulation output can ever guarantee convergence, all such dianostics can be fooled\textsuperscript{34}. However theoretical bounds on simulation time are very difficult to obtain; although relevant work in this direction is ongoing\textsuperscript{31}.

### 2.3 Parameter Adaptation

The energetics used in molecular dynamics simulations involve a large number of parameters which must be specified in advance. These include the parameters of the AMBER potential\textsuperscript{19}, given by the covalently-bonded and non-bonded terms:

\[
U(x) = \sum_{\text{bonds}} K_r (r-r_{eq})^2 + \sum_{\text{angles}} K_{\theta} (\theta-\theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_\varphi}{2} [1 + \cos(n\varphi - \gamma)]
\]

\[
+ \sum_{i<j} [A_{ij} \frac{r_{ij}}{r_{ij}^n} - B_{ij} \frac{r_{ij}^n}{r_{ij}^{n+1}} + q_i q_j \epsilon r_{ij}] + \sum_{i<j\in\Omega_{1-4}} \left[ \frac{1}{S_{\text{nb}}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \frac{1}{S_{\text{ee}}} \frac{q_i q_j}{r_{ij}} \right]
\]

where $\Omega_{1-4}$ is the set of atom pairs $(i, j)$ which are separated by three bonds. For example, the parameter $S_{\text{ee}}$ weights the electrostatic interactions in $\Omega_{1-4}$, $S_{\text{nb}}$ weights the corresponding van der Waals interactions, and the dielectric constant $\epsilon$ affects all longer range electrostatic terms.

In addition, the implicit-solvent model given by the generalized Born approximation\textsuperscript{20,21} has associated parameters:

\[
\Delta G_{\text{solv}} = \sum_{a \in A} \beta_a \alpha_a(x) + \Delta G_{\text{pol}}.
\]

where $A$ is the set of atom types, $\alpha_a$ is the total solvent-accessible surface area of atoms of type $a$ in configuration $x$, $\beta_a$ are solvation parameters, and the electrostatic polarization component of the free energy of solvation is given by:

\[
\Delta G_{\text{pol}} = -\frac{1}{2} \sum_{i,j} \left( \frac{1}{\epsilon_{\text{mol}}} - \frac{1}{\epsilon_{\text{water}}} \right) \frac{q_i q_j}{f_{\text{GB}}(r_{ij})},
\]

which involves parameters such as the intra-molecular dielectric constant $\epsilon_{\text{mol}}$ and the solvent dielectric $\epsilon_{\text{water}}$. (In AMBER, these parameters are specified as $\epsilon_{\text{in}}, \epsilon_{\text{ext}},$ and $\epsilon_{\text{dielec}},$ with $\epsilon_{\text{water}} = \epsilon_{\text{ext}} \epsilon_{\text{dielec}}$, $\epsilon_{\text{mol}} = \epsilon_{\text{in}} \epsilon_{\text{dielec}},$ and $\epsilon = \epsilon_{\text{in}} \epsilon_{\text{dielec}}$.) Although in principal these parameters represent physical quantities whose values can be known, in practice they are approximations with values determined individually in empirical or theoretical studies.

The simulations of Section 3 utilize a default set of parameters given in Table 2, chosen to comply with standard practice as described in Section 2. Nevertheless, there is significant variation in the literature in values chosen for some of these parameters. Since the ensemble simulated is defined by these parameters, the simulation averages obtained and their comparison with experimental values will be a function of these parameter choices. It is therefore important to understand the how differences in these parameter values...
may be propagated into the resulting thermodynamic quantities estimated, and to determine the impact on the conclusions obtained. Sensitivity analysis of these parameters is described in Section 3.2.

2.4 Bayesian estimation of forcefield parameters

Given the sensitivity of simulation results to certain forcefield parameters as demonstrated in Section 3.2, we would like to identify a standard set of values that could be used by various researchers to ensure consistency and comparability of simulations across different studies. A natural approach to determine such values is to optimize the parameter values with reference to experimental data. However, it is important to do so in such a way that the resulting parameter values are generalizable to other systems. A criticism commonly leveled at simulation research is that with the large number of parameters involved in specifying a potential energy function, a solvation model, and a simulation algorithm, the simulation may be adjusted to produce almost any behavior the investigator desires.

Such concerns can be addressed by standardized use of a common set of parameters, but the adaptation of these parameters to better match experimental observations remains important. The danger is that optimizing parameters on a specific set of data may provide good results, but generalize poorly to the study of other systems, a phenomenon known as overfitting. One solution to overfitting is the use of large datasets relative to the number of parameters, where the simultaneous adaptation to multiple experimental measurements ensures that no particular measurements are well-described at the expense of others. In fact we have large quantities of experimental helicity data available, but we are currently limited by the fact that each data point requires parallel simulations at nanosecond or greater timescales for inclusion. However overfitting is also avoided by a variety of parameter estimation and predictive validation techniques developed in statistics, such as Bayesian analysis and regularization. These techniques, which penalize large parameter changes when insufficient data is available to justify them, regularly allow the adaptation of complex many-parameter models to relatively small data sets while avoiding overfitting and producing parameters which generalize well. In this paper we adopt a Bayesian approach, using prior information to adapt the parameters by Bayesian inference. Because computational considerations limit us to the use of the eight peptides in Table 1, we perform only a small example of this approach, adapting only one parameter at a time.

To evaluate generalizability of the parameters resulting from this adaptation approach, we apply the statistical method of cross-validation. This allows us to estimate out-of-sample prediction accuracy, or how accurately we can expect these parameters to perform when simulating a new peptide to predict its experimental helicity.

We first specify a simple statistical error model for the experimental data, which says that the measured helicities may be described as a combination of the theoretical equilibrium helicity under our forcefield, plus some experimental noise:

\[ h_{\text{exp}}^R = \langle h_R \rangle_\theta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2) \] (7)

Here \( h_{\text{exp}}^R \) denotes the experimentally measured helicity of peptide \( R \), and we now denote explicitly the dependence of \( \langle h \rangle \) on the peptide sequence \( R \) as well as the forcefield parameters \( \theta \). If \( \langle h_R \rangle_\theta \) was a linear function of the parameters \( \theta \), then (7) would simply be a linear regression model. Instead, the ensemble helicity \( \langle h_R \rangle_\theta \) is a complicated function given by the configuration integral under the Boltzmann distribution with potential function (4) parameterized by \( \theta \). But similar statistical principles apply, allowing us to estimate the parameters \( \theta \) from data, with a slightly unique aspect arising from the difficulty in calculating \( \langle h_R \rangle_\theta \).
which can only be done approximately by the simulation average from a finite length simulation as described in Section 2.2. The assumption of normally distributed noise can be justified by the previously described central limit theorem for \( \langle \hat{h}_R \rangle \) as well as standard usage for experimental noise; the quantile plot in Figure 4b shows that this assumption is quite reasonable.

The Bayesian approach next specifies a prior distribution for the parameter \( P(\theta) \) which captures any background or biophysical knowledge we may have about the parameter, to supplement the information contained in the experimental data. We then base our inference about the parameter on the posterior distribution:

\[
P(\theta \mid \text{Data}) = \frac{P(\text{Data} \mid \theta) P(\theta)}{\int P(\text{Data} \mid \theta') P(\theta') d\theta'} \propto P(\theta) \prod_{i=1}^{p} \phi \left( \frac{\hat{h}_i^\theta - h_i^{\exp}}{\sigma_{\exp} + \sigma_{\hat{h}_i^\theta}} \right) \tag{8}
\]

where \( p \) is the number of peptides, \( \phi \) is standard normal density function, and \( \hat{h}_i^\theta \) is the simulated helicity (3) for peptide \( i \) run at parameter value \( \theta \). Note that \( \sigma_{\exp}^2 \) reflects the variance in \( h_i^{\exp} \) arising from experimental noise, and \( \sigma_{\hat{h}_i^\theta} \) is the remaining simulation uncertainty in \( \hat{h}_i^\theta \) given in (3). We use the estimated value \( \sigma_{\exp} = .07 \) for convenience; more formally \( \sigma_{\exp} \) could be estimated or integrated out to obtain the marginal posterior distribution.

3 Results

3.1 Predicting peptide helicity by simulation

REMD simulations were performed for the eight peptides given in Table 1. For each peptide, four REMD simulations were run in parallel, with each REMD simulation utilizing 30 temperatures according to the protocol of Section 2. Initial configurations for the four REMD runs were generated as follows for each peptide: one ideal helix; one extended conformation; and two random configurations, one generated by uniformly sampling \((\phi, \psi)\) angles within the helical range and one generated by uniformly sampling \((\phi, \psi)\) outside of the helical range. Figure 1(a) shows the starting configurations for a particular peptide at \( T_0 \). Initial velocities were generated randomly and independently for each configuration. Quantities monitored for convergence included backbone \( \phi \) and \( \psi \) angles of each amino acid, helicity of the peptide, and total energy. Convergence to equilibrium was declared when the Gelman-Rubin shrink factor for these quantities reached 1.1, and the sample paths up to this time (equilibration phase) discarded. Sample paths from this time on (production phase) were included in computing time-averaged quantities \( \hat{h} \). Figure 2 shows plots of the Gelman-Rubin shrink factor for simulations of the eight peptides; for comparison a standard MD simulation (without replica-exchange) of one of the peptides is shown.

[Figure 1 about here.]

[Figure 2 about here.]
Helicity of a peptide configuration was defined as the fraction of amino acid \((\phi, \psi)\) pairs lying in a predefined helical range with the potential to form hydrogen bonds:

\[
h(x) = \frac{1}{(l-2)} \sum_{i=1}^{l} \prod_{j=i-1}^{i+1} 1(\phi_{i+1} \leq \phi(x_j) \leq \phi_{hi}) 1(\psi_{i+1} \leq \psi(x_j) \leq \psi_{hi})
\]  

for configuration \(x\) of a peptide of length \(l\), where \(1(\cdot)\) is an indicator function. We use a standard range for defining helical angles: \(\{\phi_{lo}, \phi_{hi}\} = \{-87, -27\}\) and \(\{\psi_{lo}, \psi_{hi}\} = \{-77, -17\}\); see Section 3.3.1 for the effect of changing these boundaries on the resulting helicities.

The total simulation time required to reach convergence for each peptide is shown in Table 3. Following equilibration, each simulation was continued until the estimated variance \(\hat{\sigma}^2_h\) of the combined estimate of equilibrium helicity \((3)\) decreased below 0.001. The total simulation times required in the production phase to meet this criteria are also shown in Table 3.

[Table 3 about here.]

Figure 3 shows the simulated equilibrium helicities versus the published experimental helicities in Table 1. All experimental helicities are derived from mean residue \(\theta_{222}\) ellipticity measured by circular dichroism. Simulated helicities are shown with 95% confidence intervals obtained as described in Section 2.2. For each peptide, we show the helicity interval obtained from each of the four independent REMD runs as well as the combined estimate \(\hat{h}\) (3). Simulated helicities are correlated with the experimental helicities but are not within perfect agreement even within sampling error of the simulations. Recent estimates place the standard deviation of experimental noise to be around .05, so the simulations may agree within the tolerance of combined noise due to finite simulation sampling and experimental error.

[Figure 3 about here.]

[Figure 4 about here.]

### 3.2 Sensitivity Analysis

To address the questions of sensitivity of simulation results to choice of simulation parameters described in Section 2.3, we performed a 1-way sensitivity analysis of six of the forcefield parameters described there: the external and internal dielectrics \(\epsilon_{ext}\) and \(\epsilon_{in}\), the salt concentration constant used with implicit solvent, the weight terms \(S_{ce}\) and \(S_{nb}\), for atom pairs in \(\Omega_{1-4}\), and the dielectric constant \(\epsilon_{dielec}\), in addition to a simulation parameter: the non-bonded cutoff distance. Ranges for these parameters were chosen based on the variation in use in published simulation studies and recommended values in forcefield documentation.

Sensitivity analysis was performed via short REMD simulations of two peptides (DGAEAAAKAAAGR and SAEDAMRTAGGA) starting from equilibrium states obtained from the longer simulations of Section 3. Parameters given in Table 2 were used as reference values, and each parameter was varied individually while holding the others constant, performing short REMD simulations of 200ps for each peptide at 273K. Four copies of each were run from different equilibrium starting configurations to monitor convergence as described in Section 2.2.2. The resulting sensitivity of helicity to perturbations of these seven parameters is shown in Table 4. Of the parameters examined, the internal dielectric constant \(\epsilon_{in}\) has the most dramatic effect on the helicity obtained from simulation.
These results in Table 4 suggest that the variability among choices for both $\epsilon_{in}$ and $S_{nb}$ observed in the literature may significantly impact the thermodynamic quantities measured from these simulations. In order to evaluate the ability of this potential and solvent model to reproduce experimental helicities, we must obtain an appropriate consistent value for this and other parameters.

3.2.1 Effect of $\epsilon_{in}$ and $S_{nb}$ Parameters

Variation of the two forcefield parameters $\epsilon_{in}$ and $S_{nb}$ showed a significant effect impact on the helicity of the two peptides studied in the one-way sensitivity analysis. We focus on determining an appropriate value for $\epsilon_{in}$ in Section 3.3, but first we explore here the manner in which these parameters affect helicity. $\epsilon_{in}$ affects both the solute-solvent electrostatic polarization term in (6) and non-bonded electrostatic interactions in (4). The $\Delta G_{pol}$ term represents a difference in electrostatic interaction energy resulting from solvent screening of charges. As $\epsilon_{in}$ increases towards $\epsilon_{ext}$ this difference shrinks, effectively increasing internal charge screening by making the interior of the molecule more polar. The electrostatic interactions in (4) also decrease as $\epsilon_{in}$ increases, reducing the favorability of hydrogen bonds formed in helix formation. Thus we expect that increasing $\epsilon_{in}$ will produce lower simulation helicity levels, as observed in Table 4. (Sensitivity to much larger changes in the structure of the solvation model has been previously reported\textsuperscript{38,24}, but our results show that even for a given solvation model the choice of parameter values may have a large impact.)

In contrast, Table 4 shows that as $S_{nb}$ increases, peptide helicity decreases. $S_{nb}$ scales the non-bonded Van der Waals interactions in the potential energy $U$ (4). To interpret the effect of this parameter on helicity we examined its effect on each of the 1-4 interactions along the peptide backbone, as well as the relation of these 1-4 distances to amino acid helicity. The effect of the $S_{nb}$ parameter on most 1-4 interactions had little effect on helicity, with the notable exceptions of nitrogen-to-nitrogen (N-N) and hydrogen-to-carbon (H-C) distances, as pictured in Figure 5.

Equilibrium values of both of these distances (N-N and H-C) decrease as $S_{nb}$ increases, which in turn lowers the helicity of the associated amino acid as shown in Figure 6 in both peptides for which this sensitivity analysis was performed. Changes in other 1-4 atom pair distances induced by $S_{nb}$ increase had little effect on helicity; $C_\alpha-C_\alpha$ is shown as an example.

3.3 Bayesian estimation of internal dielectric $\epsilon_{in}$

To demonstrate the approach to parameter adaptation described in Section 2.4, we applied it to estimate the parameter $\epsilon_{in}$ which was shown in Table 4 to have the most significant impact on equilibrium helicity. To estimate an optimal value for internal dielectric we discretized this parameter into a set of plausible values $\epsilon_{in} \in \{1, 2, 3, 4, 5\}$ spanning the range of values which have been used previously in the literature\textsuperscript{24,39}. A non-informative uniform prior distribution was assigned for $\epsilon_{in}$ by giving each possible value of $\epsilon_{in}$ equal prior probability. To obtain the posterior distribution (8) for $\epsilon_{in}$, REMD simulations as described in Section 2.1 were then run for each peptide at each discrete value of the dielectric constant with all other parameters...
fixed at their default values given in Table 2. The resulting posterior distribution over the discrete values of $\epsilon_{in}$ is shown in Figure 7a.

Given this shape, we decided to refine the discretization of $\epsilon_{in}$ by adding another value at $\epsilon_{in} = 4.1$ to obtain a more detailed look at the posterior in the high probability region. (Each such point requires 8 peptides $\times$ 4 REMD runs $\times$ 30 MD simulations run to convergence to obtain equilibrium helicities, hence the coarseness of the original discretization.) The new posterior is shown in Figure 7b.

Figure 8 shows estimated helicities obtain by simulation versus experimental values for all eight peptides for the six different values of $\epsilon_{in}$, where it can be seen that the highest posterior probability value of 4.1 is the one which gives the closest approximation to the set of experimental values. Figure 9 plots the simulated helicity versus the experimental helicity individually for the eight peptides plotted at the different values of $\epsilon_{in}$. Figure 7b shows that resulting posterior distribution is clearly peaked at 4.1. (Note that further refinement of the discretization may well lead to an improvement between 4.0-4.1 or 4.1-5.) As more experimental data and associated simulations are included, this posterior distribution can be updated to reflect the information in the larger data set, further refining the optimal value.

3.3.1 Helical backbone angles

As a second example, we consider the boundaries of the helical $(\phi, \psi)$ region which define a “helical” backbone conformation in Equation 9. Although a general region may be defined based on Ramachandran plots to be approximately $\phi \in \{\phi_{\min}, \phi_{\max}\} = \{-87, -27\}$ and $\psi \in \{\psi_{\min}, \psi_{\max}\} = \{-77, -17\}$, the exact region measured by CD at ellipticity $\theta_{225}$ is somewhat ambiguous. We may view the boundaries of this region to be parameters of the statistical mechanical model and estimate them by Bayesian inference as above. In this case evaluation of the posterior at a range of $\phi_{\max}^H, \phi_{\min}^H, \psi_{\max}^H, \psi_{\min}^H$ may be done more easily than for $\epsilon_{in}$ since these are parameters of the statistical mechanical model for helicity but not of the forcefield which determines the simulation ensemble; thus we need simply re-analyze the trajectories rather than re-simulate for each value.

For simplicity, we again discretize and construct a four dimensional grid for possible values of $(\phi_{\max}^H, \phi_{\min}^H, \psi_{\max}^H, \psi_{\min}^H)$. The marginal posterior distributions obtained for $(\phi_{\min}^H, \phi_{\max}^H)$ under a uniform prior are shown in Figure 10a, and for $(\psi_{\min}^H, \psi_{\max}^H)$ in Figure 10b.

Peaks representing high probability values of $(\phi_{\min}, \phi_{\max})$ are seen at $(-80, -50)$ and $(-90, -40)$, with a ridge for $\phi_{\min}$ between $-100$ and $-70$. The joint distribution for $(\psi_{\min}, \psi_{\max})$ exhibits a sharp peak at $(-60, -40)$ and a minor peak at $(-50, -30)$ The ranges of dihedral angles with the largest peaks for both $\phi$ and $\psi$ contain the values for an ideal helix $(\phi, \psi) = (-57, -47)$, but the joint mode of $(\phi_{\min}, \phi_{\max}, \psi_{\min}, \psi_{\max}) = (-90, -40, -60, -40)$ yields a narrower range than that generally accepted for helical angles $(-57 \pm 30, -47 \pm 30)$. Ridges are centered near ideal values of -57 for $\phi$ and -47 for $\psi$, further increasing the probability in these regions. The ridge suggests that the precise value of $\phi_{\max}$ is poorly identified, likely due to peptide
backbone geometry where for \( \phi \) values below -100 and \( \psi \) values above -40, backbone steric clashes prevent configurations being sampled at all.

3.4 Cross Validation

In order to run further simulations, we must choose a particular value for the internal dielectric parameter; a natural choice is the mode of the posterior distribution at \( \epsilon_{in} = 4.1 \). This value gives the best agreement between the experimental helicities and the simulated helicities for our observed peptides.

However we wish to know how sensitive this result is to the particular set of peptides chosen, and thus how well we can expect this choice to generalize to accurately simulate the helicity of new peptides outside our data set. Simply taking the accuracy of \( \epsilon_{in} = 4.1 \) in predicting these eight peptides will tend to overestimate this accuracy, because this value has been optimized to perform well on those peptides. Nevertheless, we can estimate the future (out-of-sample) predictive accuracy from the current set of peptides using the statistical method of cross-validation. We measure predictive accuracy via the mean squared error (MSE) between the predicted and experimentally measure helicity values.

Cross-validation proceeds by removing one peptide (say the \( i^{th} \) one) from the dataset, and using the other seven to estimate/optimize the parameter \( \epsilon_{in} \). Denote the resulting parameter value by \( \hat{\epsilon}_{[-i]} \). We then use this value to simulate the removed peptide and predict its helicity, calculating the squared error between predicted and experimental values. This process is then repeated to obtain similar predicted accuracies for each of the peptides in turn, always using the parameter value optimized over the other seven peptides, to calculate the overall estimated predictive accuracy:

\[
\text{MSE}_{CV} = \frac{1}{p} \sum_{i=1}^{p} (h_{i}^{exp} - \hat{h}_{i}^{\epsilon_{[-i]}})^2
\]

This procedure has well established properties as an unbiased estimator of the out-of-sample predictive accuracy of our parameter adaptation method\(^{40}\). Because the vast majority of our computational work is done upfront in running the simulations of each peptide at each value of \( \epsilon_{in} \), the expense of calculating the cross-validated prediction accuracy is negligible.

Table 5 shows the MSE for each value of \( \epsilon_{in} \) and the cross-validated MSE; in this case the cross-validated MSE is equal to the MSE for \( \epsilon_{in} = 4.1 \) since each \( \hat{\epsilon}_{[-i]} \) was equal to 4.1.

4 Conclusions

Macromolecular simulation is becoming a widely used tool in structural and molecular biology. As usage grows and computational resources continue to accelerate, the development of true macromolecular computer experiments which can accurately and reproducibly calculate thermodynamic or kinetic quantities which agree with experiment is within sight. For researchers developing or utilizing macromolecular simulation it is an exciting time.

Here we have attempted to focus attention on use of simulation in this quantitative, prediction fashion. We have described several statistical methods useful for addressing the challenges in doing so: quantitative measures of simulation convergence, construction of uncertainty intervals for simulated quantities, Bayesian
and shrinkage estimation for parameter adaptation, and the use of cross-validation to evaluate predictive accuracy. We have also demonstrated the use of this approach to evaluating and improving molecular dynamics simulations of helical peptides, and explored the sensitivity of such simulations to small changes in parameter values. The tools described here are broadly applicable and we hope they will be adopted by other researchers and will help encourage further progress towards quantitative, predictive simulations of macromolecular systems.

The results described here are only a first step and may be improved in a number of ways. We have used relatively small amounts of data representing only equilibrium helicity in evaluating the predictive accuracy of our simulations, and more sophisticated comparisons to detailed experimental data will add significantly more information to evaluate and improve ensemble properties of simulations. Interesting recent examples have been applied to short timescale peptide kinetics\textsuperscript{41,42} and comparison with NMR\textsuperscript{43,44} measurements. Even when considering equilibrium helicity, equation (9) is crude and may be more accurate if replaced by a calculation of ellipticity\textsuperscript{45} or an entire CD spectrum for comparing with CD measurements; however, calculation of CD spectra from configurations is itself difficult. In any case, in the face of more detailed experimental information the methods described in the current paper become even more important to enable reliable, quantitative comparisons and for improving forcefield models on multiple systems simultaneously to avoid overfitting and improve predictive accuracy.

Similarly, the purpose of a molecular simulation is rarely to predict only a single quantity such as helicity, and when such predictions are desired statistical models may often be developed which are significantly more accurate across a wider range of input molecules\textsuperscript{8}. The advantage of a simulation is the ability to examine many different ensemble quantities calculated from a single simulation output. Nevertheless, predictive evaluation of measured quantities will help improve the underlying forcefields and algorithms, thus improving the accuracy, relevance of, and confidence in other quantities obtained from simulation output.

Acknowledgments

SCS and BC were supported in part by NSF grant DMS-0204690 (SCS). Computing resources were provided by NSF infrastructure grant DMS-0112340 (SCS).

References


List of Figures

1 Three configuration snapshots from the four parallel REMD simulations of peptide SAEDAMRTAGGA. Shown are (a) the four starting configurations, (b) four configurations observed at time of convergence to equilibrium, and (c) four configurations from the experiment phase of the simulation. .................................................. 20

2 Convergence of REMD simulations of the eight peptides from Table 1, as measured by the Gelman-Rubin shrink factor for helical content. Plots represent convergence between 4 parallel simulations started from diverse initial configurations using parameters given in Table 2. Shown for comparison is a convergence plot for 4 standard MD simulations of peptide AE simulated without replica-exchange. .............................................. 21

3 Peptide helicity as estimated from simulation versus experimentally measured helicity for the eight peptides in Table 1. The diagonal line $y = x$ is shown as a reference. Simulation results are shown as 95% confidence intervals using standard errors estimates described in text, and are shown both for individual REMD runs and for the pooled estimates. .................. 22

4 Quantile plots of standardized residuals (a) $(\hat{h}_{ij} - \bar{h}_i)/\sigma_{\hat{h}_{ij}}$ for the $8 \times 4 = 32$ individual REMD simulations, and (b) $(\hat{h}_i - \hat{h}_i^{exp})/\sigma_{\hat{h}_i}$ for the eight combined simulation peptide helicities versus experiment. The lack of significant deviation from the diagonal suggests the assumptions of normally distributed noise is reasonable in each case. .................. 23

5 The N-N (blue) and H-C (red) 1-4 interactions along the peptide backbone, which are most affected by changes in the $S_{nb}$ scaling constant in the AMBER potential. The effect of equilibrium distances for these atom pairs has a significant effect on the $(\phi, \psi)$ angles of their respective amino acids, and hence on peptide helicity. .................. 24

6 Effects of the non-bonded scaling parameter $S_{nb}$ on the equilibrium distances of successive backbone nitrogen atoms (N-N), hydrogen-carbon atom pairs (H-C), and $\alpha$-carbons ($C_\alpha-C_\alpha$). Line represents the ensemble-mean helicity for the $i^{th}$ amino acid as a function of the $N_i-N_{i+1}$ distance (plots a,b), $H_i-C_i$ distance (c,d), or $C_\alpha-C_\alpha$ distance (e,f) for the two peptides DGAEAAKAAGAR (a,c,e) and SAEDAMRTAGGA (b,d,f). Individually labeled points give the average N-N or H-C distance for simulations with $S_{nb} = \{0.5, 1, 1.2, 1.5, 2, 3, 5\}$. Helicity changes in response to varying $S_{nb}$ can be explained by sensitivity to N-N and H-C distances; other 1-4 atom pairs have little effect on helicity as demonstrated here for $C_\alpha-C_\alpha$.................. 25

7 Posterior distributions for the dielectric constant $\epsilon_{in}$ evaluated at discrete values, obtained using Bayesian parameter updating described in Section 2.4 under uniform prior. (a) Posterior over $\epsilon_{in} \in \{1, 2, 3, 4, 5\}$. (b) An additional simulation was run at $\epsilon_{in} = 4.1$ to help identify the mode of the posterior distribution. .................. 26

8 Simulated helicity versus experimental helicity for the peptides in Table 1 evaluated at a range of values of the internal dielectric parameter $\epsilon_{in}$. .................. 27

9 Helicity versus $\epsilon_{in}$ for each peptide in Table 1 at values of $\epsilon_{in} \in \{1, 2, 3, 4, 1.4, 5\}$. The experimentally measured helicity for each peptide is plotted as a horizontal line. .................. 28

10 Marginal posterior distributions of boundaries of the helical angle region (a) $\phi_{min}$ and $\phi_{max}$, and (b) $\psi_{min}$ and $\psi_{max}$. .................. 29
Figure 1: Three configuration snapshots from the four parallel REMD simulations of peptide SAEDAM-RTAGGA. Shown are (a) the four starting configurations, (b) four configurations observed at time of convergence to equilibrium, and (c) four configurations from the experiment phase of the simulation.
Figure 2: Convergence of REMD simulations of the eight peptides from Table 1, as measured by the Gelman-Rubin shrink factor for helical content. Plots represent convergence between 4 parallel simulations started from diverse initial configurations using parameters given in Table 2. Shown for comparison is a convergence plot for 4 standard MD simulations of peptide AE simulated without replica-exchange.
Figure 3: Peptide helicity as estimated from simulation versus experimentally measured helicity for the eight peptides in Table 1. The diagonal line $y = x$ is shown as a reference. Simulation results are shown as 95% confidence intervals using standard errors estimates described in text, and are shown both for individual REMD runs and for the pooled estimates.
Figure 4: Quantile plots of standardized residuals (a) $(\hat{h}_{ij} - \bar{h}_i) / \sigma_{\hat{h}_{ij}}$ for the $8 \times 4 = 32$ individual REMD simulations, and (b) $(\hat{h}_i - h_i^{\text{exp}}) / \sigma_{\hat{h}_i}$ for the eight combined simulation peptide helicities versus experiment. The lack of significant deviation from the diagonal suggests the assumptions of normally distributed noise is reasonable in each case.
Figure 5: The N-N (blue) and H-C (red) 1-4 interactions along the peptide backbone, which are most affected by changes in the $S_{ab}$ scaling constant in the AMBER potential. The effect of equilibrium distances for these atom pairs has a significant effect on the $(\phi, \psi)$ angles of their respective amino acids, and hence on peptide helicity.
Figure 6: Effects of the non-bonded scaling parameter $S_{nb}$ on the equilibrium distances of successive backbone nitrogen atoms (N-N), hydrogen-carbon atom pairs (H-C), and $\alpha$-carbons ($C_\alpha-C_\alpha$). Line represents the ensemble-mean helicity for the $i^{th}$ amino acid as a function of the $N_i-N_{i+1}$ distance (plots a,b), $H_i-C_i$ distance (c,d), or $C_\alpha-C_\alpha$ distance (e,f) for the two peptides DGAEAAKAAAGR (a,c,e) and SAEDAMRTAGGA (b,d,f). Individually labeled points give the average N-N or H-C distance for simulations with $S_{nb} = \{0.5, 1, 1.2, 1.5, 2, 3, 5\}$. Helicity changes in response to varying $S_{nb}$ can be explained by sensitivity to N-N and H-C distances; other 1-4 atom pairs have little effect on helicity as demonstrated here for $C_\alpha-C_\alpha$. 
Figure 7: Posterior distributions for the dielectric constant $\epsilon_{in}$ evaluated at discrete values, obtained using Bayesian parameter updating described in Section 2.4 under uniform prior. (a) Posterior over $\epsilon_{in} \in \{1, 2, 3, 4, 5\}$. (b) An additional simulation was run at $\epsilon_{in} = 4.1$ to help identify the mode of the posterior distribution.
Figure 8: Simulated helicity versus experimental helicity for the peptides in Table 1 evaluated at a range of values of the internal dielectric parameter $\epsilon_{in}$. 
Figure 9: Helicity versus $\epsilon_{in}$ for each peptide in Table 1 at values of $\epsilon_{in} \in \{1, 2, 3, 4, 1.5\}$. The experimentally measured helicity for each peptide is plotted as a horizontal line.
Figure 10: Marginal posterior distributions of boundaries of the helical angle region (a) $\phi_{\text{min}}$ and $\phi_{\text{max}}$, and (b) $\psi_{\text{min}}$ and $\psi_{\text{max}}$. 
List of Tables

1 Helical peptides studied by simulation in this paper, along with original experimental characterization and conditions. Peptides are either unblocked or have an N-terminal acetyl group (Ace) and/or a C-terminal amide group (Nhe). ID provides the peptide identifier used in other figures in this paper. ................................................................. 31

2 Forcefield and simulation parameters used in the helical peptide replica-exchange simulations, and as default values for the parameter sensitivity analysis. Parameters values are those used previously for simulating a helical peptide $^{24}$. ................................................................. 32

3 Time length of equilibration and production phases for REMD simulations of each peptide; peptide identifiers are given in Table 1. Equilibration and production times were determined according to the statistical convergence criteria described in Section 2.2. Due to the use of replica-exchange, equilibration is significantly faster than physical timescales (see Figure 2). . 33

4 One-way sensitivity analysis of helicity as a function of simulation parameters. Shown are mean helicity and variance obtained for two peptides DGAEAAKAAAGR and SAEDAM-RTAGGA at a range of values for each parameter studied. ................................................................. 34

5 Mean-squared error (MSE) for each value of $\epsilon_{in}$, along with estimated out-of-sample prediction accuracy given by MSE obtained from cross-validation. ................................................................. 35
<table>
<thead>
<tr>
<th>ID</th>
<th>N-</th>
<th>Peptide Sequence</th>
<th>C-</th>
<th>Experimental Helicity</th>
<th>Temp (°K)</th>
<th>pH</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DG</td>
<td>-</td>
<td>DGAEAAKAAAGR</td>
<td>Nhe</td>
<td>0.196</td>
<td>273</td>
<td>7</td>
<td>46</td>
</tr>
<tr>
<td>SA</td>
<td>Ace</td>
<td>SAEDAMRTAGGA</td>
<td>-</td>
<td>0.168</td>
<td>273</td>
<td>7</td>
<td>47</td>
</tr>
<tr>
<td>RD</td>
<td>-</td>
<td>RDGWKRLIDIL</td>
<td>-</td>
<td>0.050</td>
<td>277</td>
<td>7</td>
<td>48</td>
</tr>
<tr>
<td>ES</td>
<td>-</td>
<td>ESSLERITRKL</td>
<td>-</td>
<td>0.217</td>
<td>277</td>
<td>7</td>
<td>48</td>
</tr>
<tr>
<td>LK</td>
<td>Ace</td>
<td>LKEDIDAFLAGGA</td>
<td>Nhe</td>
<td>0.150</td>
<td>298</td>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>PS</td>
<td>Ace</td>
<td>PSVRKYAREKGV</td>
<td>Nhe</td>
<td>0.097</td>
<td>298</td>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>RE</td>
<td>Ace</td>
<td>REKGVDIRLVQG</td>
<td>Nhe</td>
<td>0.134</td>
<td>298</td>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>AE</td>
<td>Ace</td>
<td>AETAGAKFLRAHA</td>
<td>Nhe</td>
<td>0.126</td>
<td>276</td>
<td>7</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 1: Helical peptides studied by simulation in this paper, along with original experimental characterization and conditions. Peptides are either unblocked or have an N-terminal acetyl group (Ace) and/or a C-terminal amide group (Nhe). ID provides the peptide identifier used in other figures in this paper.
<table>
<thead>
<tr>
<th>Simulation parameter</th>
<th>$\epsilon_{ext}$</th>
<th>$\epsilon_{in}$</th>
<th>Salt concentration</th>
<th>Non-bonded cut-off</th>
<th>$S_{ee}$</th>
<th>$S_{ab}$</th>
<th>$\epsilon_{dielec}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default value</td>
<td>78.5</td>
<td>4.0</td>
<td>0.0M</td>
<td>8.0Å</td>
<td>1.2</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 2: Forcefield and simulation parameters used in the helical peptide replica-exchange simulations, and as default values for the parameter sensitivity analysis. Parameters values are those used previously for simulating a helical peptide$^{24}$.  


<table>
<thead>
<tr>
<th>Peptide</th>
<th>DG</th>
<th>SA</th>
<th>RD</th>
<th>ES</th>
<th>LK</th>
<th>PS</th>
<th>RE</th>
<th>AE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equilibration phase (ns)</td>
<td>0.8</td>
<td>0.6</td>
<td>0.8</td>
<td>0.8</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.4</td>
</tr>
<tr>
<td>Production phase (ns)</td>
<td>1.0</td>
<td>0.8</td>
<td>1.4</td>
<td>1.6</td>
<td>2.4</td>
<td>2.4</td>
<td>1.6</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Table 3: Time length of equilibration and production phases for REMD simulations of each peptide; peptide identifiers are given in Table 1. Equilibration and production times were determined according to the statistical convergence criteria described in Section 2.2. Due to the use of replica-exchange, equilibration is significantly faster than physical timescales (see Figure 2).
Table 4: One-way sensitivity analysis of helicity as a function of simulation parameters. Shown are mean helicity and variance obtained for two peptides DGAEAAKAAAAGR and SAEDAMRRTAGGA at a range of values for each parameter studied.
Table 5: Mean-squared error (MSE) for each value of $\epsilon_{in}$, along with estimated out-of-sample prediction accuracy given by MSE obtained from cross-validation.

<table>
<thead>
<tr>
<th>$\epsilon_{in}$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>4.1</th>
<th>5</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.0405</td>
<td>0.0120</td>
<td>0.0025</td>
<td>0.0016</td>
<td>0.0013</td>
<td>0.0025</td>
<td>0.0013</td>
</tr>
</tbody>
</table>