# Robust conformational space exploration of cyclic peptides by combining different MD protocols and force fields

Samuel Murail,† Jaysen Sawmynaden,‡ Akli Zemirli,† Maud Jusot,‡ Fabio Pietrucci,‡ Jacques Chomilier,‡ Pierre Tufféry,† and Dirk Stratmann*,†,‡

†*Université Paris Cité, CNRS UMR 8251, INSERM ERL U1133, Unité de Biologie Fonctionnelle et Adaptative, BFA, F-75013 Paris, France*

‡*Sorbonne Université, Faculté des Sciences et Ingénierie, UFR 925, MNHN, UMR CNRS 7590, Institut de Minéralogie de Physique des Matériaux et de Cosmochimie, IMPMC, F-75005 Paris, France*

*corresponding author: dirk.stratmann@sorbonne-universite.fr

March 31, 2025

## Abstract

Cyclic peptides are an important class of pharmaceutical drugs. We used here replica-exchange molecular dynamics (REMD) and simulated tempering (ST) simulations to explore the conformational landscape of a set of nine cyclic peptides. The N-ter to C-ter backbone cyclized peptides of seven to ten residues were previously designed for a high conformational stability with a mixture of L and D amino acids. Their experimental NMR structures were available in the protein data bank (PDB). For each peptide we tested several force fields, Amber96, Amber14, RSFF2C and Charmm36 in implicit and explicit solvent. We find that the variability of the free energy maps obtained from several protocols is larger than the variability obtained by just repeating the same protocol. Running multiple protocols is therefore important for the convergence assessment of REMD or ST simulations. The majority of the free energy maps showed clusters with a high RMSD to the native structures, revealing the residual flexibility of this set of cyclic peptides. The high RMSD clusters had in some cases the lowest free energy, rendering the prediction of the native structure more difficult with a single protocol.

Fortunately, the combination of four implicit solvent REMD and ST simulations, mixing the Amber96 and Amber14 force fields, predicted robustly the native structure.

As implicit solvent in the REMD or ST setup are up to one hundred times faster than explicit solvent simulations, so running four implicit solvent simulations is a good practical choice. We checked that the use of an explicit solvent REMD or ST simulation, taken alone or combined with implicit solvent simulations, did not significantly improved our results. It results that our combination of four implicit solvent simulations is tied in terms of success rate with much more expensive combinations that include explicit solvent simulations. This may be used as guideline for further studies of cyclic peptide conformations.

## Introduction

Peptides have attracted increasing attention over the past decade as a viable alternative to small molecules for developing drug molecules capable of interfering with protein-protein interactions particularly well.[3–7] While linear peptides composed of only natural amino acids suffer from degradation, cyclic peptides display a stronger resistance to the proteases.[8] In addition, their conformational space is usually more restricted than that of linear peptides,[9] reducing the entropy loss upon binding, and leading to a stronger binding affinity.[10,11] Cyclisation is usually achieved by a peptide bond between the N-ter and the C-ter amino acid or by disulfide bonds between two cysteine residues. Combining both approaches allows the design of hyperstable cyclic peptides.[12] Conformational stability can also be obtained by a combination of L and D-amino acids,[13] without the need for disulfide bonds at least for small cyclic peptides up to about ten residues.[2] Combining the cyclisation with chemical modifications, like non-canonical amino acids[14,15] or N-methylation, can increase the specificity to the target[16] and reduce the conformation flexibility.[17] Cyclic peptides that incorporate chemical modifications are named *macrocycles*, they represent a class of drugs on their own right,[18] with an important field of research. The chemical space available to macrocycles is huge and as for small molecules, virtual libraries of billions of compounds have been developed.[19] To interfere with intracellular targets, drugs must pass the cell membrane. Fortunately, cyclic peptides can pass this barrier, as documented in the recent database CycPeptMPDB gathering experimental membrane permeability measurements of over 7000 cyclic peptides, of which more than the half present a high membrane permeability,[20] and *in-vitro* workflows have been developed recently for generating orally available cyclic peptides.[21] Overall, there is a bright future for cyclic peptides and macrocycles in therapeutic drug development[22] with an average of about one cyclic peptide drug approved per year.[7] The numerous possibilities for in-vitro and in-silico design approaches for peptide therapeutics in general are summarized in several recent reviews.[23–29]

While cyclic peptides are less flexible than linear peptides, some degree of flexibility remains depending on the length and the composition of the amino acid sequence, as well as the use of chemical modifications.[17,30] In a drug design context, it is important to be able to access the entire conformational landscape accessible to a cyclic peptide, for two main reasons: First, a poorly flexible cyclic peptide will likely lose less conformational entropy upon binding than a more flexible peptide.[10,11] To quantify this, the accessible conformational space can be a possible indicator. Second, a flexible

Table 1: Cyclic peptide data set. All structures of the peptides in free form (i.e. not bound to a protein) have been solved by NMR.. Upper case letters are L-amino acids and lower case letters are D-amino acids. The all L-amino acids peptide 8C was published in Ref.,[1] whereas the other peptides were published in Ref.[2] As reference structure we used the best representative conformer of the NMR ensemble which is here the first model for all nine PDB files. The conformational variability among the NMR ensemble is quantified with the average RMSD of the 19 models compared to the first model.

| PDB code | Peptide | sequence | atoms | NMR ensemble RMSD [Å] |
|---|---|---|---|---|
| 6BE9 | 7.A | TkNDTnp | 104 | 0.9 ± 0.38 |
| 6BEW | 7.B | hPdqseP | 100 | 0.74 ± 0.22 |
| 6BF3 | 7.C | QDPpKtd | 105 | 0.45 ± 0.087 |
| 6BE7 | 8.A | DDPTprQq | 124 | 0.56 ± 0.15 |
| 6BEN | 8.B | rQpqRePQ | 142 | 0.44 ± 0.077 |
| 6WPV | 8.C | GTVAVQFL | 119 | 0.23 ± 0.056 |
| 6BEO | 9.A | pPYhPKDLq | 150 | 0.51 ± 0.2 |
| 6BEQ | 10.A | AARvpRltPE | 160 | 0.65 ± 0.18 |
| 6BER | 10.B | EvDPehpNap | 141 | 0.8 ± 0.18 |

cyclic peptide has usually a different conformation in free form than in its bound form inside the protein-peptide complex, due to the additional protein-peptide interactions.[31] For an optimal binding affinity, it is important that the bound form is already explored inside the conformational ensemble of the free form (*conformational selection*) or that the free form can be converted into the bound conformation (*induced fit*) without having to cross a too high free energy barrier.[32] For some applications, a high flexibility is needed. For example, it is sometimes necessary to travel "a tunnel" to the binding site. So cyclic peptides cannot always be designed as "rocks", but have to incorporate an appropriate level of flexibility. Experimental structure determination methods like NMR or X-ray crystallography will resolve the lowest free energy structure, and perhaps a second minor conformation, but not the whole conformational space and the free energy barriers among several meta-stable conformations. Therefore conformational space explorations are usually done with in-silico methods.

The conformational space explored by cyclic peptides can be characterized *in-silico* with molecular dynamics (reviewed in[33]), Monte-Carlo simulations and other sampling methods. The Rosetta software suite[34] uses Monte-Carlo simulations and kinematic closure algorithms to sample cyclic peptides conformations.[12] A similar kinematic closure algorithm has been implemented in EGSCyP (Exhaustive Grid Search for Cyclic Peptides)[35] that uses an exhaustive exploration for cyclic pentapeptides composed of standard, D-amino acid and N-methylated amino acids. Conventional molecular dynamics simulations (cMD) are in general not sufficient to pass the high free energy barriers that separate different conformations of cyclic peptides. Therefore, enhanced sampling molecular dynamics simulations are used, like replica-exchange MD,[35–43] bias-exchanged metadynamics (BE-MetaD).[41,42,44–53] Machine learning can even accelerate further the exploration of the conformational space of cyclic peptides.[54,55] AlphaFold2[56] has also been applied to cyclic peptides,[57–59] but it is limited by its design to the prediction of the 3D structure and not to the identification of a conformational ensemble. Whereas AlphaFold2 is limited to natural amino acids, the recent AlphaFold3[60] may be used to predict the structure of cyclic peptides that incorporate modified amino acids, as it can be applied to any small molecule in general, but this remains the subject for furhter investigations.

In the present study, we assess the performance and convergence of several force fields in molecular dynamics (MD) simulations for the exploration of the conformational space of a set of nine cyclic peptides, ranging from seven to ten amino acids. We run in parallel two enhanced sampling methods, replica exchange molecular dynamics (REMD) and simulated tempering (ST), to exclude bias from either sampling method.

A very recent study[61] of the Yu-Shan Lin group tested several force fields on a set of 12 smaller cyclic peptides, ranging from five to seven amino acids. The smaller size stabilized the eight sequences without D-amino acids. The remaining four pentapeptides had one or two D-amino acids. They performed BE-MetaD simulations in explicit solvent for most of the force fields tested. In our study we focused on implicit solvent simulations to assess if this kind of solvent model is sufficient for the elucidation of the conformational space of free-form cyclic peptides. We focused also on medium sized cyclic peptides, as the conformational space can be more difficult to be resolved. Therefore our study is complementary to the similar study[61] of the Yu-Shan Lin group.

## Materials & Methods

### Cyclic peptide dataset

Most cyclic peptides present in the Protein Data Bank (PDB) are bound with other molecules or have chemical modification (like N-methyl). In the present study we choose a set of nine cyclic peptides available in free form in the PDB and having a medium size from seven to ten amino acids. All but one are peptides with a mixture of L and D amino acids, optimized for conformational stability by the David Baker group using Rosetta.[2] We employ here the same peptide names as in their paper, like "7.B" for a peptide that has seven amino acids and that is the second design of this size (see table 1). The peptide 8.C in table 1 is not present in the study of Hosseinzadeh et al. It is composed of standard amino acids only.[1] We added it to see if a cyclic peptide of this size can be as stable as the peptides that are stabilized by D-amino acids and prolines.

### Implicit solvent REMD protocol

Our REMD protocol is similar as the one we used in a former study on cyclic pentapeptides[35] and inspired from the protocol presented in the work of Wakefield et al.[40]

Table 2: List of the eight combinations of sampling protocol, force field and solvent model that have been run in the present study. Each combination was run on the nine peptides from table 1. The trajectory length for REMD is 1$\mu$s and for ST 10$\mu$s. The cumulated trajectory time is the product of this trajectory length with the number of replica and with the number of repeated runs. The total cumulated trajectory time of the implicit and explicit solvent runs is 1440$\mu$s and 378$\mu$s, respectively.

| Protocol | Force Field | Solvent | replica | runs | Cumulated trajectory time [$\mu$s] |
|---|---|---|---|---|---|
| REMD | Amber96 | implicit | 8 | 5 | 360 |
| REMD | Amber14 | implicit | 8 | 5 | 360 |
| REMD | RSFF2C | implicit | 8 | 5 | 360 |
| REMD | RSFF2C | explicit | 32 | 1 | 288 |
| ST | Amber96 | implicit | 1 | 2 | 180 |
| ST | Amber14 | implicit | 1 | 1 | 90 |
| ST | Amber14 | explicit | 1 | 1 | 90 |
| ST | Charmm36 | implicit | 1 | 1 | 90 |

The starting structures of the cyclic peptides were constructed manually with ChimeraX.[62] Topology files were generated with Ambertools[63] and converted in GROMACS format with acpype.[64]

Temperature REMD simulations were setup with eight replica with OBC (Onufriev, Bashford, and Case)[65] GBSA implicit solvent and run with GROMACS 5.1.2,[66] more recent GROMACS versions do not include implicit solvents any more. A short minimization was first done with an alternation of one step of steepest descent every 500 conjugate gradient steps. The maximum number of steps was set to 50,000 with a step size of 0.01 nm and an energy convergence criterion of 10 kJ/(mol*nm). The thermalizations and simulations were done in a temperature range from 300K to 455K. The temperature values of the eight replica were chosen to keep the probabilities of accepting exchanges the same (300, 318, 337.97, 358.81, 380.85, 404.27, 429.12 and 455.50K). The acceptance ratio was high, as 55 to 65% of all exchange attempts were successful, thanks to the small number of atoms (100 to 150). Only a short phase of thermalization in NVT was realized during 100 ps (50,000 steps, 2 fs time step). The modified Berendsen thermostat was employed with the time constant $\tau = 0.1$ ps.

A good sampling in REMD depends on the time step between exchange attempts between neighbouring replicas.[67] No consensus choices exist for this parameter. The time step used in the literature is between 0.01 to 100 ps[36 68] and it depends on the system and the process studied. Here we chose 1 ps as time step between exchange attempts. For each force field and peptide we produced trajectories of 1 $\mu$s with a 2 fs time step for each replica, i.e. 8 $\mu$s for each REMD simulation. The atomic coordinates were recorded every 10 ps. We repeated each REMD simulation five times (run 1 to 5) to access convergence. In total we produced in this study with the REMD protocol 8$\mu$s * 5 runs * 9 peptides * 3 force fields = 1080 $\mu$s of cumulative trajectory time (see table 2).

To produce 1 $\mu s$ per replica the cost is between 1000 to 2000 CPU-hours per REMD simulation. So in total we consumed about 5 runs * 9 peptides * 3 force fields * 1000-2000 CPU-hours = 135k-270k CPU-hours, which is quite reasonable thanks to the use of an implicit solvent.

Three force fields were tested with the REMD protocol: Amber96,[69] Amber14SB[70] and the coil-library-based Residue Specific Force Field RSFF2C[71] which is derived from the Amber14SB force field and has shown better performance for peptides,[61] proteins and loops.[72] We included the older Amber96 force field, as we employed it in our former study[35] and as it showed still a good performance on cyclic peptides.

## Explicit solvent REMD protocol

We setup also an explicit solvent REMD protocol with the RSFF2C force field, which has been tested only in explicit solvent in previous studies. Using the standard TIP3P explicit solvent model, a cubic water box with a minimum distance between the peptide and the box walls of 1.0 nm was equilibrated with GROMACS 2018.8.[66] In the case that the peptide has not a zero net electric charge, the total charge of the peptide + water system has been neutralized by adding sodium or chloride ions to the solvent. The energy of the whole system was then minimized with a steepest descent algorithm and each replica was equilibrated to its temperature in two steps, first with a simulation of 100 ps in the NVT ensemble followed by a simulation of 200 ps in the NPT ensemble. The time step of the equilibration and production simulations was set to 2 fs. A non-bonded cutoff of 1.0 nm was used with the Particle Mesh Ewald (PME) algorithm for long range electrostatic interactions. The modified Berendsen thermostat was employed with the time constant $\tau = 0.1$ ps and the Parrinello-Rahman pressure coupling in the NPT ensemble with the time constant $\tau = 2.0$ ps.

We employed 32 replica, as this value fitted well with our computer architecture. The temperature values were generated with the web-server[73] using the algorithm published in Ref.[74] using the same temperature range as for the implicit solvent protocol (see supplementary materials and methods for the temperatures list).

We obtained an acceptance ratio of 25 to 35% for the replica exchanges, depending on the number of atoms, which varied from 4300 to 6900 atoms (solvent + peptide) in function of the size of the peptide. Usually an acceptance ratio of around 30% is targeted and we were pleased to see that the final acceptance ratios were about twice the values (12-15%) predicted by the web-server mentioned above.

With two 32 cores CPUs per replica we obtained 50 to 75 ns/day on two recent 32 cores AMD EPYC 7452 @ 2.35GHz CPUs. Running all nine peptides on eighteen 32 cores CPUs in parallel, we obtained after about 13 to 20 days a single 1000 ns trajectory per peptide. In total this single run consumed about 200k CPU hours (15 days * 64 cores * 9 peptides). Repeating the simulation with five runs like for the implicit solvent was out of reach for this study, neither doing the whole benchmark of the three force fields in explicit solvent with REMD, as taken together it would have consumed 15 times more CPU hours, i.e. about 3 million CPU hours.

## ST simulations

Protein structure were prepared using the pdbfixer module of the OpenMM package, where a bond was specifically added in the peptide topology between the N backbone atom of the first residue and the C backbone atom of the last residue.[75] For simulation run with explicit solvent, peptides were solvated in truncated octahedron boxes with a padding of 1.5 nm and the TIP3P water model. Counter ions were added to counter the charge of the peptide, together with a 150 mM concentration of $Na^+Cl^-$ ions. Simulation were conducted with the OpenMM 7.7 Molecular Dynamics software.[75]

Energy minimization was run for up to 10.000 steps, and followed by an equilibration step of 10 ns in the NPT ensemble. During production and equilibration, a 4 fs integration time step was used with the help of constraining all bonds involding a hydrogen atom, as well as using heavy hydrogens assigned to a 3 atomic mass units. A non-bonded cutoff of 1.0 nm was used with the Particle Mesh Ewald (PME) algorithm for long range electrostatic interactions. A pressure of 1.0 bar was applied during equilibration and production using the Monte-Carlo algorithm and a 25 time steps interval. Temperature was equilibrated using the Langevin Middle integrator and 1 $ps^{-1}$ friction coefficient.

During the equilibration the temperature was fixed to 300.0 K. Simulated Tempering (ST) simulation were conducted using a protocol inspired from a previous work[76] using a python script written by Peter Eastman and modified to implement weight calculation of Park and Pande[77] and on-the-fly weight calculation of Nguyen et al.[78] During ST simulations, temperature exchanges were attempted every 10 ps and depending with the size of the peptide 15 to 20 temperature ladders spaced exponentially between 300.0 and 500.0 K were used. The ST simulation production run had a length of 10 µs.

Simulation run with implicit solvent were run in the NVT ensemble with no cutoff for non-bonded interaction. We used only 8 temperature ladders, and the temperature exchange were attempted every 6 ps. Implicit solvent models were GBN2[79] for simulation run with Amber14 and Charmm36 force field simulation, as we used OBC model[80] with Amber96 force field.

## Analysis of REMD and ST simulations

All analyses in this article were done on the REMD trajectory at 300K and the parts of the ST trajectory that were at 300K. We skipped the first 10% of the trajectory to reduce the bias of the starting conformation. We compared each frame to the experimental reference structure which is here the best representative conformer of the NMR ensemble (first model here). Two measures were calculated for each frame: the backbone RMSD to the experimental reference structure and the radius of gyration. Both were calculated with the mdtraj[81] python package on the heavy atoms of the backbone (N, $C_\alpha$, C and O). Free energy maps in function of both metrics were calculated and plotted with python using a custom function adapted from the PyEMMA package.[82] Therefore a 2D histogram is generated with numpy (100 bins for each axis) and translated into free energy values according the formula: $F = -kT * \log(\rho)$, with $\rho$ being the density at each point of the free energy map.

## Convergence analysis

To determine the convergence of a MD simulation is not a trivial problem. As MD is a random walk, short simulations cannot sample enough conformational landscape and explore all free energy minima. Moreover the populations of the clusters can change during longer trajectories, as shown in figure 1a).

To assess quantitatively the convergence of our implicit solvent REMD simulations, we performed five independent runs for each peptide and force field combination. If the RMSD density profiles of all five runs superpose within a certain error margin, we consider that the REMD simulations converged. The RMSD values are calculated to the experimental reference structure, but in principle any structure could be used as reference here. To quantify this superposition we build first a mean density profile combining all five runs (see figure 1b)). To quantify how much each RMSD profile differ from the mean RMSD profile, the Jensen–Shannon divergence (JSD[83]) is used:

$$\mathrm{JSD}(P||Q) = \frac{1}{2}D_{\mathrm{KL}}(P||M) + \frac{1}{2}D_{\mathrm{KL}}(Q||M) \qquad (1)$$

with $P$: RMSD density profile, $Q$: Mean RMSD density profile, $M = \frac{1}{2}(P+Q)$ and $D_{\mathrm{KL}}$: Kullback-Leibler (KL) divergence.

Unlike KL, JSD is a symmetric metric ($\mathrm{JSD}(A,B) = \mathrm{JSD}(B,A)$) and the range of values is between 0 (similar profiles) to 1 (completely different profiles). If the JSD value is above 0.05, we consider that the RMSD profile is different to the mean RMSD profile (see figure 1b) as an example of JSD values around 0.05).

For each REMD run 100 RMSD density profiles are calculated along the trajectory (0-$t_1$, 0-$t_2$, 0-$t_3$, ..., 0-$t_{100}$) to capture the simulation time evolution of the convergence, which allowed us to plot the JSD curves shown in figure 1c).

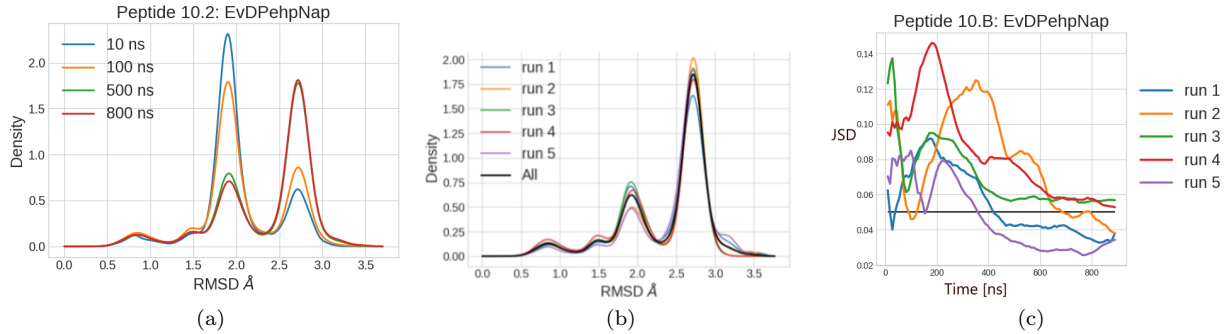
## 2D projections of the conformational landscape

The conformation landscape is projected here to a 2D map, using two complementary projections: the first one has on the x-axis the radius of gyration of the cyclic peptide backbone and on the y-axis the RMSD to the experimental reference structure calculated over the backbone atoms. This projection gives a good idea where the highest populated conformations are situated compared to the experimental reference structure. The radius of gyration is a good measure for the overall shape of the cyclic peptide ring. The second projection is obtained through a principal component analysis (PCA) of the cyclic peptide backbone atoms positions of the molecular dynamics trajectory (see below for more details). We employed the "MDanalysis" python package[84] for the PCA calculation.

For the identification of low free energy clusters we employed two techniques: k-means clustering[85] and HDBscan clustering.[86] See supplementary materials and methods for more details.

## Prediction of the experimental reference structure

Although our molecular dynamics simulations yield an ensemble of conformations and not just a single structure, the most populated cluster of conformations is expected to




(a) (b) (c)

Figure 1: (a) REMD is a random walk, conformational space exploration is not the same with short (<100 ns) and long (> 500 ns) simulations. The determination of the simulated time necessary to get a converged simulation is important to conclude which clusters are most populated. In this example the population of the clusters converged after 500 ns. The RMSD values of the shown profiles are calculated to the experimental reference structure. (b) Comparison of the RMSD profiles of five REMD runs of $1\mu s$ (first 10% skipped) to the mean RMSD profile which combines all five profiles.(c) The Jensen-Shannon divergence (JSD) is used as metric to quantify the difference between two profiles. If the JSD value is above 0.05, we consider that the RMSD profile is different to the mean RMSD profile. Convergence of all five REMD runs is reached if all five curves are below 0.05, here only three of five runs converged, but the other two runs (run 3 and 4 here) are not far from the chosen threshold.

be near to the experimental reference structure. This can be evaluted immediately by looking on the 2D maps that project the conformational landscape onto the RMSD to the experimental reference structure, by reporting the RMSD of the centroid structure of the most populated cluster, i.e. the one with the lowest free energy in general (not always, as the HDBscan clustering algorithm can merge together two clusters of low free energy that do not include the lowest free energy peak).

In real cases the experimental reference structure is in general not available. Therefore we setup a protocol to predict it without the use of the RMSD/radius of gyration projection, but by performing a principal component analysis (PCA) directly from the coordinates of the backbone atoms sampled during the MD simulation (REMD or ST protocol). To do a PCA, the MD trajectory needs first to be aligned to one reference frame. We simply choose the first frame as reference frame. From the PCA we retained as many components as necessary to get a cumulated variance of 80% of the total variance. In our case about four to six principal components (PC) had to be retained. We clustered this multidimensional PCA space with the HDBscan algorithm using the same parameters as described above. The PCA space is projected onto the first two principal components to give the 2D maps shown in figure S5. The centroid of the most populated and lowest free energy cluster is indicated by a green star. In the rare case that the position of the lowest free energy value in the PCA maps is different from the position of the most populated cluster, we plotted a cyan star for the former and a yellow star for the latter. For each star, its RMSD to the experimental reference structure and radius of gyration are calculated, and this coordinate is indicated again by a star of the same color on the 2D maps shown in figure 9.

We tested also a variant of this protocol, where the frame number corresponding to the centroid of the most populated cluster from the first PCA run is used as reference frame to realign the MD trajectory and to perform a second PCA run, that should be less dependent on the arbitrary choice of first frame as reference frame from the first run. As the results did not change significantly (results not shown), we retained the first version of our protocol.

## Combining different force fields and simulation protocols

As will be shown in the results section, each force field or simulation protocol yields a slightly different free energy map for the same peptide. Even though different force fields cannot be combined directly, their resulting free energy maps can be combined to obtain *consensus* free energy maps which should give a more complete view of the conformational space of a given peptide. We combined simulations having either a different force field or a different simulation protocol (REMD vs ST, implicit vs explicit solvent) or both. These combinations were done by concatenating the trajectories of the simulations. As the number of frames of each trajectory is not the same for each simulation, we determined its minimum value and subsampled the other trajectories with the same number of frames, which yielded equal weights of each simulation that contributed to the concatenated trajectory. As before, we constructed the free energy maps from the concatenated trajectories and applied the same protocol described above for the prediction of the experimental reference structure. We excluded the ST-Charmm simulation, due to its bad results on our dataset. All possible combinations over the other seven simulations were done, by joining together two, three or more simulations.

The performance of each combination in the prediction of the experimental reference structure is quantified by a single value, the number of "good" predictions among all nine peptides of our dataset. A good prediction is defined as a RMSD value to the experimental reference structure below a given threshold. This threshold is specific to each peptide and is chosen to be the minimum RMSD value among all seven simulations for a given peptide plus a constant margin of 0.3Å (see figure 10a).

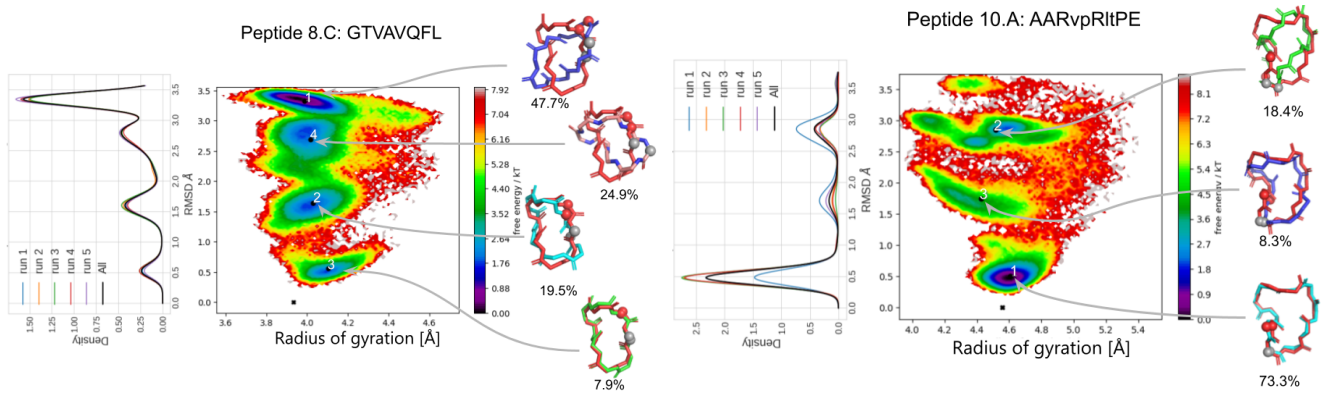



Figure 2: Convergence of simulations assessed by repeating simulations. The Amber96 force field in implicit solvent is employed here. Results for two peptides of REMD simulations repeated with five runs. For each run the density profile of the RMSD to the experimental reference structure is shown on the left. The density profile of the five concatenated trajectories is shown in black. The free energy map is also obtained by the assembly of all five runs. The map is clustered with the k-means algorithm and the centroid structures are superposed to the experimental reference structure (in red). The two first residues are indicated with a red and a gray ball. The size of each cluster is indicated below the 3D structures. For the other peptides the results are shown in the supplementary figure S1.

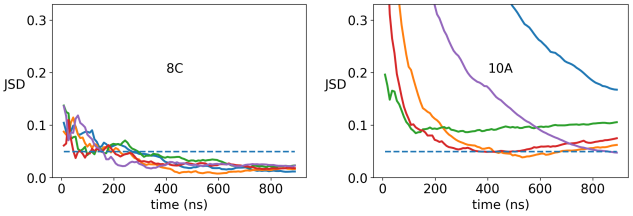



Figure 3: The speed of convergence is shown with the evolution of the Jensen-Shannon distance (JSD) of each RMSD density profile to the average RMSD density. Same REMD simulations as in figure 2. For the other peptides the time evolution of the JSD values is shown in figure S2b. Below the threshold of a JSD of 0.05 two RMSD profiles are nearly identical. Convergence is obtained, if all curves fall and stay below this threshold.

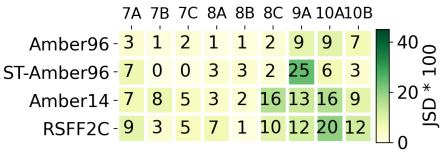



Figure 4: Individual convergence of REMD and ST simulations as measured by the mean JSD of all runs at the end of the trajectories (i.e. the mean of the last points of the curves in figure 3 and figure S2).

## Results & Discussion

### Convergence assessed by repeating simulations

Usually convergence of molecular dynamics simulations can be assessed by repeating the same simulation with different initial conditions (initial velocities for example). We assessed the convergence of REMD simulations as shown for the Amber96 force field for two peptides in figure 2. For the other peptides the results are shown in the supplementary figure S1. We compared the density profiles of each of the 5 runs of the RMSD to the experimental reference structure to the average density profile. For almost all peptides the density profiles converged at the end of the simulation (1000 ns, or more exactly 900 ns as we reject the first 10% to avoid bias from initial conditions), as shown in figure 2 for the peptide 8.C (see figure S1 for the other peptides). In terms of convergence, the worst case was peptide 10.A, but as can be seen from figure 2 the ranking of the three major clusters is not affected by the slow convergence, only the absolute values of the cluster sizes did not converge at the end of the simulation. The speed of convergence can be assessed quantitatively by tracing the evolution of the profile-profile distance measured by the Jensen-Shannon distance (JSD) (figure 3). For the 8C peptide all five simulations converged after only 300 ns, as all Jensen-Shannon distances are under the threshold that we defined (0.05) and that correspond to very high similarity of the density profiles. Overall the six smaller peptides 7A to 8C show a much faster convergence than the three larger peptides 9A to 10B, as demonstrated by the JSD curves in figure S2b and the first line of mean JSD values at the end of the trajectories in figure 4.

As the comparison of the 8C and 10A peptide results shows (figure 2), the convergence is not a guarantee to get the highest population for the cluster with the lowest RMSD to the experimental reference structure. Even though the REMD simulations of the 8C peptide converged very rapidly, all five runs yielded the highest population in a cluster more than 3Å away from the reference structure. In contrary, the five non converged runs of the 10A peptide all ranked the lowest RMSD cluster at the first position in terms of free energy/population size.

For the ST simulations we assessed the convergence by repeating the simulations with the Amber96 force field with two runs, run 1 and run 2. Figures 4 and S3 show that the convergence is here also very good among almost all peptides. Only one case showed a poor convergence (peptide 9A). This case is a bit surprising, as the cluster that accumulated 58% of the trajectory frames in run 1, did not even appear with a small percentage on the run 2. It seems that run 1 explored a cluster that has not been explored by run 2, whereas the other two clusters are both explored by each

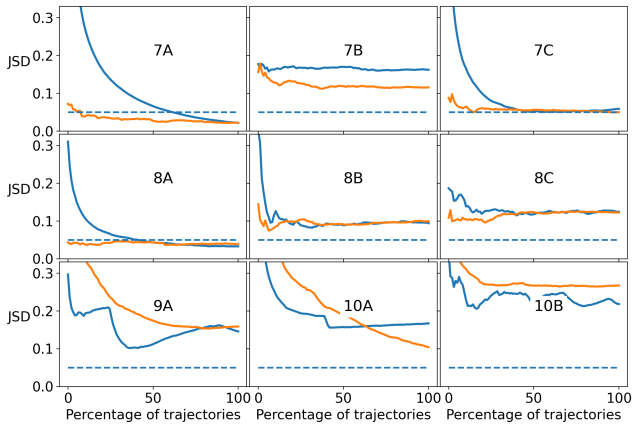



Figure 5: Speed of REMD with ST agreement as measured by the JSD of the RMSD profile of a simulation protocol (ST in blue and REMD in orange) to the averaged RMSD profile (combining ST and REMD profiles). The results obtained with the Amber96 force field are shown here.

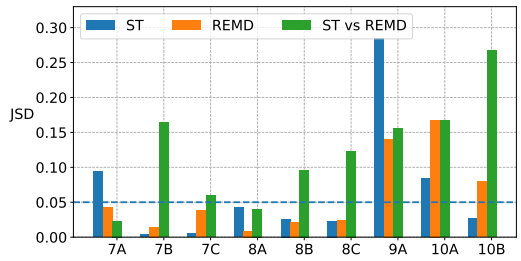



Figure 6: Final convergence or agreement as measured at the end of the trajectories by the maximum JSD of the RMSD profile of a run or simulation protocol to the averaged RMSD profile. The results obtained with the Amber96 force field are shown here. The individual convergence of each simulation protocol (ST in blue and REMD in orange) is compared to the agreement of both taken together ("ST vs REMD" in green). The time evolution of the JSD values is shown in figure S2.

run. Among both runs we selected the run 1 for further analysis which is less favourable in terms of the ranking of the clusters, so the performance of ST with Amber96 is a bit lower rated for 9A. Unlike the REMD results, the convergence speed does not show a clear dependence on peptide size with the ST protocol. As can be seen in figure S2a the convergence speed is quite different among the nine peptides and also among the peptides with the same number of amino acids : peptide 7A shows a horizontal line above the 0.05 JSD threshold. 7B converged very rapidly and 7C needs about 4000 ns simulation time to converge both curves together. For the three larger peptides of eight amino acids, the convergence speed shows a similar heterogeneity. For the largest peptides of ten amino acids (10A and 10B), the convergence is average for 10B (JSD $<$ 0.05 after 4500 ns) and slow for 10A due to a horizontal line like in the 7A case. Nevertheless a JSD threshold of 0.1 instead of 0.05 would classify both cases, 7A and 10A, as being converged after 6000 ns.

The individual convergence of REMD Amber96 and ST-Amber96 is summarized in figure 4 with the mean JSD values at the end of the trajectories. The convergence of the REMD Amber14 and RSFF2C simulations is shown in the same figure. The JSD values are higher here than for the Amber96 force field, especially for the larger or more flexible peptides (8C-10B), but still relatively contained. We cannot explain why the Amber14 and its derived RSFF2C force fields yield a lower convergence among repeated REMD simulations than the Amber96 force field. Maybe the fact that Amber14 and RSFF2C were developed for an explicit solvent plays a role here.

## Agreement across different simulation conditions

Even though each simulation protocol taken alone shows overall very good convergence properties, the agreement among different simulation protocols or force fields is not as good. A quick comparison of the results shown in figures 2 and S3 for peptide 8C illustrates this fact: while the most populated cluster of the REMD simulations is at about 3.3Å from the experimental reference structure, the ST simulations yield a better ranking with the most populated cluster at about 1.7Å. This is more due to different population sizes of each cluster than to "shape" differences of the free-energy maps, i.e. the explored regions. The position of the clusters of the free-energy maps are quite similar among REMD and ST, but each simulation stays a different time in each cluster. To have a more quantitative analysis of the *inter-protocol* agreement, we traced the evolution of RMSD profile-profile JSD distances comparing both protocols, as we compared the runs inside a single protocol (*intra-protocol* convergence). Therefore we merged all available runs into one large trajectory per protocol and took the same number of frames per protocol by sub-sampling the trajectory with a higher number of frames. With the data from each protocol we generated an average RMSD profile which each protocol should reproduce. The JSD curves in figure 5 show that four peptides (7A, 7C, 8A, 8B) showed a good agreement among the REMD and ST protocols, whereas the others, mainly larger peptides, do not. For the cases with a lower agreement (7B, 8C, 9A, 10A, 10B) the curves are almost horizontal at the second half of the trajectories, indicating that longer simulation times would not help here to obtain a better agreement among both protocols.

To summarize the number of cases where each protocol converges independently and taken together, we calculated the maximum JSD value of the endpoints of each JSD curve, as a metric to assess intra-protocol convergence or inter-protocol agreement at the end of the simulated trajectories. The barplot shown in figure 6 confirms that convergence is obtained more often by simply repeating a simulation with the same protocol (REMD or ST here) than by doing a simulation with different MD simulation methods, as REMD and ST here. Nevertheless, both methods agreed quite well, especially for the six smaller peptides (7A-8C). The three larger peptides (9A, 10A, 10B) challenged intra-protocol convergence and inter-protocol agreement.

For the analyses shown up to here, it has to be reminded, that a single value, as here the maximum JSD value at the endpoint, is quite reductive to describe the similarity of two conformational ensembles. As we will see below, the visual comparison of 2D free energy maps is more informative. As an example, while the peptide 7B showed a high maximum JSD of 0.16 in the comparison ST vs REMD in figure 6, the difference is mainly due to high RMSD clusters, present

in the REMD simulations, but not in the ST ones for the Amber96 force field (see figure 9). The highest populated cluster was identical among both simulation protocols. On the other hand, the peptide 9A had the same maximum JSD value as peptide 7B, but the comparison of the free energy maps shows that the REMD and ST maps of peptide 7B for the Amber96 force field look more similar than the ones of the peptide 9A. This visual difference in 2D is not well represented by a single JSD value.

Doing the same analysis as in figure 6 to quantify the inter-protocol and inter-force field agreement for all tested simulation conditions (MD protocol and force field), we obtained the nine heatmaps in figure 7, one for each peptide. The average and standard deviation over all nine peptides are shown in figure 8. The individual heatmaps of each peptide should help in the comparison of the free energy maps shown in figure 9. The order of the columns is the same in figures 7 to 9. The first four columns correspond to the REMD protocols and the last four columns correspond to the ST protocols (as indicated by "ST-" before the force field name, absence of it means REMD protocol). For each protocol, one simulation is done with an explicit solvent (columns 4 and 7, as indicated by "-exp"), whereas the others are done with an implicit solvent. Figure 7 show the singular behaviour of the Charmm-36 force field with the high JSD values in the corresponding "ST-Charmm" row or column. For the peptide 8A the two REMD simulations in implicit solvent with the Amber14 and the RSFF2C force fields show quite high JSD values to the other simulations, as will be discussed in the next section analysing the free energy maps of figure 9. For the peptide 7C the RMSD profile of the explicit solvent REMD simulation with the RSFF2C force field is only similar to the other explicit solvent simulation in our test set: ST-Amber14-exp, even though the force field and the simulation protocol are different.

Overall figure 8 in comparison to figure 4 confirms that agreement across force fields and/or simulation protocols (i.e. *inter-protocol*) is lower than the convergence of repeated simulations under the same conditions (i.e. *intra-protocol*). The lowest JSD values, i.e. best agreement, are observed for the Amber96 force field, comparing REMD with ST simulation protocol (JSD = 0.12±0.07 (mean±SD)). The fact that the RSFF2C force field is derived from the Amber14 force field is confirmed by the second best JSD value of 0.14±0.07 comparing the REMD simulations with Amber14 and RSFF2C in an implicit solvent. The two explicit solvent simulations, REMD RSFF2C-exp and ST-Amber14-exp, hold the third best JSD value of 0.15±0.05, hinting to the visual similarity of the free energy maps obtained in the two corresponding columns in figure 9.

The switch in MD protocol (REMD ↔ ST) had overall a lower impact (average JSD values: 0.12±0.07 for Amber96 and 0.18±0.11 for Amber14) than the switch of the force field family from Amber96 to Amber14 (or RSFF2C): average JSD of 0.23±0.09 for REMD and 0.19±0.10 for ST. The switching from implicit to explicit solvent had a greater impact on the RSFF2C force field with REMD than on the Amber14 force field with ST, as quantified by the JSD values of 0.24±0.08 (RSFF2C ↔ RSFF2C-exp) and 0.17±0.09 (ST-Amber14 ↔ ST-Amber14-exp), respectively.

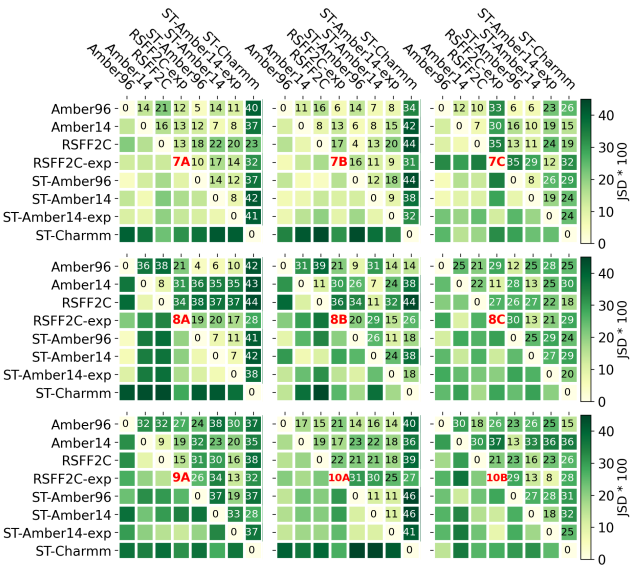



Figure 7: Inter-protocol and force field agreement as measured by the mean JSD of the RMSD profiles to the averaged RMSD profile. The mean JSD is calculated here at the end of the trajectories, as a measure of the agreement obtained at the end of the REMD or ST simulations. Agreement between force fields and simulation protocols (ST where indicated, REMD otherwise). "exp" stands for explicit solvent, all other simulations were done in implicit solvent.

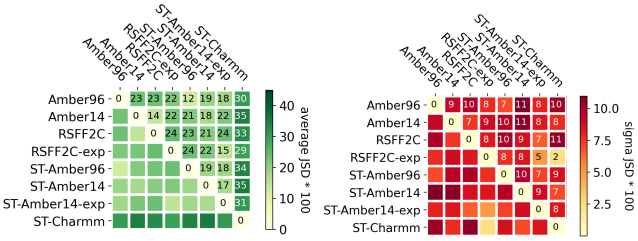



Figure 8: By averaging through the heatmap values of the nine peptides shown in figure 7, the "average JSD" values are reported here on the left with their respective standard deviations on the right.

## Free energy maps obtained with REMD and ST on several force fields

Figure 9 shows the free energy maps of seven different simulation conditions over all nine peptides of our dataset. Overall, the three force fields Amber96, Amber14 and the Amber14 variant, RSFF2C, produce similar free energy maps whereas the Charmm36 force field gives quite different results (see figure S4). As Charmm36 seems to have important difficulties in reproducing structures at low RMSD to the experimental reference structure for our dataset, we do not discuss results using this force field any longer in this study.

The observation of multiple free energy minima in almost each map shows that the cyclic peptides of our dataset have some conformational flexibility, despite the fact that all of them, except peptide 8C, were designed to have a high conformational stability. Only the smallest cyclic peptides of our dataset with seven amino acids (7A-C) show in most

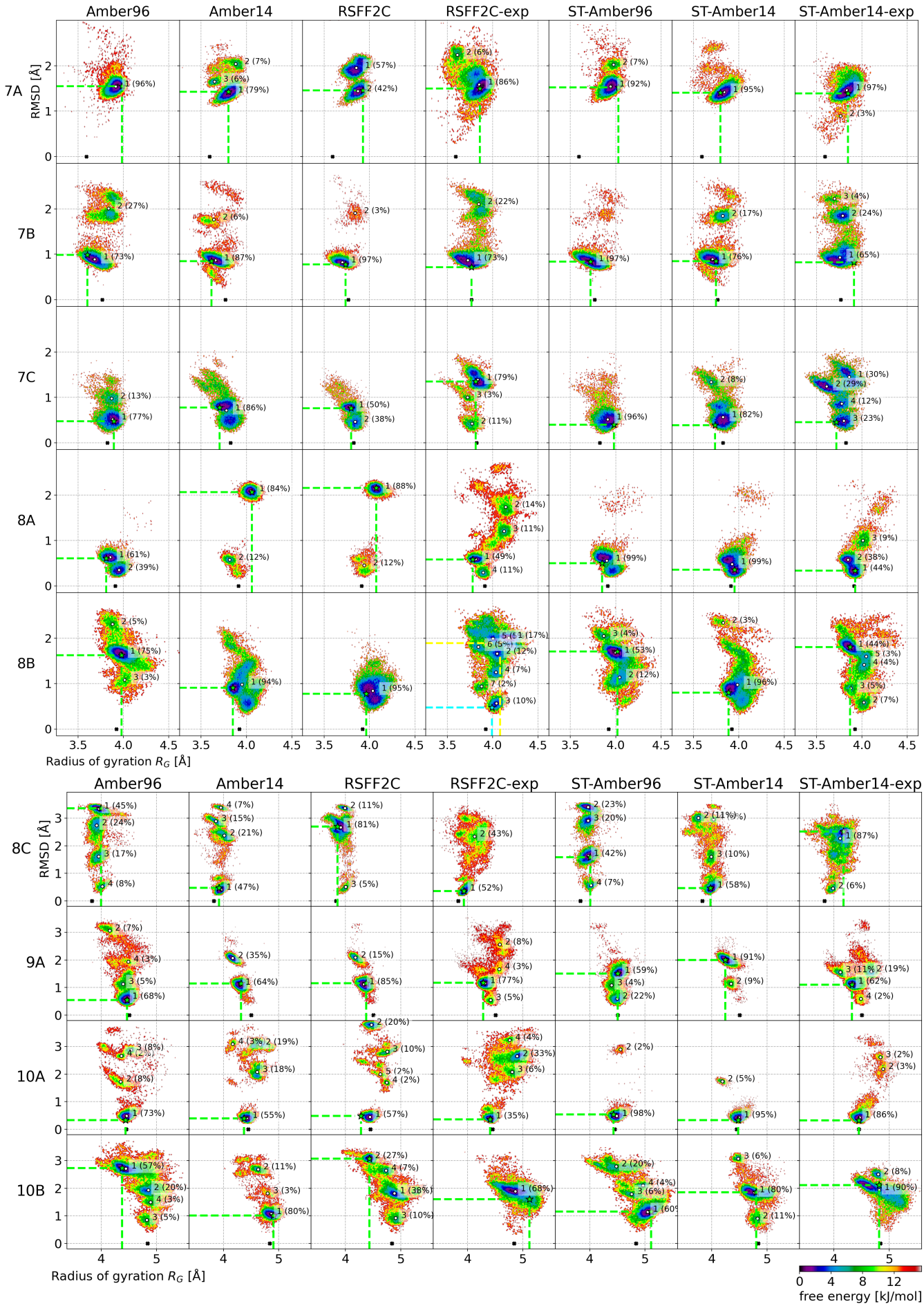



Figure 9: (Caption on next page.)

Figure 9: (Previous page.) Free energy maps projected onto the radius of gyration (horizontal axis) and the RMSD to the reference structure. The units of the axes are in Å and the free energy unit is in kJ/mol. The first four columns were obtained with the REMD protocol and the last four columns with the ST protocol. The force field employed is indicated on top of each column. RSFF2C-exp and ST-Amber14-exp are explicit solvent simulations. The other columns are implicit solvent simulations. The clustering results are indicated by the cluster ids and their fraction of the total number of frames, from the most populated cluster (id = 1) to the less populated cluster (clusters with a size below 2% are not shown). The centroid of each cluster is indicated by white dots. The centroid position of the most populated and lowest free energy cluster of the PCA projection (see Figure S5) is indicated by a green star and green dashed lines. In the rare case that the position of the lowest free energy value in the PCA maps is different from the position of the most populated cluster, we plotted a cyan star for the former and a yellow star for the latter. The radius of gyration of the reference experimental structure is indicated by a black cross on the 0.0 RMSD line.

cases only one wide free energy minimum. It is difficult to prove that the alternative free energy minima with high RMSD distances to the experimental reference structure exist *in vitro*, as they have not been observed in the NMR studies that produced the reference structures. But some of these alternative free energy minima were observed across several simulation conditions (MD protocols and force fields), which should give a strong evidence for their existence. For example the peptide 8B has a low free energy region centred at a RMSD of 1.7Å and $R_G = 4.0$Å that is observed in all but two (Amber14 and RSFF2C) simulation conditions. The more flexible peptide 8C has a low free energy "band" spanning a large interval of RMSD values from 0.2Å to 3.5Å almost without interruption, i.e. white areas in the maps. Several free energy minima are observed in this "band", across all simulation conditions. The larger peptide 10B shows a similar free energy landscape, with the difference that the high RMSD regions are shifted to lower $R_G$ values than the one of the experimental reference structure. Only the two explicit solvent simulations (RSFF2C-exp and ST-Amber14-exp) do not sample significantly the highest RMSD region above 2.8Å and the lowest RMSD region below 1Å, they are both centred in a more restricted RMSD interval of $[1, 2.8]$Å, compared to the implicit solvent simulations that span a large RMSD interval of $[0.5, 3.5]$Å. This may be explained by the slowdown of conformational changes of the peptide due to the explicit water molecules, as found already in.[87] But for the other ten residues peptide 10A the opposite is observed, when comparing ST-Amber14 with ST-Amber14-exp. The latter has spans to higher RMSD regions.

Looking at the impact of the MD protocol (REMD vs ST), one observes that overall, the maps are quite similar for a given force field, showing that the simulations are well converged and that the sampling method does not influence much the obtained free energy maps. Some small differences can be observed nevertheless: For some peptides, regions with high RMSD conformations compared to the experimental reference structure are more populated by the REMD simulations compared to the ST simulations. This can be seen the best on peptides 9A (Amber 96 only) and 10A (Amber 96 and Amber 14). The lowest free energy cluster is very similar for a given force field, except for some cases where this cluster is lower in RMSD for one method compared to the other. REMD yields a lower RMSD for the peptides 9A and 10B (Amber 14 both) and ST yields a lower RMSD for the peptides 8A (Amber 14), 8C and 10B (Amber 96), see figure 10b.

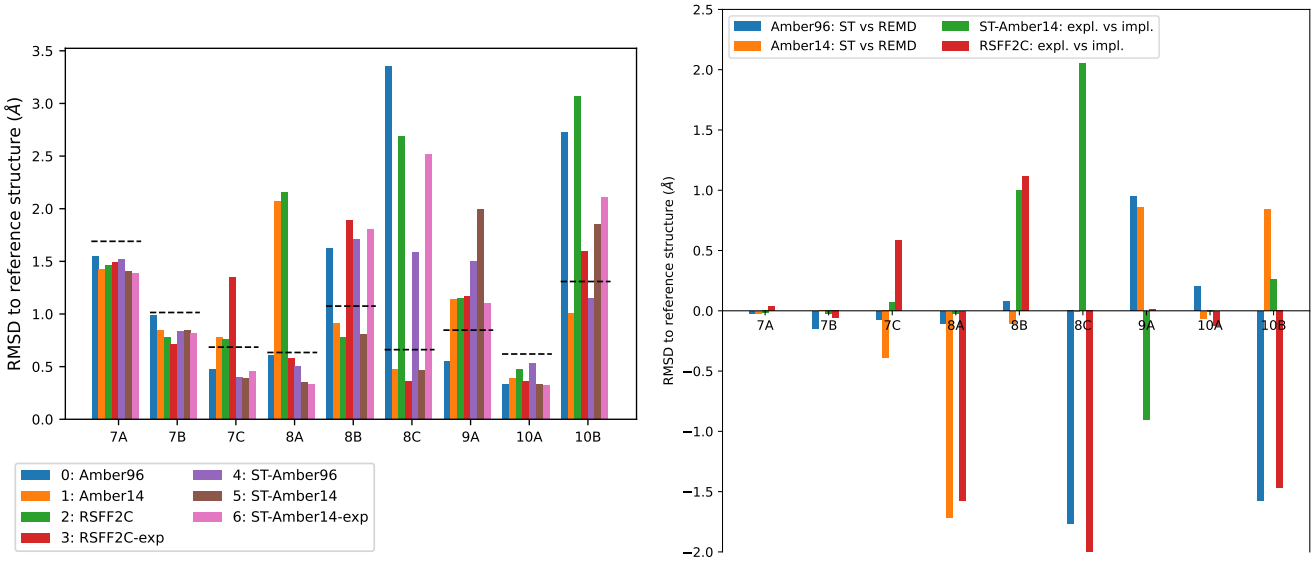
## Prediction of the experimental reference structure

Applying the protocol described in the materials & methods section to predict the experimental reference structure from our MD simulations we obtained the best clusters indicated by the green stars and green dashed lines in figure 9. The green stars correspond to the most populated and lowest free energy cluster in the PCA maps (figure S5) which are constructed without the knowledge of the experimental reference structure. As can be seen in figure 9 the positions of the green stars in the RMSD/radius of gyration projection superpose very well with the global free energy minima of these maps. As described in the materials & methods section, in the case that the position of the lowest free energy value in the PCA maps is different from the position of the most populated cluster, we plotted a cyan star for the former and a yellow star for the latter. This rare case happened only in one map, namely for the peptide 8B with the RSFF2C-exp force-field: the cyan star gave here a much lower RMSD compared to the yellow star (0.5Å vs 1.8Å). In the further analysis we retained the position of the yellow star for peptide 8B, as this single case cannot validate one prediction method over the other.

In figure 10a we reported the RMSD values of the green (or yellow) stars for each MD simulation. To see the full context of these RMSD values, one should keep a look also on the full free energy maps, shown in figure 9. For the smallest cyclic peptides of the set, with seven amino acids (7A-7C), the different simulations converged to similar RMSD values (figure 10a). Comparing the RMSD values among these three peptides 7A to 7C, one can observe that the peptide 7A has a threefold higher RMSD than the peptide 7C (1.5Å vs 0.5Å). This shows that the notion of "low" RMSD is specific to each peptide. We define a peptide specific RMSD threshold below which a prediction is classified as "good" (see dashed lines in figure 10a). This threshold is chosen to be the minimum RMSD value among all seven simulations for a given peptide plus a constant margin of 0.3Å.

For the other peptides, only the peptide 10A shows a very good agreement among the force fields and simulation protocols: All simulations yield the same global free energy minimum at only 0.4 - 0.5Å (figure 10a), despite the relatively large size of the peptide 10A of 10 amino acids, that allows a large range of conformations with up to 4Å of RMSD to the reference structure (figure 9).

The performance on the peptide 10A is even more impressive, if it is compared to the second peptide with 10 amino acids in the set: the peptide 10B. Here the lowest RSMD is only 1.0Å obtained with two simulations: ST-Amber96 and REMD Amber14 (figure 10a). Nevertheless, all simulations sampled with a significant frequency the sub 1Å region




(a) (b)

Figure 10: (a) RMSD to the reference structure for the best ranked cluster for each peptide. The threshold below which a RMSD value is considered as a good prediction is indicated by black horizontal dashed lines (see methods). Results shown for each peptide. (b) Comparison of the two simulation methods, REMD and ST, for two force fields Amber96 (blue) and Amber14 (orange) in implicit solvent. The RMSD values of the lowest free energy cluster obtained in either method are compared by subtracting the RMSD value obtained by REMD from the one obtained by ST. Positive values indicate cases where REMD give better results and negative values indicate cases where ST give better results. The green bars compare the explicit solvent ST-Amber14 simulation to its implicit solvent equivalent. The red bars compare the explicit solvent REMD-RSFF2C simulation to its implicit solvent equivalent. For the green and red bars positive values indicate cases where the implicit solvent give better results and negative values indicate cases where the explicit solvent give better results.

(see figure 9). The differences among the simulations in the population sizes of the clusters may come from insufficient convergence, as REMD and ST show here quite different results, even if the same force field is used (figure 10b). It may come also simply from a relatively flat free energy landscape for the peptide 10B, that does not give a clear preference for the experimental reference conformation.

The results obtained on the peptide 8A are quite similar to the peptide 7C, except for the REMD Amber14 and RSFF2C simulations in implicit solvent. The other five simulations of peptide 8A give a very low RMSD of $0.4 - 0.6$Å (figure 10a) and do not sample much the higher RMSD regions (figure 9). The REMD Amber14 and RSSF2C simulations sample the same low RMSD regions as the other simulations, but at a lower frequency than the unique cluster at 2.2Å (figure 9), that is not sampled at all by the other simulations.

For the peptide 8B the Amber14 force field and its derivative RSFF2C give the best results in terms of low RMSD with a good value of 0.9Å. The use of an explicit solvent degraded this good result here for both force fields (figure 10b). Here the two simulation protocols, REMD and ST, give exactly the same results (figure 10b), also the free energy maps look very similar (figure 9).

The peptide 8C is particular as it contains only L-amino acids and is therefore probably more flexible than the other peptides that contain a mixture of L and D-amino acids. In fact, the free energy maps of 8C show a large landscape from 0.5Å to 3.5Å of RMSD (figure 9). For such a flexible peptide it is difficult to rank the lowest RMSD cluster in the first position. The Amber14 force field pick up the right cluster, with a very low RMSD of 0.4Å. The use of an explicit solvent degrades this performance, as the ST-Amber14 explicit solvent simulations gives a high RMSD cluster of 2.5Å as the most populated cluster. Interestingly, for the RSFF2C force field it is the opposite case (figure 10b): the explicit solvent improved the ranking with the lowest observed RMSD for 8C of 0.3Å, instead of a high RMSD of 2.7Å for the implicit solvent simulation. The use of the Amber96 force field gives here even worse results, with 3.4Å using the REMD protocol. The ST protocol yields also a quite low free energy basin at 3.4Å, but the basin at 1.5Å has here a slightly lower free energy, explaining the great difference of the two protocols using the Amber96 force field for the peptide 8C (figure 10a). Both protocols result here in a quite similar free energy map (figure 9, peptide 8C), showing that the differences observed in figure 10a should not be over-interpreted, a look on the full free energy maps is still necessary.

The last peptide to be analyzed is the peptide 9A: Here the REMD Amber96 simulation gives the best results of 0.6Å. The Amber14 force field give a doubled RMSD value (1.2Å) for the REMD protocol and for the ST protocol, but only in explicit solvent. The ST-Amber14 implicit solvent simulation ranks the higher RMSD cluster of 2Å at the first position. The RSFF2C force field give almost identical free energy maps as with Amber14 force field using the REMD protocol. The explicit solvent RSFF2C simulation populates the 0.6Å cluster a bit more than its implicit solvent counterpart (figure 9).

## Explicit vs implicit solvent for structure prediction

The Amber14 force field has been optimized for the use of an explicit solvent, therefore we expected better results using this force force field with explicit solvent than with implicit solvent. Especially the comparison of the ST-Amber14 implicit solvent simulations with its explicit solvent counterpart (ST-Amber14-exp column in figure 9) should show the improvements by the use of an explicit solvent vs an implicit solvent. But surprisingly this is observed only for the peptides 7A (lowest free energy cluster does not improve, but lower RMSD areas are explored, see figure 9) and 9A (negative green bar in figure 10b). If the comparison for the peptide 9A is done against the other implicit solvent simulations, the lowest RMSD cluster is as well or better ranked with the three REMD simulations (Amber96, Amber14 and RSFF2C), showing that an explicit solvent is not a requirement to obtain a good ranking. In contrary, we observed more cases where the use of an explicit solvent degrades the ranking for ST-Amber14, namely 8B, 8C and 10B (positive green bars in figure 10b).

## RSFF2C force field

Finally, we turn to looking at the free energy maps obtained with RSFF2C which is a derivative from the Amber14 force field. The MD simulations are done in implicit and explicit solvent using the REMD protocol (RSFF2C and RSFF2C-exp, respectively). For most peptides the results obtained with RSFF2C are quite similar to the results obtained with Amber14 in implicit solvent for both sets, REMD and ST (see figure 9). The peptide 8A is a special case, as only the two REMD simulations with Amber14 and RSFF2C found a large cluster at a high RSMD of 2.2Å. This special case underlines the similarity of the two forces fields Amber 14 and RSFF2C, at least in implicit solvent conditions. The RSFF2C simulations in explicit solvent (RSFF2C-exp) gave better results for three peptides (8A, 8C and 10B, see negative red bars in figure 10b) and gave only for two peptides, 7C and 8B (positive red bars in figure 10b), worse results, compared to the implicit solvent simulation (RSFF2C). This is the opposite result for what we found when comparing implicit vs explicit solvent for the Amber14 force field (see above), despite the fact that the RSFF2C force field is derived from the Amber14 force field. It seems therefore more important to have an explicit solvent with the RSFF2C force field, than with the Amber14 force field.

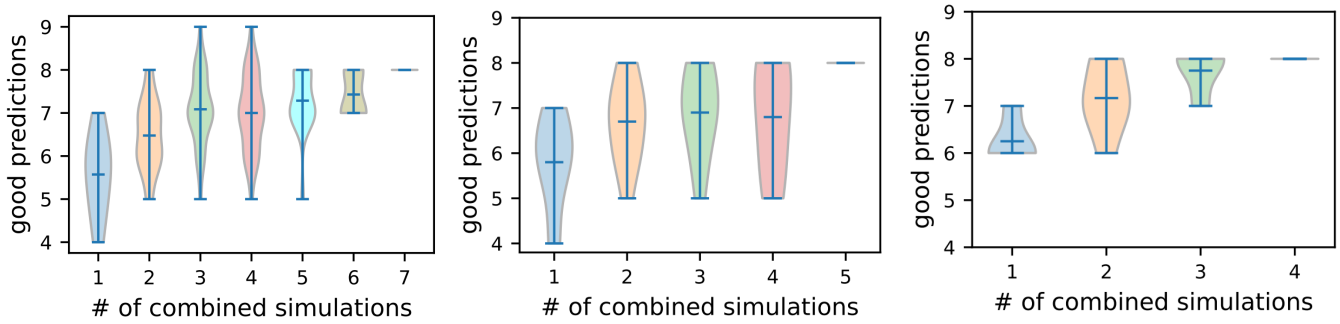
## Combining different force fields and simulation protocols

As described in the methods section, we combined different simulations to achieve more robust results, especially in the prediction of the experimental reference structure. The rationale behind this approach is the observation that while the low RMSD clusters are often the same among the different simulations, high RMSD clusters are more spread. Combining the simulations should increase the populations of the low RMSD clusters, while decreasing the ones of high RMSD clusters, and therefore increasing the chance to rank the low RMSD clusters at the first position. The figure S6 shows that if the simulations are taken individually, the number of good predictions oscillates from four for the REMD RSFF2C simulation to seven for the ST-Amber14 simulation (implicit solvent for both), but none of the simulations gives a good prediction for all nine peptides of our dataset.

The values of figure S6 are reported on the top left in figure S7a as "single". If only a second simulation is added to one of the seven simulations, the number of good predictions increases significantly as can be seen by the comparison of the "twins" curve with the "single" curve in figure S7a and also with the violin plots in figure 11a. Figure S8 gives the details of the 21 possible combinations of twins among the seven simulations (binomial coefficient: $\binom{n}{k} = \frac{n!}{k!(n-k)!} = 21$ with $n = 7$ and $k = 2$ here). Two twin combinations (id 4 and 18 in figure S8), Amber14 + ST-Amber14 and ST-Amber96 + ST-Amber14, yield eight good predictions, one more than the ST-Amber14 simulation alone. Beside the gain of one good prediction by the two optimal twins, it is interesting to note that the average number of good predictions increased also by one good prediction from 5.5 to 6.5 good predictions, see the violin plots in figure 11a. Figure S8a shows that both combinations failed for the same peptide, 9A, to rank the lowest RMSD cluster in the first position. The free energy maps of the two twin combinations (figure S14) shows that ST-Amber96 + ST-Amber14 populates the sub-1Å RMSD region more than Amber14 + ST-Amber14, but with a cluster size of only 11% it is not selected. The Amber14 + ST-Amber14 twin shows a higher cluster size of 36% at a good RMSD value of 1.1Å, but the cluster at 2.0Å had here a higher population of 63%.

It needs three simulations combined to obtain one particular combination among all 35 triplets that achieves good predictions for all nine peptides: Amber96 + Amber14 + ST-Amber14-exp (see figure S9b). For the peptide 9A this combination has a cluster at 0.6Å RMSD with a population of 22%, which is lower than its second best cluster at 1.1Å RMSD with a population of 40% (figure S16), but the PCA projection, shown in figure S17, has a most populated cluster that corresponds to the 0.6Å RMSD cluster which is therefore selected here. It may be surprising that the ST-Amber14 simulation is not in this triplet combination, but it can be explained by the high RMSD value of this simulation for the peptide 9A (see figure S6a). The individual three simulations of this triplet have the lowest three RMSD values for the peptide 9A (figure S6a), explaining the good performance of this particular triplet.

The violin plots in figure 11a shows that the average number of good predictions is again increased from 6.5 for twins to 7.0 for triplets. A further increase of the number of simulations combined to four simulations does not improve the average number of good predictions. A small increase is observed when passing from four to five simulations and again when passing from five to six and to seven simulations. Beside this increase in the average number of good predictions, only one combination that have less than seven good predictions is observed for five combined simulations, and for six combined simulations all combinations are between seven and eight good predictions. Combining all seven simulations yield eight good predictions, only the peptide 9A has again a RMSD value above the threshold, but a very close RMSD value of 1.0Å compared to the threshold of 0.8Å (see figure S13a). A general rule seems to emerge here: the higher the number of different simulations (not just runs) combined together, the more robust the results of the prediction of the experimental structure. The robustness stems from the observation that any combination of a high number of different simulations (five and higher) yields for almost all peptides good prediction results.




(a) (b) (c)

Figure 11: Combining simulations to improve the number of good predictions of the reference structure. See supplementary figures for more details. (a) Number of good predictions for all possible combinations of simulations. The maximum number of good predictions is nine, as the dataset contains nine peptides. (b) Number of good predictions for all combinations of simulations with implicit solvent only. (c) Number of good predictions for all combinations of simulations with implicit solvent only and excluding the RSFF2C force field.

From all the 127 tested combinations (from one to seven combined simulations), only two combinations yielded a good prediction for all nine peptides. The triplet Amber96 + Amber14 + ST-Amber14-exp discussed above and the quadruplet Amber96 + RSFF2C-exp + ST-Amber96 + ST-Amber14 (combination id = 16 in figure S10b). Only the Amber96 simulation is included in both top combinations. It is therefore difficult to find a general rule which combination of simulations gives the best results. This underlines the importance of the robustness feature found here for a high number of different simulations, as it renders the need to find the optimal combination(s) to a secondary goal.

Nevertheless this analysis can be done also on subsets of simulations that perform the best individually or simulations that take less computational time, like the simulations in implicit solvent. If the same analysis is done on the simulations with implicit solvent only, we obtain a similar improvement of the number of good predictions, if two simulations are combined (see figures 11b and S7c). The triplets or quadruplets did not improve the number of good predictions further, only taking all five simulations together yield eight good predictions. Adding the two explicit solvent simulations did not improve this performance. The figure S6b shows that the RSFF2C force field in implicit solvent does not perform as well as the Amber96 and Amber14 force fields in implicit solvent. Therefore we filtered it out in another combinations analysis (see figures 11c and S7d), where already 3/4 of the triplets achieve eight good predictions and the triplet (Amber96 + Amber14 + ST-Amber96) achieves still seven good predictions. This shows that a limited number of implicit solvent simulations with the two Amber force fields, Amber96 and Amber14, and with the two simulation methods, REMD and ST yield already robust results, predicting the experimental reference structure for 8 out of 9 cyclic peptides. As the peptides are not simulated in interaction with a binding partner, there is no need here to use explicit solvent simulations to generate accurate free energy landscapes. Three or four implicit solvent simulations will take less computational time than a single explicit solvent simulation and more crucially will give robust and accurate results, which is not achieved with a single explicit solvent simulation.

## Conclusion

Our study explored various molecular dynamics "ingredients" for the cartography of the conformationnal landscape of cyclic peptides and the prediction of their reference experimental structures. Three types of ingredients were investigated: the force field, the solvent model and the simulation method. For the force fields, no single best force field could be observed, older force fields like Amber96 and Amber14 performed as well as the recent RSFF2C force field. The use of an explicit solvent instead of an implicit solvent showed a mixed picture: for RSFF2C the prediction of the native structure worked better in explicit solvent, while for Amber14 we observed the opposite. Under the same force field and solvent model both simulation methods, REMD and ST, give very similar free energy maps, especially for the lowest free energy clusters. Some differences could be observed in higher free energy clusters and REMD tended to produce more high free energy clusters than ST. No clear advantage for one method vs. the other in the prediction of the native structure could be observed. The *intra-protocol* convergence of repeated simulations under the same conditions has been observed to be higher than the *inter-protocol* agreement of simulations using different conditions, i.e. force field, solvent model or simulation method. This finding is important to underline, as convergence of simulations is often only evaluated by repeating simulations under the same conditions, which seems to be an insufficient convergence criteria.

Among the nine peptides, the native structure of on average six peptides were predicted correctly with a single MD simulation. The ST-Amber14 (implicit solvent) simulation yielded the best result here with seven peptides correctly predicted and the REMD RSFF2C (implicit solvent) simulation the worst with only four peptides correctly predicted. To improve this sucess rate we combined systematically all seven simulations, merging two, three or more simulations together. The average number of good predictions increased gradually from six to eight with the number of merged simulations, which can be explained by the observation that high RMSD clusters are less conserved among the seven simulations than the lowest RMSD cluster. The merging of the simulations enforces therefore the populations of the lowest RMSD cluster which becomes the most populated cluster more often. The two explicit solvent simulations performed

here, are not necessary to obtain eight good predictions, the five implicit solvent simulations are sufficient for this. As the RSFF2C simulation in implicit solvent yielded taken alone only four good predictions, we excluded it and obtained again eight good predictions with only four simulations in implicit solvent: REMD Amber96 + Amber14 and ST Amber96 + Amber14. Removing one of the four simulations, decreases the number of good predictions by one for only one triplet, for the other three triplets the performance is not altered.

The combination of different "ingredients" of MD simulations, here the force field and the simulation method, seems to yield more robust results, with a better native structure prediction performance. Three or four implicit solvent simulations are already sufficient and will take less computational time than a single explicit solvent simulation. In addition to the combination of complementary MD simulations, our native structure prediction method is based on a PCA projection of the free energy landscape and robustly yield good predictions for eight out of the nine peptides. It would be interesting to apply our method also to the structure prediction of proteins and protein-petide complexes, as well as cyclic peptides with chemical modifications, like N-methylation.[35]

## Data and Software Availability

ST simulation protocol:
`https://github.com/samuelmurail/SST2/tree/main/bin`

Trajectory data and python scripts for analysis:
`https://doi.org/10.5281/zenodo.15112951`

**Acknowledgement** We would like to thank Fan Jiang and Wei Kang from Peking University, Beijing, for sending us their RSFF2C force field parameters files. This project was provided with computer and storage resources by GENCI at CINES thanks to the grant 2021-A0100707641 on the supercomputer Occigen. The RPBS and iPOP-UP platforms also contributed with calculation time.

## Supporting Information Available

Supplementary materials and methods and supplementary figures are available for download.

## References

(1) Fisher, M. F.; Payne, C.; Chetty, T.; Crayn, D.; Berkowitz, O.; Whelan, J.; Rosengren, K. J.; Mylne, J. S. The genetic origin of evolidine, the first cyclopeptide discovered in plants, and related orbitides. *bioRxiv* **2020**,

(2) Hosseinzadeh, P. et al. Comprehensive computational design of ordered peptide macrocycles. *Science* **2017**, *358*, 1461–1466.

(3) Robinson, J. A.; DeMarco, S.; Gombert, F.; Moehle, K.; Obrecht, D. The design, structures and therapeutic potential of protein epitope mimetics. *Drug discovery today* **2008**, *13*, 944–951.

(4) Olmez, E. O.; Akbulut, B. S. *Binding Protein*; InTech, 2012.

(5) Gao, M.; Cheng, K.; Yin, H. Targeting protein- protein interfaces using macrocyclic peptides. *Pept. Sci.* **2015**, *104*, 310–316.

(6) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic peptides: current applications and future directions. *Sig Transduct Target Ther* **2022**, *7*, 48.

(7) Ji, X.; Nielsen, A. L.; Heinis, C. Cyclic Peptides for Drug Development. *Angewandte Chemie International Edition* **2024**, *63*, e202308251.

(8) Lamers, C. Overcoming the shortcomings of peptide-based therapeutics. *Future Drug Discovery* **2022**, *4*, FDD75.

(9) Cowell, S.; Lee, Y.; Cain, J.; Hruby, V. Exploring Ramachandran and chi space: conformationally constrained amino acids and peptides in the design of bioactive polypeptide ligands. *Current medicinal chemistry* **2004**, *11*, 2785–2798.

(10) Workman, R. J.; Gorle, S.; Pettitt, B. M. Effects of Conformational Constraint on Peptide Solubility Limits. *J. Phys. Chem. B* **2022**, *126*, 10510–10518, Publisher: American Chemical Society.

(11) Tran, H. L.; Lexa, K. W.; Julien, O.; Young, T. S.; Walsh, C. T.; Jacobson, M. P.; Wells, J. A. Structure–Activity Relationship and Molecular Mechanics Reveal the Importance of Ring Entropy in the Biosynthesis and Activity of a Natural Product. *J. Am. Chem. Soc.* **2017**, *139*, 2541–2544, Publisher: American Chemical Society.

(12) Bhardwaj, G. et al. Accurate de novo design of hyperstable constrained peptides. *Nature* **2016**, *538*, 329.

(13) Yongye, A. B.; Li, Y.; Giulianotti, M. A.; Yu, Y.; Houghten, R. A.; Martínez-Mayorga, K. Modeling of peptides containing D-amino acids: implications on cyclization. *Journal of computer-aided molecular design* **2009**, *23*, 677–689.

(14) Hickey, J. L.; Sindhikara, D.; Zultanski, S. L.; Schultz, D. M. Beyond 20 in the 21st Century: Prospects and Challenges of Non-canonical Amino Acids in Peptide Drug Discovery. *ACS Med. Chem. Lett.* **2023**, *14*, 557–565.

(15) Gentilucci, L.; De Marco, R.; Cerisoli, L. Chemical modifications designed to improve peptide stability: incorporation of non-natural amino acids, pseudo-peptide bonds, and cyclization. *Current pharmaceutical design* **2010**, *16*, 3185–3203.

(16) Mas-Moruno, C.; Beck, J. G.; Doedens, L.; Frank, A. O.; Marinelli, L.; Cosconati, S.; Novellino, E.; Kessler, H. Increasing $\alpha v\beta 3$ Selectivity of the Anti-Angiogenic Drug Cilengitide by N-Methylation. *Angew. Chem., Int. Ed.* **2011**, *50*, 9496–9500.

(17) Jwad, R.; Weissberger, D.; Hunter, L. Strategies for Fine-Tuning the Conformations of Cyclic Peptides. *Chem. Rev.* **2020**, *120*, 9743–9789, Publisher: American Chemical Society.

(18) Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. The exploration of macrocycles for drug discovery—an underexploited structural class. *Nat. Rev. Drug Discovery* **2008**, *7*, 608.

(19) Saha, I.; Dang, E. K.; Svatunek, D.; Houk, K. N.; Harran, P. G. Computational generation of an annotated gigalibrary of synthesizable, composite peptidic macrocycles. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 24679–24690.

(20) Li, J.; Yanagisawa, K.; Sugita, M.; Fujie, T.; Ohue, M.; Akiyama, Y. CycPeptMPDB: A Comprehensive Database of Membrane Permeability of Cyclic Peptides. *J. Chem. Inf. Model.* **2023**, *63*, 2240–2250.

(21) Merz, M. L.; Habeshian, S.; Li, B.; David, J.-A. G. L.; Nielsen, A. L.; Ji, X.; Il Khwildy, K.; Duany Benitez, M. M.; Phothirath, P.; Heinis, C. De novo development of small cyclic peptides that are orally bioavailable. *Nat Chem Biol* **2023**, 1–10, Publisher: Nature Publishing Group.

(22) Costa, L.; Sousa, E.; Fernandes, C. Cyclic Peptides in Pipeline: What Future for These Great Molecules? *Pharmaceuticals* **2023**, *16*, 996, Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

(23) Barman, P.; Joshi, S.; Sharma, S.; Preet, S.; Sharma, S.; Saini, A. Strategic Approaches to Improvise Peptide Drugs as Next Generation Therapeutics. *Int J Pept Res Ther* **2023**, *29*, 61.

(24) Hellinger, R.; Sigurdsson, A.; Wu, W.; Romanova, E. V.; Li, L.; Sweedler, J. V.; Süssmuth, R. D.; Gruber, C. W. Peptidomics. *Nat Rev Methods Primers* **2023**, *3*, 1–21, Publisher: Nature Publishing Group.

(25) Rossino, G.; Marchese, E.; Galli, G.; Verde, F.; Finizio, M.; Serra, M.; Linciano, P.; Collina, S. Peptides as Therapeutic Agents: Challenges and Opportunities in the Green Transition Era. *Molecules* **2023**, *28*, 7165, Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

(26) Moulahoum, H.; Ghorbanizamani, F.; Tok, K.; Zihnioglu, F. On the Cyclization of Non-cyclic Peptides for Biological Applications: Inspiration from Naturally Cyclic Peptides. *ChemistrySelect* **2023**, *8*, e202301335.

(27) Adiga, U. Trends in Peptide Drug Discovery: Unleashing the Therapeutic Potential of Peptides: A Review. **2023**,

(28) Ramadhani, D.; Maharani, R.; Gazzali, A. M.; Muchtaridi, M. Cyclic Peptides for the Treatment of Cancers: A Review. *Molecules* **2022**, *27*, 4428, Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

(29) Muttenthaler, M.; King, G. F.; Adams, D. J.; Alewood, P. F. Trends in peptide drug discovery. *Nat Rev Drug Discov* **2021**, *20*, 309–325, Publisher: Nature Publishing Group.

(30) Hill, T. A.; Shepherd, N. E.; Diness, F.; Fairlie, D. P. Constraining Cyclic Peptides To Mimic Protein Structure Motifs. *Angewandte Chemie International Edition* **2014**, *53*, 13020–13041.

(31) Malde, A. K.; Hill, T. A.; Iyer, A.; Fairlie, D. P. Crystal Structures of Protein-Bound Cyclic Peptides. *Chem. Rev.* **2019**, *119*, 9861–9914, Publisher: American Chemical Society.

(32) Kessler, H. Conformation and Biological Activity of Cyclic Peptides. *Angew. Chem. Int. Ed. Engl.* **1982**, *21*, 512–523.

(33) Damjanovic, J.; Miao, J.; Huang, H.; Lin, Y.-S. Elucidating Solution Structures of Cyclic Peptides Using Molecular Dynamics Simulations. *Chem. Rev.* **2021**, *121*, 2292–2324.

(34) Leman, J. K. et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods* **2020**, *17*, 665–680, Publisher: Nature Publishing Group.

(35) Jusot, M.; Stratmann, D.; Vaisset, M.; Chomilier, J.; Cortés, J. Exhaustive exploration of the conformational landscape of small cyclic peptides using a robotics approach. *Journal of Chemical Information and Modeling* **2018**, *58*, 2355–2368.

(36) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* **1999**, *314*, 141–151.

(37) Earl, D. J.; Deem, M. W. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910–3916, Publisher: The Royal Society of Chemistry.

(38) Voelz, V. A.; Dill, K. A.; Chorny, I. Peptoid conformational free energy landscapes from implicit-solvent molecular simulations in AMBER. *Peptide Science* **2011**, *96*, 639–650.

(39) Butterfoss, G. L.; Yoo, B.; Jaworski, J. N.; Chorny, I.; Dill, K. A.; Zuckermann, R. N.; Bonneau, R.; Kirshenbaum, K.; Voelz, V. A. De novo structure prediction and experimental characterization of folded peptoid oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 14320–14325.

(40) Wakefield, A. E.; Wuest, W. M.; Voelz, V. A. Molecular Simulation of Conformational Pre-Organization in Cyclic RGD Peptides. *Journal of chemical information and modeling* **2015**, *55*, 806–813.

(41) Damas, J. M.; Filipe, L. C.; Campos, S. R.; Lousa, D.; Victor, B. L.; Baptista, A. M.; Soares, C. M. Predicting the Thermodynamics and Kinetics of Helix Formation in a Cyclic Peptide Model. *J. Chem. Theory Comput.* **2013**, *9*, 5148–5157, Publisher: American Chemical Society.

(42) Yu, H.; Lin, Y.-S. Toward structure prediction of cyclic peptides. *Phys. Chem. Chem. Phys.* **2015**, *17*, 4210–4219.

(43) Geng, H.; Jiang, F.; Wu, Y.-D. Accurate structure prediction and conformational analysis of cyclic peptides with residue-specific force fields. *The journal of physical chemistry letters* **2016**, *7*, 1805–1810.

(44) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553–4559, Publisher: American Chemical Society.

(45) Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. Substrate Binding Mechanism of HIV-1 Protease from Explicit-Solvent Atomistic Simulations. *J. Am. Chem. Soc.* **2009**, *131*, 11811–11818, Publisher: American Chemical Society.

(46) Cossio, P.; Marinelli, F.; Laio, A.; Pietrucci, F. Optimizing the Performance of Bias-Exchange Metadynamics: Folding a 48-Residue LysM Domain Using a Coarse-Grained Model. *J. Phys. Chem. B* **2010**, *114*, 3259–3265, Publisher: American Chemical Society.

(47) McHugh, S. M.; Rogers, J. R.; Yu, H.; Lin, Y.-S. Insights into How Cyclic Peptides Switch Conformations. *J. Chem. Theory Comput.* **2016**, *12*, 2480–2488, Publisher: American Chemical Society.

(48) McHugh, S. M.; Yu, H.; Slough, D. P.; Lin, Y.-S. Mapping the sequence–structure relationships of simple cyclic hexapeptides. *Phys. Chem. Chem. Phys.* **2017**, *19*, 3315–3324, Publisher: The Royal Society of Chemistry.

(49) Slough, D. P.; McHugh, S. M.; Lin, Y.-S. Understanding and designing head-to-tail cyclic peptides. *Biopolymers* **2018**, *109*, e23113.

(50) Slough, D. P.; McHugh, S. M.; Cummings, A. E.; Dai, P.; Pentelute, B. L.; Kritzer, J. A.; Lin, Y.-S. Designing Well-Structured Cyclic Pentapeptides Based on Sequence–Structure Relationships. *J. Phys. Chem. B* **2018**, *122*, 3908–3919, Publisher: American Chemical Society.

(51) Slough, D. P.; Yu, H.; McHugh, S. M.; Lin, Y.-S. Toward accurately modeling N-methylated cyclic peptides. *Phys. Chem. Chem. Phys.* **2017**, *19*, 5377–5388.

(52) Cummings, A. E.; Miao, J.; Slough, D. P.; McHugh, S. M.; Kritzer, J. A.; Lin, Y.-S. $\beta$-Branched Amino Acids Stabilize Specific Conformations of Cyclic Hexapeptides. *Biophysical journal* **2019**, *116*, 433–444.

(53) Fonseca Lopez, F.; Miao, J.; Damjanovic, J.; Bischof, L.; Braun, M. B.; Ling, Y.; Hartmann, M. D.; Lin, Y.-S.; Kritzer, J. A. Computational Prediction of Cyclic Peptide Structural Ensembles and Application to the Design of Keap1 Binders. *J. Chem. Inf. Model.* **2023**, *63*, 6925–6937.

(54) Miao, J.; Descoteaux, M. L.; Lin, Y.-S. Structure prediction of cyclic peptides by molecular dynamics + machine learning. *Chem. Sci.* **2021**, *12*, 14927–14936.

(55) Hui, T.; Descoteaux, M. L.; Miao, J.; Lin, Y.-S. Training Neural Network Models Using Molecular Dynamics Simulation Results to Efficiently Predict Cyclic Hexapeptide Structural Ensembles. *J. Chem. Theory Comput.* **2023**, *19*, 4757–4769.

(56) Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589, Publisher: Nature Publishing Group.

(57) Rettie, S. A. et al. Cyclic peptide structure prediction and design using AlphaFold. 2023; `http://biorxiv.org/lookup/doi/10.1101/2023.02.25.529956`.

(58) Zhang, C.; Zhang, C.; Shang, T.; Wu, X.; Duan, H. Highfold: accurately predicting cyclic peptide monomers and complexes with AlphaFold. 2023; `https://www.biorxiv.org/content/10.1101/2023.08.27.554979v1`, Pages: 2023.08.27.554979 Section: New Results.

(59) Kosugi, T.; Ohue, M. Design of Cyclic Peptides Targeting Protein–Protein Interactions Using AlphaFold. *International Journal of Molecular Sciences* **2023**, *24*, 13257, Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.

(60) Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, 1–3, Publisher: Nature Publishing Group.

(61) Miao, J.; Ghosh, A. P.; Ho, M. N.; Li, C.; Huang, X.; Pentelute, B. L.; Baleja, J. D.; Lin, Y.-S. Assessing the Performance of Peptide Force Fields for Modeling the Solution Structural Ensembles of Cyclic Peptides. *J. Phys. Chem. B* **2024**, acs.jpcb.4c00157.

(62) Meng, E. C.; Goddard, T. D.; Pettersen, E. F.; Couch, G. S.; Pearson, Z. J.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* **2023**, *32*, e4792.

(63) Case, D. A. et al. AmberTools. *J. Chem. Inf. Model.* **2023**, *63*, 6183–6191, Publisher: American Chemical Society.

(64) Da Silva, A. W. S.; Vranken, W. F. ACPYPE-Antechamber python parser interface. *BMC research notes* **2012**, *5*, 367.

(65) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born radii in the generalized Born approximation: the importance of being perfect. *Journal of computational chemistry* **2002**, *23*, 1297–1304.

(66) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(67) Sindhikara, D.; Meng, Y.; Roitberg, A. E. Exchange frequency in replica exchange molecular dynamics. *The Journal of chemical physics* **2008**, *128*, 01B609.

(68) Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *The Journal of chemical physics* **2004**, *121*, 10748–10756.

(69) Kollman, P. A. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Accounts of Chemical Research* **1996**, *29*, 461–469.

(70) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713, Publisher: American Chemical Society.

(71) Jiang, F.; Wu, H.-N.; Kang, W.; Wu, Y.-D. Developments and Applications of Coil-Library-Based Residue-Specific Force Fields for Molecular Dynamics Simulations of Peptides and Proteins. *J. Chem. Theory Comput.* **2019**, *15*, 2761–2773, Publisher: American Chemical Society.

(72) Feng, J.-J.; Chen, J.-N.; Kang, W.; Wu, Y.-D. Accurate Structure Prediction for Protein Loops Based on Molecular Dynamics Simulations with RSFF2C. *J. Chem. Theory Comput.* **2021**, *17*, 4614–4628, Publisher: American Chemical Society.

(73) https://virtualchemistry.org/remd-temperature-generator/.

(74) Patriksson, A.; Spoel, D. v. d. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073–2077, Publisher: The Royal Society of Chemistry.

(75) Eastman, P. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **2017**, *13*, e1005659, Publisher: Public Library of Science.

(76) Dorgham, K.; Murail, S.; Tuffery, P.; Savier, E.; Bravo, J.; Rebollo, A. Binding and Kinetic Analysis of Human Protein Phosphatase PP2A Interactions with Caspase 9 Protein and the Interfering Peptide C9h. *Pharmaceutics* **2022**, *14*, 2055, Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

(77) Park, S.; Pande, V. S. Choosing weights for simulated tempering. *Phys. Rev. E* **2007**, *76*, 016703, Publisher: American Physical Society.

(78) Nguyen, P. H.; Okamoto, Y.; Derreumaux, P. Communication: Simulated tempering with fast on-the-fly weight determination. *The Journal of Chemical Physics* **2013**, *138*, 061102.

(79) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034, Publisher: American Chemical Society.

(80) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **2004**, *55*, 383–394.

(81) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109*, 1528 – 1532.

(82) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* **2015**, *11*, 5525–5542.

(83) Lin, J. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* **1991**, *37*, 145–151.

(84) Naughton, F. B. et al. MDAnalysis 2.0 and beyond: fast and interoperable, community driven simulation analysis. *Biophysical Journal* **2022**, *121*, 272a–273a, Publisher: Elsevier.

(85) MacQueen, J.; others Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 1967; pp 281–297.

(86) McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software* **2017**, *2*, 205.

(87) Anandakrishnan, R.; Drozdetski, A.; Walker, R. C.; Onufriev, A. V. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophysical Journal* **2015**, *108*, 1153–1164, Publisher: Elsevier.