

# Metal Cation-Binding Mechanisms of Q-Proline Peptoid Macrocycles in Solution

Matthew F. D. Hurley, Justin D. Northrup, Yunhui Ge, Christian E.  
Schafmeister, and Vincent A. Voelz\*

*Department of Chemistry, Temple University, Philadelphia, PA 19122, USA*

E-mail: voelz@temple.edu

## Abstract

The rational design of foldable and functionalizable peptidomimetic scaffolds requires the concerted application of both computational and experimental methods. Recently, a new class of designed peptoid macrocycle incorporating spiro oligomer proline mimics (Q-prolines) has been found to pre-organize when bound by monovalent metal cations. To determine the solution-state structure of these cation-bound macrocycles, we employ a Bayesian inference method (BICePs) to reconcile enhanced-sampling molecular simulations with sparse ROESY correlations from experimental NMR studies. Conformations predicted to be most populated in solution were then simulated in the presence of explicit cations to yield trajectories with observed binding events, revealing a highly-preorganized all-*trans* amide conformation, whose formation is likely limited by the slow rate of *cis/trans* isomerization. Interestingly, this conformation differs from a racemic crystal structure solved in the absence of cation. Free energies of cation binding computed from distance-dependent potentials of mean force suggest  $\text{Na}^+$  has higher affinity to the macrocycle than  $\text{K}^+$ , with both cations binding much more strongly in acetonitrile than water. The simulated affinities are able to correctly

rank the extent to which different macrocycle sequences exhibit preorganization in the presence of different metal cations and solvents, suggesting our approach is suitable for solution-state computational design.

## Introduction

Non-biological peptide mimics, or *peptidomimetics*, have desirable properties that include a high degree of biocompatibility, diversity of chemical space, resistance to proteolysis, and cell permeability. While many different peptidomimetics now exist, it remains a key challenge to rationally design novel synthetic scaffolds that, like proteins, exhibit preorganized structure in solution. Preorganization is crucial for functional properties such as high-affinity binding to desired targets,<sup>1</sup> or efficient catalysis through the positioning of specific chemical groups.<sup>2</sup>

### Spiroligomer-peptoid macrocycles

Spiroligomers are shape-programmable rigid molecular scaffolds of bis-peptide rings whose stereochemistry can be exquisitely controlled.<sup>3</sup> Spiroligomers are highly functionalizable,<sup>4,5</sup> and can present preorganized arrangements of chemical groups to mimic peptide ligands,<sup>6</sup> artificial enzymes,<sup>7,8</sup> and form supermolecular metal-binding assemblies.<sup>9</sup>

Peptoids (*N*-substituted oligoglycines) are another class of highly functionalizable and sequence-programmable peptide mimics.<sup>10</sup> Unlike peptides, peptoids lack a backbone amide hydrogen bond donor and can populate both *cis* and *trans* amide conformations. Nevertheless, many peptoids are *foldamers*,<sup>11,12</sup> peptidomimetic molecules able to self-assemble into preorganized three-dimensional structures such as helices<sup>13–16</sup> and other architectures<sup>17–19</sup> for judiciously chosen side chains.

Just as cyclization is known to stabilize folded peptide conformations,<sup>20–24</sup> so have similar strategies been successful at achieving pre-organized spiroligomer<sup>25</sup> and peptoid macrocycles.<sup>26–34</sup> Indeed, preorganization due to cyclization helps explain the prevalence of cyclic

designs in the modest corpus of peptoid crystal structures that have been solved to date.<sup>1</sup> Interestingly, it has been found that cyclic peptoids can host metal cation adducts through the coordination of backbone carbonyl groups,<sup>31,35</sup> reminiscent of natural products such as mycotoxins.<sup>33</sup>

Northrup et al. has recently synthesized cyclic hexamers of spiroligomer-peptoid hybrids<sup>36</sup> that can bind monovalent metal cations (Figure 1).<sup>37</sup> The architecture of the macrocycle is highly functionalizable, with alternating spiroligomer and peptoid residues. The portion of the spiroligomer residues participating in the macrocycle backbone resembles a  $\beta$ -substituted proline, which has led to these structures being called “enhanced prolines” or “Q-prolines”, while the complete macrocycles are named “Q-proline macrocycles (QPM)”. Each Q-proline residue contains a selectable stereocenter and two functionalization sites (substituents  $R_1$  and  $R_2$ ). Each peptoid residue can also be *N*-substituted ( $R_3$ ). These spiroligomer-peptoid macrocycles are structurally similar to proline-substituted cyclic peptoid hexamers that have been shown to coordinate metal cations with high affinity.<sup>31</sup> While crystal structures of the proline-substituted macrocycles show cations bound in the center of the macrocycle, the cation-bound structure of spiroligomer-peptoid macrocycles in solution is unknown.

Northrup et al. reports several features of spiroligomer-peptoid macrocycles warranting further exploration through computational modeling. First, solution NMR studies show a sharpening of spectra after metal cation is added, suggesting rearrangement from a heterogeneous structural ensemble to a single conformational population. The spectral resolution occurs over a timescale of minutes to hours, implying that isomerization of backbone amides may be responsible for this slow conformational rearrangement. Second, different cation affinities and extents of preorganization are observed depending on the sequence of Q-Pro and peptoid residues employed in the design. This suggests that cation-binding function may be tuned by controlling the distribution of conformational populations, namely by rewarding

---

<sup>1</sup>A comprehensive database of published peptoid structures can be found at: <https://wp.nyu.edu/kirshenbaumlab/peptoid-data-bank/>

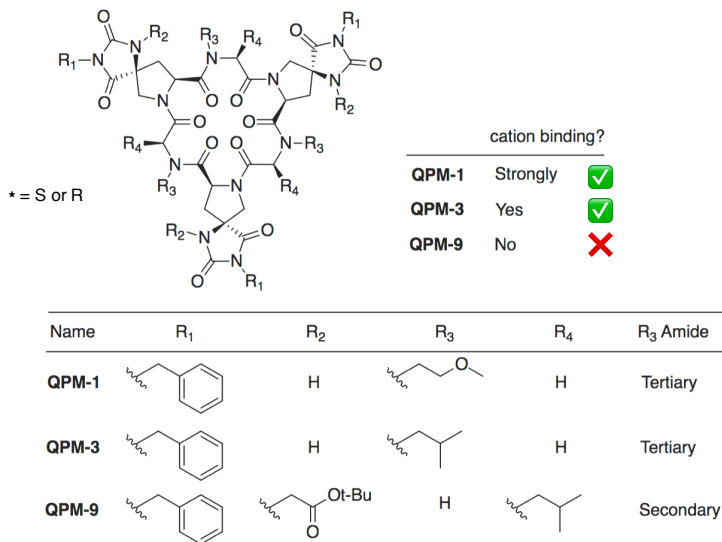


Figure 1: The Q-proline macrocycle (QPM) scaffold as described by Northrup et al. Shown are the three macrocycle sequences simulated in this study. QPM-1 and QPM-3 bind metal cations, while QPM-9 (the control) does not. All are three-fold symmetric.

one or more binding-competent conformations.

In this work, we use molecular simulation alongside sparse NMR restraints to infer the solution structure of cation-bound spiroligomer-peptoid macrocycles using a Bayesian inference approach, and then perform further simulations to investigate the mechanism and thermodynamics of cation binding. Our modeling correctly ranks the sequence preferences for cation-binding to the macrocycles, and predicts a conformational selection binding model that requires isomerization to an all-*trans* amide backbone in solution, a conformation distinct from the *apo* conformation(s) recently obtained from a racemic crystal structure of QPM-3.<sup>37</sup> The methods we present provide a new path towards the rational design of spiroligomer-peptoid macrocycles, and have implications for the general problem of foldamer prediction design.

# Methods and Results

## An integrated approach to modeling foldamers

While protein design efforts have benefited from vast databases of known protein structures, the rational design of non-biological foldamers faces the dual challenges of novel synthesis and limited knowledge of architectural rules. Progress in foldamer design has thus relied much on computational modeling, with successful design efforts resulting from the tight collaboration of both experiment and theory.<sup>30,38–41</sup> One approach to foldamer prediction and design has been to use *ab initio* modeling to construct rotamer libraries of residue types that can be used with existing tools for protein design such as Rosetta.<sup>38,42–44</sup> While this enables fast searching of sequence and structure space, the main drawback is the difficulty of parameterization. The rotamers of each new residue must be calculated separately, and it is challenging to empirically parameterize inter-residue interactions without a large number of known structures.

Alternatively, molecular simulation approaches offer more thorough sampling of foldamer conformational landscapes,<sup>30,45</sup> albeit at the expense of efficient searching through sequence space. The main challenges to simulation-based approaches are: (1) sampling all relevant conformations to obtain accurate estimates of conformational free energies, and (2) the accuracy of the force field parameters, which is essential to foldamer prediction and design. To address the first challenge, we perform enhanced sampling using temperature replica exchange molecular dynamics (tREMD),<sup>46</sup> which is able to efficiently overcome *cis/trans* isomerization barriers.

The second of these challenges—force field accuracy—is arguably more difficult. There have been many attempts to parameterize all-atom force fields for peptoids<sup>15,47,48</sup> and other foldamers,<sup>49,50</sup> but the chemical diversity of peptidomimetics makes the development of a general-purpose force field for foldamers difficult. Here, we have opted to pursue a slightly different philosophy for foldamer simulation which utilizes our Bayesian Inference of Con-

formational Populations (BICePs) algorithm.<sup>51</sup> Instead of having to perform a custom force field parameterization for each new class of foldamer, we first perform molecular simulations using a general-purpose force field for organic molecules (in this case GAFF<sup>52</sup>) and then use Bayesian inference to refine the predicted conformational populations using sparse <sup>1</sup>H-NMR ROESY correlations implemented as distance restraints. An advantage of the BICePs approach is that the refinement of conformational populations can be performed as a post-processing step, with Bayesian inference used to optimally combine both computational and experimental information, implicitly taking into account unknown uncertainties in the experimental measurements.

## Simulation and modeling protocol

An overview of our simulation and modeling protocol is shown in Figure 2. Briefly, we use implicit-solvent temperature replica-exchange molecular dynamics (tREMD) in the absence of ions to sample the populations of backbone amide *cis/trans* isomers.<sup>15,45,46</sup> Conformations sampled at 300 K, which favor a mixture of *cis*- and *trans*-amide backbone states, are then clustered based on time-lagged Independent Component Analysis (tICA), using the amide dihedral angles as input features.<sup>53</sup> This resulted in 100 conformational states and their corresponding populations, which were used as input for the BICePS algorithm<sup>51,54</sup> which further refined these predictions using a number of NMR ROESY correlations measured for the ion-bound macrocycles.

The conformations with the highest populations predicted by BICePs were then simulated in explicit solvent (water and acetonitrile) in the presence of ions (Na<sup>+</sup> and K<sup>+</sup>), with the aim of generating trajectories containing cation binding events. The analysis of the trajectory data allows determination of binding pathways and affinities of monovalent cations to the spiroligomer-peptoid macrocycles.

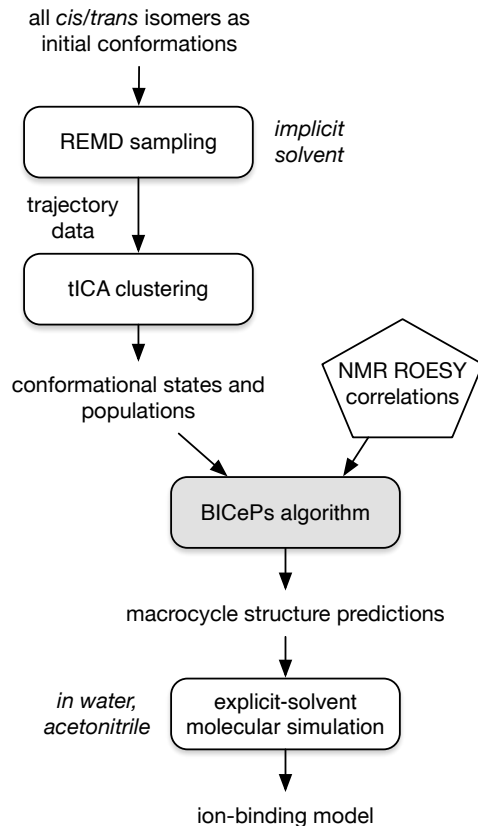


Figure 2: A flowchart of methods used to model macrocycle cation-binding, starting from the generation of initial structures.

## Macrocycle sequences

Three spiroligomer-peptoid macrocycle sequences from Northrup et al.<sup>37</sup> were studied in this work; all are three-fold symmetric (Figure 1). Two of the sequences, QPM-1 and QPM-3, were found to bind metal cations in solution. The third sequence, QPM-9, does not bind cations; it is included as a negative control. These structures differ in both the stereochemistry of the spiroligomer residue as well as the substituents on the N-substituted glycine.

## Molecular topologies

Molecular topologies for peptoid and spiroligomer residues were parameterized using the *AmberTools16* software package.<sup>55</sup> Partial charges for acetyl- and *N*-dimethyl-capped residues

were calculated using the semi-empirical AM1-BCC method using *antechamber*, and assigned atom types from the General Amber Force Field (GAFF).<sup>52</sup> Residue topologies were bonded and cyclized using *tleap*, and converted for use with the GROMACS molecular dynamics simulation package<sup>56</sup> using *ACPYPE*.<sup>57</sup> For explicit water-solvated trajectories, the TIP3P water model was used. Acetonitrile was parameterized using the Automated Topology Builder running DFT (via B3LYP/6-31G\*). Charges were computed with a Merz-Singh-Kollman (MK) scheme and non-bonded parameters were computed from the QM potential. MD simulation files for acetonitrile can be found at <https://atb.uq.edu.au/molecule.py?molid=1004>.

## Temperature replica exchange molecular dynamics (tREMD) simulation.

A significant challenge in simulating peptoids and enhanced proline residues is overcoming the large energy barriers associated with *cis/trans* isomerization of backbone amides. At room temperature, isomerization of these bonds can occur on the timescale of seconds or longer.<sup>58</sup> As done in previous peptoid simulation studies,<sup>15,16,45</sup> we overcome amide dihedral ( $\omega$ -angle) barriers by using temperature replica-exchange molecular dynamics (tREMD),<sup>30</sup> with 36 replicas ranging from 300–800 K, spaced linearly in temperature. To ensure thorough sampling, initial replica conformations were chosen to include all possible *cis/trans* isomers. In general, six amide bonds should require  $2^6 = 64$  initial replicas; due to three-fold symmetry, however, this number of unique isomers reduces to 26 initial replicas.

tREMD simulations were performed on the CB2RR high performance computing cluster at Temple University using GROMACS 5.1.2<sup>56</sup> with a Generalized Born implicit solvent model.<sup>59</sup> Stochastic dynamics, using Langevin integration and water-like viscosity, was used to generate trajectories of between 667 and 677 ns for each replica, resulting in 24.38  $\mu$ s and 24.01  $\mu$ s aggregate trajectory data for molecules QPM-1 and QPM-3 respectively. Replica exchanges were attempted every 10 ps, with average acceptance ratios of 78.9%. Trajectory



snapshots were saved every 10 ps. The three lowest temperature replicas for each macrocycle were used for conformational analysis, and clustered into 100 conformational states using a  $k$ -centers algorithm.

To validate converged sampling of backbone amide conformations, an autocorrelation analysis was performed on tREMD replicas (i.e. following a given replica’s conformation as it heats and cools in a tREMD simulation). Autocorrelation functions  $f(\tau) = \langle \chi(t)\chi(t+\tau) \rangle_t$  were computed for the indicator variable  $\chi(t)$ , where  $\chi$  takes the value 0 for *cis*-amides ( $-90^\circ < \omega < +90^\circ$ ) and the value 1 for *trans*-amides ( $-90^\circ < \omega < +90^\circ$ ). Plotting the function  $g(\tau) = (f(\tau) - \langle \chi \rangle^2) / (\langle \chi^2 \rangle - \langle \chi \rangle^2)$  reveals dihedral angle decorrelation times to be around 20 ns, much shorter than the trajectory lengths ( $> 600$  ns) of simulations (Figure 3).

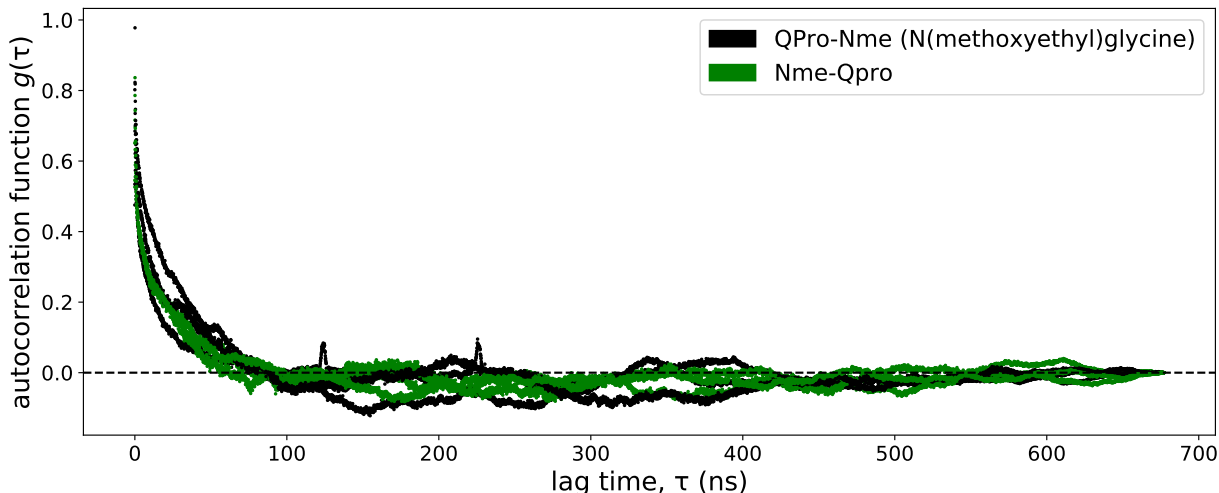


Figure 3: The autocorrelation function  $g(\tau)$ , shown for QPM-1, shows dihedral angle decorrelation times around 20 ns, indicating well-converged sampling of amide dihedrals over the much longer ( $> 600$  ns) tREMD trajectory lengths.

## NMR observables

$^1\text{H}$ -NMR spectra were measured as described in Northrup et al.<sup>37</sup> The complex spectra of peptoid macrocycles in solution in the absence of metal cations suggests several slowly interconverting conformations of the macrocycle at equilibrium. In contrast, spectra that were measured for macrocycles in the presence of 100 mM of monovalent trifluoromethylsulfonate

(triflate) salts show a remarkable structuring in solution over the course of 12 hours. These NMR experiments were performed with data collection every three minutes for the first 90 minutes, followed by every half hour up to 300 minutes, at which point the peaks had largely converged. Based on the timescale of the spectral sharpening, Northrup et al. hypothesized that the observed structuring corresponds to the slow isomerization of amide bonds around the macrocycle to a single predominant structure.

A sparse set of ROESY correlations (four in total) was obtained using 2D NMR for molecules QPM-1 and QPM-3, as described in Northrup et al. (Table 1). The ROEs were classified as either strong or weak, with distance restraints of 0.25 nm and 0.38 nm, respectively, implemented in the BICePs algorithm.

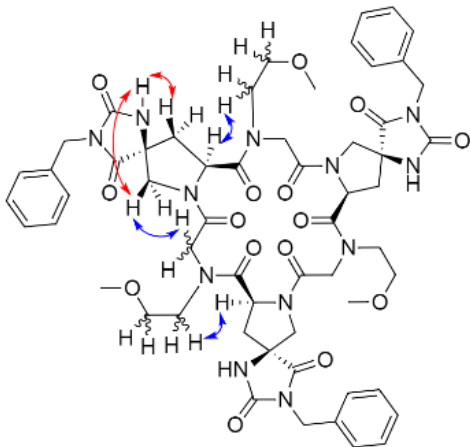


Figure 4: ROESY  $^1\text{H}$  correlations, shown here for QPM-1. Strong and weak ROEs are denoted by blue and red arrows, respectively.

Table 1: ROE correlations used as distance restraints in the BICePs calculation. Qspr = (*S*)-Q-proline; Nme = *N*-(methoxyethyl)glycine;

Proton 1	Proton 2	ROE Strength	distance restraint (nm)
Nme H $\alpha$	Qspr H5	strong	0.25
Qspr H5	Qspr H7	weak	0.38
Qspr H7	Qspr H $\beta$	weak	0.38
Qspr H $\alpha$	Nme H $\beta$	strong	0.25

# Bayesian Inference of Conformational Populations (BICePs)

The BICePs algorithm<sup>51,54</sup> was used to reconcile simulated macrocycle populations with the experimental ROESY measurements. The output of BICePs is a set of refined conformational populations optimally consistent with both experiment and theory. BICePs has previously been used to successfully infer conformational populations of macrolides,<sup>51,60</sup> peptide macrocycles<sup>23</sup> and peptoid foldamers.<sup>15,61</sup>

Briefly, BICePs is a Bayesian inference method applied to a set of discrete conformational states  $X$ . The main goal of BICePs is to infer the posterior distribution  $P(X | D)$  of state populations, given some sparse experimental data  $D$ . Results from theoretical modeling act as the prior distribution,  $P(X)$ , which is reweighted by a likelihood function  $Q(D|X)$  describing how well the conformational state  $X$  agrees with experimental restraints. By Bayes' Theorem:

$$P(X | D) \propto Q(D|X)P(X) \quad (1)$$

The likelihood function is modeled using a Gaussian error function,

$$Q(D | X) = \prod_{j=1} (2\pi\sigma_d^2)^{-1/2} e^{-(r_j(X) - \gamma' r_j^{\text{exp}})^2 / 2\sigma_d^2} \quad (2)$$

where  $r_j(X)$  represents the simulation prediction of the  $j^{\text{th}}$  observable for conformational state  $X$ , and  $r_j^{\text{exp}}$  is the  $j^{\text{th}}$  experimentally measured observable. Here, the observables are calculated from the simulation using the  $1/r^6$ -averaged distances,  $r_j(X) = \langle r_j^{-6} \rangle^{-1/6}$ , where the average is taken over all snapshots sampled from conformational state  $X$ . The experimental values are derived from the ROESY data, with  $r_j$  set to 0.25 nm and 0.38 nm for strong and weak correlations, respectively.

A so-called ‘‘nuisance’’ parameter in the likelihood function is the uncertainty  $\sigma_d$ , which here is applied to each experimental distance restraint, and reflects the combined uncertainty arising from experimental errors and conformational heterogeneity of the simulated ensemble. An additional nuisance parameter,  $\gamma'$ , arises from the uncertainty in the average interproton

distance  $d$  obtained from the experimentally measured NOE intensity  $I$  through the relation  $I = \gamma_0 \langle d^{-6} \rangle$ , for some scaling factor  $\gamma_0$ . To deal with the uncertainty, the experimental values  $r_j^{\text{exp}}$  are scaled by  $\gamma' = (\gamma/\gamma_0)^{-1/6}$ , and  $\gamma'$  is included as a variable in the posterior distribution. Since the values of  $\sigma_d$  and  $\gamma'$  are not known *a priori*, BICePs uses uninformative Jeffreys priors  $P(\sigma_d) = 1/\sigma_d$  and  $P(\gamma') = 1/\gamma'$  to include them in the posterior,

$$P(X, \sigma_d, \gamma' | D) \propto Q(D | X)P(X)P(\sigma_d)P(\gamma'). \quad (3)$$

BICePs then samples the full posterior distribution  $P(X, \sigma_d | D)$  via Markov Chain Monte Carlo (MCMC) with moves in  $(X, \sigma_d, \gamma')$ . The BICePs prediction of the posterior populations is the marginal distribution  $P(X | D) = \int P(X, \sigma_d | D) d\sigma_d d\gamma'$ . Similarly, the marginal posterior distribution of the nuisance parameters can be obtained as  $P(\sigma_d | D) = \int P(X, \sigma_d, \gamma' | D) dX d\gamma'$  and  $P(\gamma' | D) = \int P(X, \sigma_d, \gamma' | D) dX d\sigma_d$ .

## BICePs results

The results of the BICePs algorithm are presented in Figure 5. Figure 5a shows a comparison of the BICePs-computed posterior state populations  $P(X | D)$  versus populations computed using only the likelihood function  $Q(D | X)$  reflecting the agreement with experimental restraints. Three conformational states, arbitrarily indexed as 11, 33 and 92, show predicted populations above 5 percent. These states lie above the diagonal in Figure 5a, indicating that these states have both low conformational free energies, and good agreement with the experimental restraints.

The marginal distribution  $P(\sigma_d)$  corresponds the average root-mean squared deviation of simulated interproton distances from the experimentally measured values, over the entire BICePs-sampled ensemble. Using a uniform prior  $P(X)$  ('exp', no theoretical population estimates)  $P(\sigma_d)$  is peaked around 0.6 nm, whereas inclusion of the tREMD-based populations ('sim+exp') results in  $P(\sigma_d)$  peaked around 0.7 nm. This indicates some tension between the simulated and experimental ensembles, due to (as we discuss below) a prefer-

ence for *cis*-amide backbones in the tREMD simulations, versus experimental restraints that favor *trans*-amide backbones. Marginal posterior distributions  $P(\gamma')$  are centered near 1.0 in both ‘exp’ and ‘sim+exp’ scenarios, indicating reasonable values of  $r_j^{\text{exp}}$  for the experimental distance restraints.

Inspection of three highest-population microstate conformational states predicted by the BICePs algorithm reveals a strong preference for a *trans*-amide backbone. Two of these states contain only one *cis*-amide in the backbone, and the other state contains all *trans*-amides.

To further visualize the effect of the experimental restraints on the simulated ensemble, we coarse-grained the 100 conformational microstates into 26 unique macrostate *cis/trans* amide isomers, each denoted by a 6-character string (for example, *ctcttt* has two of the three Q-proline residues with *cis*-amides; all other residues have *trans*-amides). Figure 6a shows the populations of the *cis/trans* isomers predicted from all temperature replicas (cation-free) tREMD simulations, Figure 6b shows the populations of the *cis/trans* isomers predicted from the lowest-three temperature replicas, and Figure 6c shows the refined populations from BICePs after incorporating the experimental distance restraints. A shift to higher numbers of *trans*-amides can clearly be observed at each step. The conformational state with the highest predicted population from BICePs is the *tctttt* state, which is entirely *trans*-amide except for one *cis*-amide Nme residue.

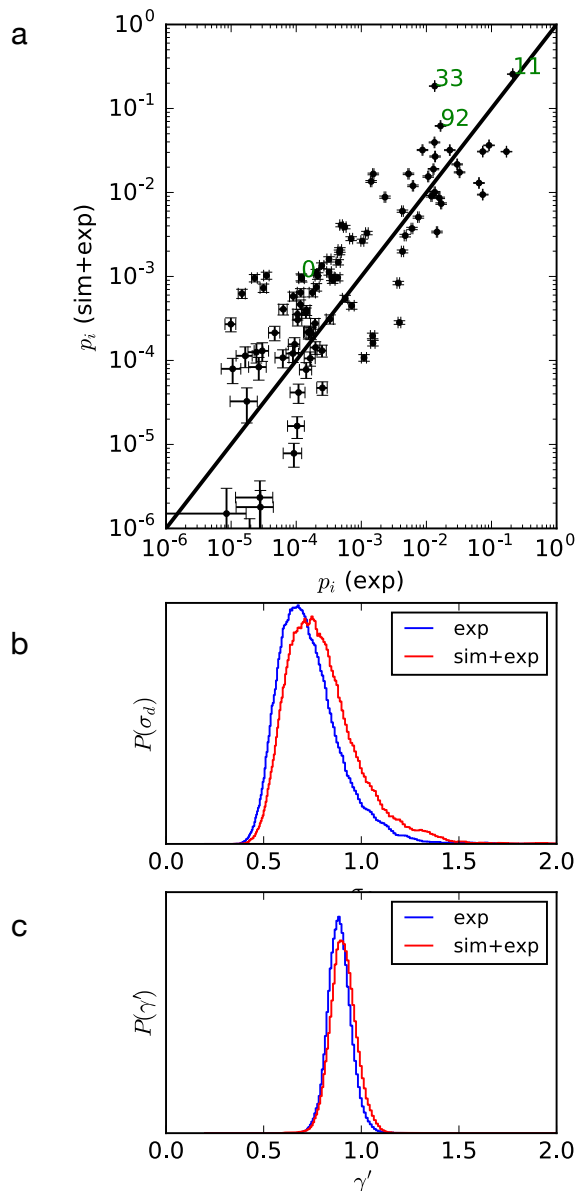


Figure 5: Results of the BICePs algorithm. (a) A comparison of estimated conformational state populations from BICePs when only experimental restraints are used (‘exp’), versus a combination of tREMD simulation and experiment (‘sim+exp’). States 11, 33, and 92 (labeled in green) have the highest predicted populations. (b) The marginal posterior distributions of  $P(\sigma_d)$  for ‘exp’ and ‘sim+exp’ scenarios. (c) The marginal posterior distributions for  $P(\gamma')$ .

## Explicit Solvent simulation of cation binding

The conformational state with the highest population predicted by BICePs (*tctttt*) was used as the initial starting structure for explicit-solvent simulations of macrocycles QPM-1,

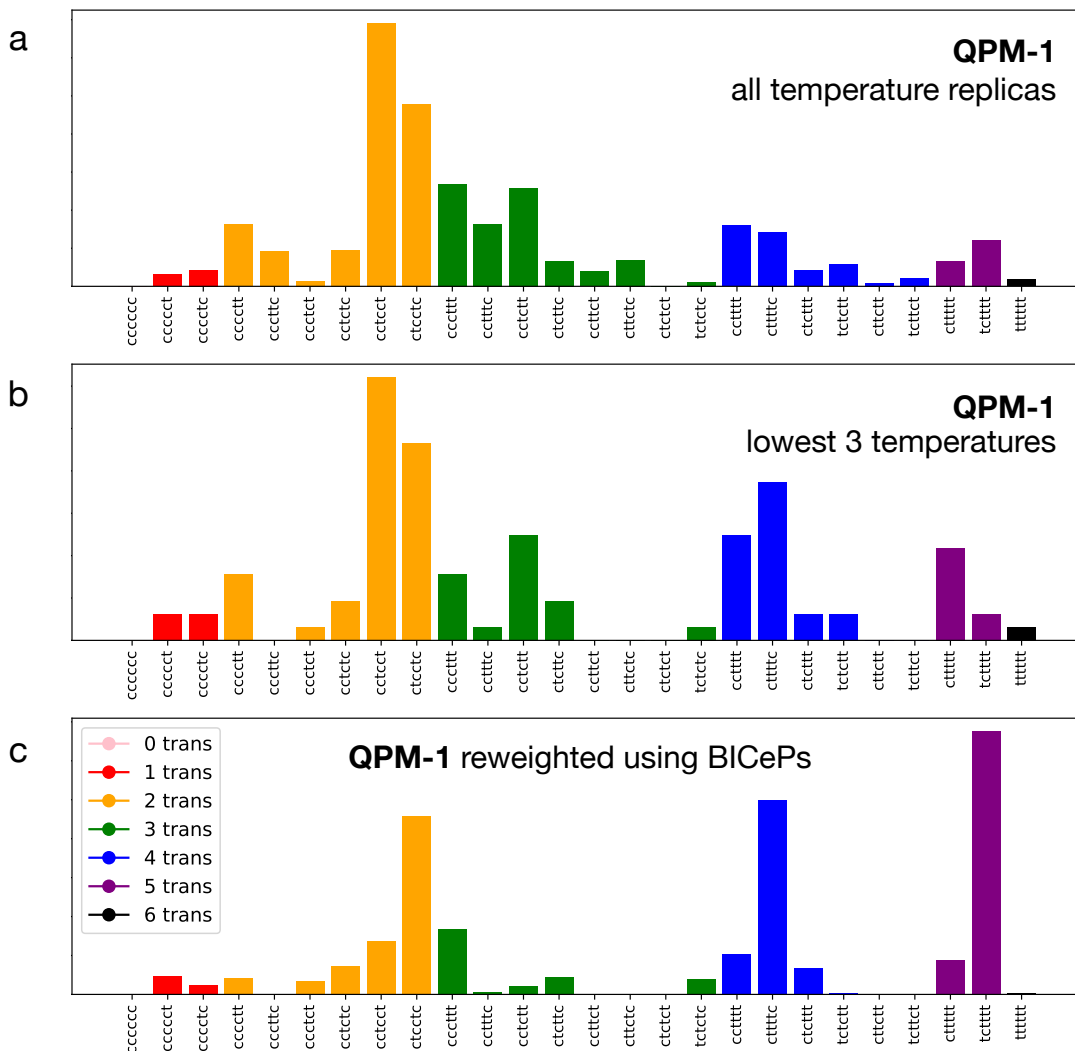


Figure 6: Distributions of backbone amide *cis/trans* isomers for QPM-1, shown for (a) all tREMD temperature replicas, (b) the three lowest-temperature replicas, and (c) after reweighting populations by the BICePs algorithm. The inclusion of experimental ROE restraints results in much higher population estimates for backbone *trans*-amides, with the mostly *trans* *tctttt* as the most populated state.

QPM-3 and QPM-9 in two different solvents (water and acetonitrile) and in the presence of two different cations ( $K^+$  or  $Na^+$ ). Simulations performed in acetonitrile were to enable comparison with the solvent conditions of the NMR experiments. Although no NMR data existed for QPM-9 (it did not exhibit ion-binding properties), it was nevertheless included as a control, to investigate whether our simulations could discriminate between binding-competent and -incompetent macrocycles.

All simulations were performed at a temperature of 300 K, with cation concentrations of 100 mM. Simulations in water used the TIP3P explicit solvent model, and chloride ( $\text{Cl}^-$ ) as the counterion. Simulations in acetonitrile used trifluoromethanesulfonate (triflate) as the counterion. Parameters for acetonitrile and triflate came from the Automated Topology Builder.<sup>62</sup>

The simulations were performed using the GROMACS molecular dynamics package<sup>56</sup> using Verlet integration on Temple’s Owlsnest HPC Cluster. For aqueous simulations, each macrocycle was placed in a cubic box of length 4.04 nm and sampled in the NVT ensemble using solvent and ion parameters from the *amber99sb-ildn* force field. The total number of particles in the system with four ion pairs and 2102 solvent molecules was 6464. Simulations in acetonitrile were performed in a cubic box of length 3.50 nm. Each of the 12 simulations (3 conformations  $\times$  2 cations  $\times$  2 solvents) produced an aggregate of nearly 5  $\mu\text{s}$  of trajectory data, with snapshots taken every 10 ps. Trajectory data was analyzed using Python3.6 packages MDTraj<sup>63</sup> and MSMBuilder<sup>64</sup> to calculate cation-macrocycle potentials of mean force (PMF), and to perform time-lagged independent component analysis (tICA).

Multiple cation binding events can be directly observed in the simulated explicit-solvent trajectories for QPM-1 and QPM-3. In agreement with experiment, simulations of the control, QPM-9, exhibited no cation binding, likely due to the bulkier spiroligomer substituents hindering cation access to the macrocycle center. Simulations of QPM-1 and QPM-3 in both water and acetonitrile show cation binding events coupled to a macrocycle conformational transition to an all-*trans* amide backbone state, which binds cations most strongly. In this conformation, electron density is concentrated in the center of the macrocycle, a model consistent with previous crystallographic studies of cation-bound hexameric cyclic peptoids.<sup>31</sup>

Potentials of mean force calculated for QPM-1 and QPM-3 along the average distance between cation and backbone carbonyls (Figure 7) shows that binding is most strongly dependent on the choice of solvent. At a simulated cation concentration of  $\sim 100$  mM in acetonitrile, cation binding is favorable, with macrocycle affinities between  $-7$  and  $-3 k_B T$ ,



while in aqueous solution, cation binding is less favorable, with affinities above  $+2 k_B T$ .

The next most important criterion for binding appears to be the macrocycle sequence. In agreement with experiment, the simulation results predict that QPM-1 has a higher cation affinity than QPM-3, by about  $1.5\text{-}2.0 k_B T$ , in both water and acetonitrile. A possible explanation for the higher affinity of cations for QPM-1 is the additional electron-donating groups made available by the methoxyethyl peptoid *N*-substituents, in contrast to the isopropyl *N*-substituents of QPM-3.

Finally, the simulations predict a preference for  $\text{Na}^+$  cations over  $\text{K}^+$  cations, in both water and acetonitrile. At the simulated cation concentrations, the magnitude of the preference is about  $1 k_B T$ . The preference for  $\text{Na}^+$  over  $\text{K}^+$  arises from the balance of two driving forces: (1) specific bound-state interactions for each ion, versus (2) the solvation free energy of each ion. Hydration free energies of  $\text{Na}^+$  and  $\text{K}^+$  are  $-365 \text{ kJ mol}^{-1}$  and  $-295 \text{ kJ mol}^{-1}$ , respectively.<sup>65</sup> Since  $\text{Na}^+$  hydration is more favorable than  $\text{K}^+$ , specific bound-state interactions between metal cation and macrocycle in water must be even *more* favorable for  $\text{Na}^+$ , likely due to the smaller ionic radius of  $\text{Na}^+$  (0.102 nm) versus  $\text{K}^+$  (0.138 nm).<sup>65</sup> Although solvation free energies of  $\text{Na}^+$  and  $\text{K}^+$  in acetonitrile are highly unfavorable ( $+665 \text{ kJ mol}^{-1}$  and  $+731 \text{ kJ mol}^{-1}$ , respectively)<sup>66</sup> the more favorable solvation of  $\text{Na}^+$  similarly implies the preference must be driven by specific bound-state interactions.

In Table 2 we convert concentration-dependent affinities derived the PMFs to standard binding free energies.

## Cation binding follows a conformational selection mechanism

To investigate the microscopic molecular mechanism by which cations bind the macrocycles, we performed tICA analysis of the trajectories simulated in explicit water (see Methods). The time-lagged independent components (tICs) correspond to the slowest dynamical motions of the cation-macrocycle binding reaction. Projection of the trajectory data to the two largest tICA components, tIC<sub>1</sub> and tIC<sub>2</sub>, shows three conformational states that slowly interconvert.

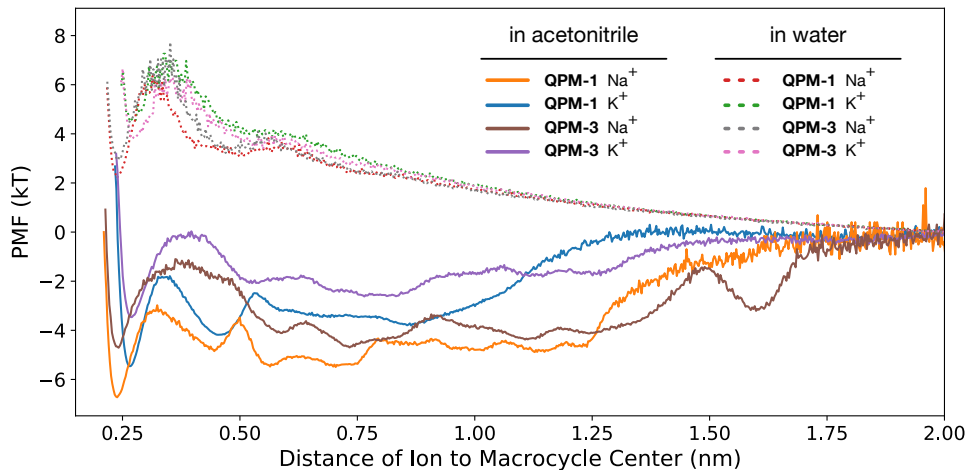


Figure 7: Potentials of mean force (PMF) as a function of cation distance to the centers of macrocycles QPM-1 and QPM-3, computed from explicit cation-binding trajectories simulated in acetonitrile and water.

In the case of water-solvated simulations, two of the states, positioned along tIC1, are cation-unbound conformations containing at least one *cis*-amide in the backbone that we designate as “open” and “closed” (Figure 8). Conversion between these states constitute the slowest conformational motions in the simulation. In the “closed” conformation, the central cation binding site is disrupted by hydrogen bonding interactions between sidechain hydantoin groups and the backbone amides. In the “open” conformation, the cation binding site is unobstructed.

A third macrocycle conformational state, which we call the “bound” state, has all *trans*-amides in the backbone. Transitions to the “bound” conformational state occur along tIC<sub>2</sub>, corresponding to the second-slowest motions in the trajectory data. The majority of observed cation binding events occur when the macrocycle is in this conformation (Figure 8, colored lines). This suggests a conformational selection mechanism, whereby strong cation binding only occurs after preorganization of the macrocycle to a binding-competent conformation. Interestingly, based on the conformational transitions observed in the trajectory data, the “open” conformation appears to be an obligate intermediate to formation of the “bound” conformation.

Table 2: Estimated free energies of cation binding to QPM-1 and QPM-3 for different counter anions and solvents, estimated from Potentials of Mean Force (PMFs) and ranked from highest to lowest binding affinity. The triflate anion (trifluoromethanesulfonate,  $\text{CF}_3\text{SO}_3^-$ ) is abbreviated  $\text{TFL}^-$ , and acetonitrile is abbreviated as ACN. Standard binding free energies  $\Delta G$  are reported in units  $k_B T$ , with uncertainties estimated using a bootstrap procedure (the standard deviation of 100 bootstrapped trajectories resampled with replacement). Dissociation constants  $K_d$  are given in units of mM.

Macrocycle	Cation	Anion	Solvent	$\Delta G$ ( $k_B T$ )	$K_d$ (mM)
QPM1	$\text{Na}^+$	$\text{TFL}^-$	ACN	$-8.02 \pm 0.52$	$0.37 \pm 0.18$
QPM1	$\text{K}^+$	$\text{TFL}^-$	ACN	$-7.52 \pm 0.25$	$0.56 \pm 0.13$
QPM3	$\text{Na}^+$	$\text{TFL}^-$	ACN	$-7.20 \pm 0.36$	$0.88 \pm 0.27$
QPM3	$\text{K}^+$	$\text{TFL}^-$	ACN	$-5.48 \pm 0.18$	$4.22 \pm 0.77$
QPM1	$\text{Na}^+$	$\text{Cl}^-$	Water	$-0.165 \pm 0.065$	$850 \pm 55$
QPM3	$\text{Na}^+$	$\text{Cl}^-$	Water	$0.520 \pm 0.086$	$1670 \pm 140$
QPM3	$\text{K}^+$	$\text{Cl}^-$	Water	$1.29 \pm 0.10$	$3650 \pm 360$
QPM1	$\text{K}^+$	$\text{Cl}^-$	Water	$1.89 \pm 0.14$	$6730 \pm 890$

Simulations of macrocycles QPM-1 and QPM-3 with cations in acetonitrile similarly show three slowly interconverting conformational states. Projection of trajectory data to the first two tICA components reveals a similar “opening” of the macrocycle along  $\text{tIC}_1$ . The “open” state quickly forms a bound-state complex with two cations from above and below the plane of the macrocycle, which induce a transition to an all-trans state (Figure 9).

The cation-bound conformational state of QPM-1 in acetonitrile is distinct from the bound state in water. On one side of the macrocycle plane, a “basket” of spiroligomer Q-prolines creates a volume in which triflate sulfonates help coordinate the cation. On the other side, a basket of peptoid residues help coordinate the second ion through their *N*-methoxyethyl groups. This two-cation bound-state conformation is observed in simulations of QPM-1 in the presence of  $\text{Na}^+$  and in  $\text{K}^+$  (Figure 10).

A general feature of the simulations of macrocycles in acetonitrile is the many compact, semi-crystalline arrangements of cations and triflate ions found near the macrocycle. This non-specific association is likely driven by the low solubility of ions in acetonitrile, just as is specific binding to the macrocycle center.

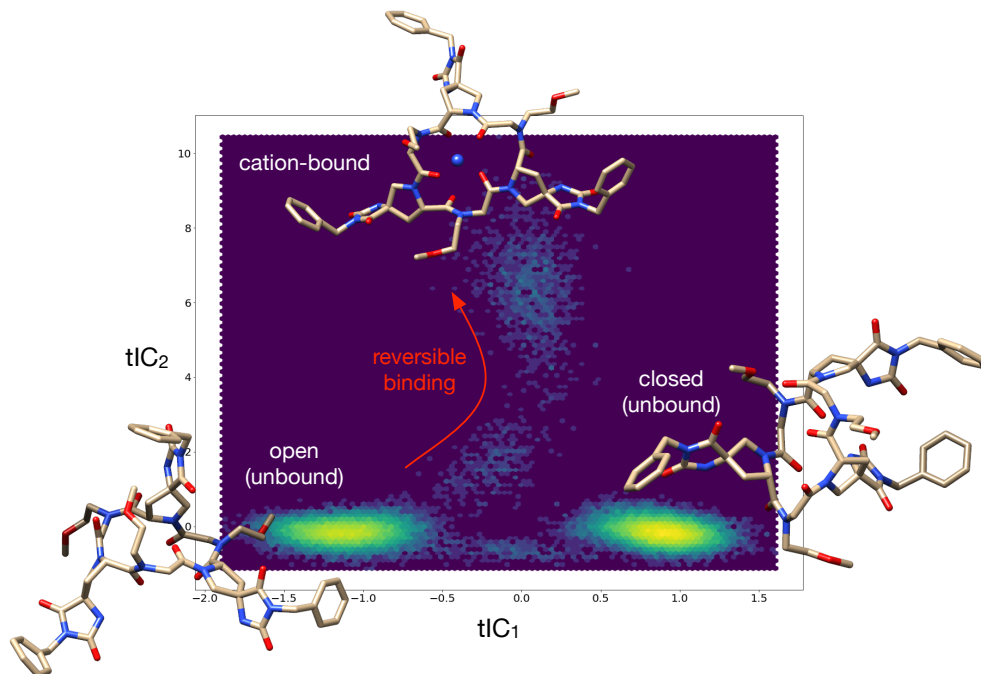


Figure 8: Projection of explicit-water QPM-1 trajectories to the first two tICA components reveals three slowly-interconverting conformational states: “open” and “closed” states with weak affinity to cations, and a “bound” conformation with stronger affinity to cation. Representative structures are taken from simulations of QPM-1 with  $\text{Na}^+$ . The slowest conformational motions (along  $\text{tIC}_1$ ) correspond to the formation of an unobstructed cation binding site, while the next slowest motion (along  $\text{tIC}_2$ ) corresponds to reversible cation binding to the “bound” conformation. Binding is defined as the cation being  $< 0.35$  nm from the center of macrocycle.

## Conclusion

In this work, we used a Bayesian inference approach called BICePs, along with implicit-solvent molecular simulation methods, to model the preorganized solution-state conformations of spirologomer-peptoid macrocycles when bound to metal cations. Our results show the BICePs approach is an extremely useful tool that reconcile simulations with sparse experimental observables, circumventing the need to bespoke force field parameterization.

Explicit-solvent simulations of the predicted macrocycle conformations show metal cation binding to an all-*trans* amide backbone. Impressively, simulations are able to correctly predict the solvent-, sequence- and ion-dependence of binding for all the macrocycles studied in this work, in agreement with experiment. No binding was observed for the negative control,

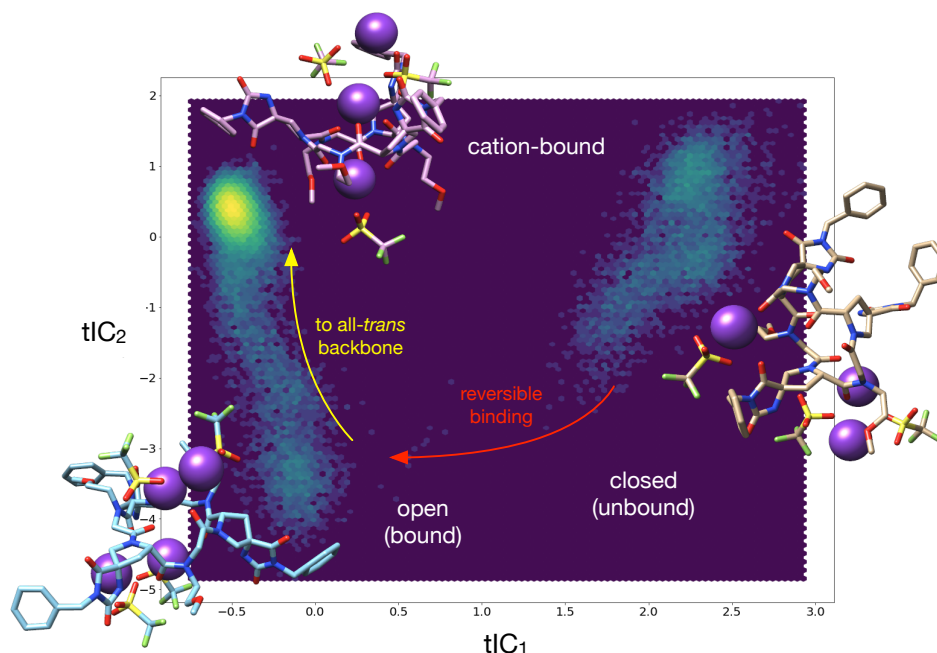


Figure 9: Projection of explicit-acetonitrile QPM-1 trajectories to the first two tICA components similarly reveals three metastable conformational states: “open” and “closed” states, which contain at least one *cis* amide, and an *all-trans* “bound” conformation with stronger affinity to cation. Representative structures are taken from simulations of QPM-1 with  $K^+$ . The slowest conformational motions (along  $tIC_1$ ) correspond to the formation of an unobstructed cation binding site, while the next slowest motion (along  $tIC_2$ ) corresponds to reversible cation binding to the “bound” conformation. Binding is defined as the cation being  $< 0.35$  nm from the center of macrocycle.

#### QPM-9.

One advantage our method has over structural determination methods such as crystallography is that it is designed to predict solution-phase properties, which may be particularly important for Q-proline macrocycles. An *apo* (racemic) crystal structure of QPM-3 was found to be distinctly different from the solution structures we found in this work.<sup>37</sup> While crystal structures of similar macrocycles show *all-trans* backbone amides, they are affected by packing artifacts (i.e. stacking interactions) that would not be present in solution.<sup>31</sup>

Overall, these results strongly suggest that simulation-based methods are up to the challenging task of virtual screening and selection for computational design of macrocycle scaffolds. Given the huge chemical diversity available for functionalizing spiroligomer-peptoid

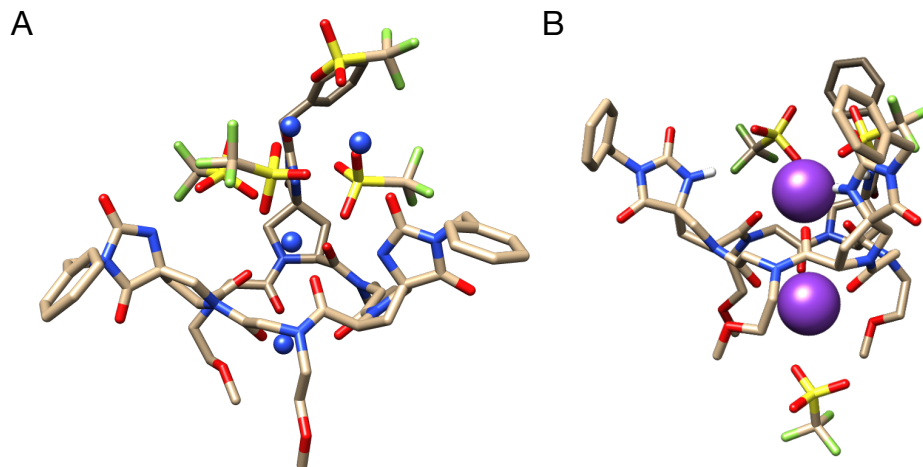


Figure 10: Simulations of QPM-1 in acetonitrile reveals a highly stable, two-cation structure that forms in the presence of both (A) sodium and (B) potassium cations. The bound conformation features all-trans backbone amides, with one cation above the macrocycle plane in a “basket” of Q-proline residues coordinated by triflate sulfonate groups, and the second cation below the plane, stabilized by coordinating peptoid *N*-methoxyethyl substituents.

macrocycles, we hope in the future to use efficient simulation-based computational design to develop macrocycles for many new applications.

## Acknowledgments

This research was supported in part by National Institutes of Health grant 1R01GM123296. Calculations were performed on HPC resources supported in part by the National Science Foundation through major research instrumentation grant number 1625061 and by the US Army Research Laboratory under contract number W911NF-16-2-0189. Research was also performed on the CB2RR cluster made possible through NIH Research Resource computer instrumentation grant S10-OD020095.

## Data and Software Availability

A complete set of scripts for preparing molecular systems and simulations, preparing and performing the BICePs analysis, and analysis of simulated trajectories is freely available at <https://github.com/yabmtm/macrocycles>. Simulations were performed with the open-source molecular dynamics package GROMACS, which can be downloaded at <http://gromacs.org>. The BICePs algorithm is freely available at <https://github.com/vvoelz/biceps>. Trajectory data is archived at <https://doi.org/10.5281/zenodo.4430604>.

## References

- (1) Martí-Centelles, V.; Pandey, M. D.; Burguete, M. I.; Luis, S. V. Macrocyclization reactions: The importance of conformational, configurational, and template-induced preorganization. 2015.
- (2) Rajagopalan, P. T.; Benkovic, S. J. Preorganization and protein dynamics in enzyme catalysis. *Chemical Record* **2002**,
- (3) Schafmeister, C. E.; Brown, Z. Z.; Gupta, S. Shape-programmable macromolecules. *Accounts of Chemical Research* **2008**,
- (4) Brown, Z. Z.; Schafmeister, C. E. Synthesis of hexa-and pentasubstituted diketopiperazines from sterically hindered amino acids. *Organic letters* **2010**, *12*, 1436–1439.
- (5) Northrup, J.; Purcell, C. R.; Schafmeister, C. E. One-Pot Synthesis of Chiral, Spirocyclic 4-Hydantoin-Proline Derivatives for Incorporation into Spiroligomer-Based Macromolecules. *The Journal of organic chemistry* **2017**, *82*, 3223–3231.
- (6) Brown, Z. Z.; Akula, K.; Arzumanyan, A.; Alleva, J.; Jackson, M.; Bichenkov, E.; Sheffield, J. B.; Feitelson, M. A.; Schafmeister, C. E. A Spiroligomer  $\alpha$ -Helix Mimic

- That Binds HDM2, Penetrates Human Cells and Stabilizes HDM2 in Cell Culture. *PLoS ONE* **2012**,
- (7) Kheirabadi, M.; Çelebi-Ölçüm, N.; Parker, M. F.; Zhao, Q.; Kiss, G.; Houk, K. N.; Schafmeister, C. E. Spiroligozymes for transesterifications: Design and relationship of structure to activity. *Journal of the American Chemical Society* **2012**,
  - (8) Parker, M. F.; Osuna, S.; Bollot, G.; Vaddypally, S.; Zdilla, M. J.; Houk, K.; Schafmeister, C. E. Acceleration of an aromatic Claisen rearrangement via a designed spiroligolyse catalyst that mimics the ketosteroid isomerase catalytic dyad. *Journal of the American Chemical Society* **2014**, *136*, 3817–3827.
  - (9) Vaddypally, S.; Xu, C.; Zhao, S.; Fan, Y.; Schafmeister, C. E.; Zdilla, M. J. Architectural spiroligomers designed for binuclear metal complex templating. *Inorganic Chemistry* **2013**,
  - (10) Sun, J.; Zuckermann, R. N. Peptoid polymers: a highly designable bioinspired material. *ACS Nano* **2013**, *7*, 4715–4732.
  - (11) Gellman, S. H. Foldamers: a manifesto. *Accounts of Chemical Research* **1998**, *31*, 173–180.
  - (12) Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. A Field Guide to Foldamers. *Chemical Reviews* **2001**, *101*, 3893–4012.
  - (13) Wu, C. W.; Sanborn, T. J.; Huang, K.; Zuckermann, R. N.; Barron, A. E. Peptoid Oligomers with  $\alpha$ -Chiral, Aromatic Side Chains: Sequence Requirements for the Formation of Stable Peptoid Helices. *Journal of the American Chemical Society* **2001**, *123*, 6778–6784.
  - (14) Stringer, J. R.; Crapster, J. A.; Guzei, I. A.; Blackwell, H. E. Extraordinarily robust polyproline type I peptoid helices generated via the incorporation of  $\alpha$ -chiral aromatic



- N-1-naphthylethyl side chains. *Journal of the American Chemical Society* **2011**, *133*, 15559–15567.
- (15) Mukherjee, S.; Zhou, G.; Michel, C.; Voelz, V. A. Insights into Peptoid Helix Folding Cooperativity from an Improved Backbone Potential. *Journal of Physical Chemistry B* **2015**, *119*, 15407–15417.
- (16) Gimenez, D.; Zhou, G.; Hurley, M. F. D.; Aguilar, J. A.; Voelz, V. A.; Cobb, S. L. Fluorinated Aromatic Monomers as Building Blocks To Control  $\alpha$ -Peptoid Conformation and Structure. *Journal of the American Chemical Society* **2019**, *141*, 3430–3434.
- (17) Crapster, J. A.; Guzei, I. A.; Blackwell, H. E. A peptoid ribbon secondary structure. *Angewandte Chemie International Edition in English* **2013**, *52*, 5079–5084.
- (18) Gorske, B. C.; Mumford, E. M.; Gerrity, C. G.; Ko, I. A Peptoid Square Helix via Synergistic Control of Backbone Dihedral Angles. *Journal of the American Chemical Society* **2017**, *139*, 8070–8073.
- (19) Wijaya, A. W.; Nguyen, A. I.; Roe, L. T.; Butterfoss, G. L.; Spencer, R. K.; Li, N. K.; Zuckermann, R. N. Cooperative Intramolecular Hydrogen Bonding Strongly Enforces cis-Peptoid Folding. *Journal of the American Chemical Society* **2019**, *141*, 19436–19447.
- (20) Schafmeister, C. E.; Po, J.; Verdine, G. L. An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. *Journal of the American Chemical Society* **2000**, *122*, 5891–5892.
- (21) Robinson, J. A.  $\beta$ -Hairpin Peptidomimetics: Design, Structures and Biological Activities. *Accounts of Chemical Research* **2008**, *41*, 1278–1288.
- (22) Chatterjee, J.; Rechenmacher, F.; Kessler, H. N-Methylation of Peptides and Proteins:

- An Important Element for Modulating Biological Functions. *Angewandte Chemie International Edition* **2012**, *52*, 254–269.
- (23) 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.
- (24) Siegert, T. R.; Bird, M. J.; Makwana, K. M.; Kritzer, J. A. Analysis of Loops that Mediate Protein-Protein Interactions and Translation into Submicromolar Inhibitors. *Journal of the American Chemical Society* **2016**, *138*, 12876–12884.
- (25) Zhao, Q.; Schafmeister, C. E. Synthesis of Spiroligomer-Containing Macrocycles. *Journal of Organic Chemistry* **2015**,
- (26) Shin, S. B. Y.; Yoo, B.; Todaro, L. J.; Kirshenbaum, K. Cyclic Peptoids. *Journal of the American Chemical Society* **2007**, *129*, 3218–3225.
- (27) Maayan, G.; Ward, M. D.; Kirshenbaum, K. Metallopeptoids. *Chemical Communications* **2008**, 56.
- (28) Yoo, B.; Shin, S. B. Y.; Huang, M. L.; Kirshenbaum, K. Peptoid macrocycles: making the rounds with peptidomimetic oligomers. *Chemistry - A European Journal* **2010**, *16*, 5528–5537.
- (29) Vollrath, S. B. L.; Bräse, S.; Kirshenbaum, K. Twice tied tight: Enforcing conformational order in bicyclic peptoid oligomers. *Chemical Science* **2012**,
- (30) 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. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 14320–14325.

- (31) Izzo, I.; Ianniello, G.; De Cola, C.; Nardone, B.; Erra, L.; Vaughan, G.; Tedesco, C.; De Riccardis, F. Structural Effects of Proline Substitution and Metal Binding on Hexameric Cyclic Peptoids. *Organic Letters* **2013**, *15*, 598–601.
- (32) Meli, A.; Macedi, E.; De Riccardis, F.; Smith, V. J.; Barbour, L. J.; Izzo, I.; Tedesco, C. Solid-State Conformational Flexibility at Work: Zipping and Unzipping within a Cyclic Peptoid Single Crystal. *Angewandte Chemie* **2016**, *128*, 4757–4760.
- (33) D’Amato, A.; Volpe, R.; Vaccaro, M. C.; Terracciano, S.; Bruno, I.; Tosolini, M.; Tedesco, C.; Pierri, G.; Tecilla, P.; Costabile, C.; Della Sala, G.; Izzo, I.; De Riccardis, F. Cyclic Peptoids as Mycotoxin Mimics: An Exploration of Their Structural and Biological Properties. *The Journal of Organic Chemistry* **2017**, *82*, 8848–8863.
- (34) Tedesco, C.; Schettini, R.; Iuliano, V.; Pierri, G.; Fitch, A. N.; De Riccardis, F.; Izzo, I. Role of Side Chains in the Solid State Assembly of Cyclic Peptoids. *Crystal Growth & Design* **2018**, *19*, 125–133.
- (35) Maulucci, N.; Izzo, I.; Bifulco, G.; Aliberti, A.; De Cola, C.; Comegna, D.; Gaeta, C.; Napolitano, A.; Pizza, C.; Tedesco, C. Synthesis, structures, and properties of nine-, twelve-, and eighteen-membered N-benzyloxyethyl cyclic  $\alpha$ -peptoids. *Chemical Communications* **2008**, 3927–3929.
- (36) Northrup, J. D.; Mancini, G.; Purcell, C. R.; Schafmeister, C. E. Development of Spiroligomer-Peptoid Hybrids. *Journal of Organic Chemistry* **2017**,
- (37) Northrup, J.; Wiener, J.; Hurley, M.; Hou, C.-F. D.; Keller, T.; Baxter, R.; Zdilla, M. J.; Voelz, V.; Schafmeister, C. Metal-Binding Q-Proline Macrocycles. 2021; [https://chemrxiv.org/articles/preprint/Metal-Binding\\_Q-Proline\\_Macrocycles/13554731/1](https://chemrxiv.org/articles/preprint/Metal-Binding_Q-Proline_Macrocycles/13554731/1).
- (38) Renfrew, P. D.; Craven, T. W.; Butterfoss, G. L.; Kirshenbaum, K.; Bonneau, R. A

- rotamer library to enable modeling and design of peptoid foldamers. *Journal of the American Chemical Society* **2014**, *136*, 8772–8782.
- (39) Butterfoss, G. L.; Renfrew, P. D.; Kuhlman, B.; Kirshenbaum, K.; Bonneau, R. A Preliminary Survey of the Peptoid Folding Landscape. *Journal of the American Chemical Society* **2009**, *131*, 16798–16807.
- (40) Nguyen, A. I.; Spencer, R. K.; Anderson, C. L.; Zuckermann, R. N. A bio-inspired approach to ligand design: folding single-chain peptoids to chelate a multimetallic cluster. *Chemical Science* **2019**, *114*, 4175–8.
- (41) Roe, L. T.; Pelton, J. G.; Edison, J. R.; Butterfoss, G. L.; Tresca, B. W.; LaFaye, B. A.; Whitelam, S.; Wemmer, D. E.; Zuckermann, R. N. Unconstrained peptoid tetramer exhibits a predominant conformation in aqueous solution. *Biopolymers* **2019**, *8*, e23267–10.
- (42) Shandler, S. J.; Shapovalov, M. V.; Dunbrack, R. L., Jr; DeGrado, W. F. Development of a rotamer library for use in  $\beta$ -peptide foldamer computational design. *Journal of the American Chemical Society* **2010**, *132*, 7312–7320.
- (43) Drew, K.; Renfrew, P. D.; Craven, T. W.; Butterfoss, G. L.; Chou, F.-C.; Lyskov, S.; Bullock, B. N.; Watkins, A.; Labonte, J. W.; Pacella, M., et al. Adding diverse non-canonical backbones to rosetta: enabling peptidomimetic design. *PLoS One* **2013**, *8*, e67051.
- (44) Watkins, A. M.; Craven, T. W.; Renfrew, P. D.; Arora, P. S.; Bonneau, R. Rotamer Libraries for the High-Resolution Design of  $\beta$ -Amino Acid Foldamers. *Structure* **2017**, *25*, 1771–1780.
- (45) 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.

- (46) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* **1999**, *314*, 141–151.
- (47) Weiser, L. J.; Santiso, E. E. A CGenFF-based force field for simulations of peptoids with both cis and trans peptide bonds. *Journal of computational chemistry* **2019**,
- (48) Mirijanian, D. T.; Mannige, R. V.; Zuckermann, R. N.; Whitlam, S. Development and use of an atomistic CHARMM-based forcefield for peptoid simulation. *Journal of computational chemistry* **2014**, *35*, 360–370.
- (49) Christianson, L. A.; Lucero, M. J.; Appella, D. H.; Klein, D. A.; Gellman, S. H. Improved treatment of cyclic  $\beta$ -amino acids and successful prediction of  $\beta$ -peptide secondary structure using a modified force field: AMBER\* C. *Journal of Computational Chemistry* **2000**, *21*, 763–773.
- (50) Galan, J. F.; Tang, C. N.; Chakrabarty, S.; Liu, Z.; Moyna, G.; Pophristic, V. Conformational preferences of furan- and thiophene-based arylamides: A combined computational and experimental study. *Physical Chemistry Chemical Physics* **2013**,
- (51) Voelz, V. A.; Zhou, G. Bayesian inference of conformational state populations from computational models and sparse experimental observables. *Journal of Computational Chemistry* **2014**, *35*, 2215–2224.
- (52) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *Journal of Computational Chemistry* **2004**,
- (53) Schwantes, C. R.; Pande, V. S. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *Journal of chemical theory and computation* **2013**, *9*, 2000–2009.
- (54) Ge, Y.; Voelz, V. A. Model Selection Using BICePs: A Bayesian Approach for Force

- Field Validation and Parameterization. *The Journal of Physical Chemistry B* **2018**, *122*, 5610–5622.
- (55) Case, D. et al. Amber 2016. 2016.
- (56) 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.
- (57) Sousa Da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* **2012**,
- (58) Sui, Q.; Borchardt, D.; Rabenstein, D. L. Kinetics and equilibria of cis/trans isomerization of backbone amide bonds in peptoids. *Journal of the American Chemical Society* **2007**,
- (59) 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 Genetics* **2004**,
- (60) Liang, H.; Zhou, G.; Ge, Y.; D’Ambrosio, E. A.; Eidem, T. M.; Blanchard, C.; Shehatou, C.; Chatare, V. K.; Dunman, P. M.; Valentine, A. M.; Voelz, V. A.; Grimes, C. L.; Andrade, R. B. Elucidating the inhibition of peptidoglycan biosynthesis in *Staphylococcus aureus* by albocycline, a macrolactone isolated from *Streptomyces maizeus*. *Bioorganic & Medicinal Chemistry* **2018**, *26*, 3453–3460.
- (61) Wan, H.; Ge, Y.; Razavi, A.; Voelz, V. Reconciling Simulated Ensembles of Apomyoglobin with Experimental Hydrogen/Deuterium Exchange Data Using Bayesian Inference and Multiensemble Markov State Models. *Journal of Chemical Theory and Computation* **2020**, *16*, 1333–1348.

- (62) Koziara, K.; Stroet, M., M.; A.K., Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. *Journal of Computationally Aided Molecular Design* **2014**, *28*, 221—233.
- (63) 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.
- (64) Harrigan, M. P.; Sultan, M. M.; Hernández, C. X.; Husic, B. E.; Eastman, P.; Schwantes, C. R.; Beauchamp, K. A.; McGibbon, R. T.; Pande, V. S. MSMBuilder: statistical models for biomolecular dynamics. *Biophysical journal* **2017**, *112*, 10–15.
- (65) Marcus, Y. Thermodynamics of solvation of ions. Part 5.—Gibbs free energy of hydration at 298.15 K. *Journal of the Chemical Society, Faraday Transactions* **1991**, *87*, 2995–2999.
- (66) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Single-ion solvation free energies and the normal hydrogen electrode potential in methanol, acetonitrile, and dimethyl sulfoxide. *The Journal of Physical Chemistry B* **2007**, *111*, 408–422.