



Published in final edited form as:

*J Chem Inf Model.* 2021 May 24; 61(5): 2353–2367. doi:10.1021/acs.jcim.1c00029.

## Solution-state preorganization of cyclic $\beta$ -hairpin ligands determines binding mechanism and affinities for MDM2

Yunui Ge<sup>†</sup>, Si Zhang<sup>†</sup>, Mate Erdelyi<sup>‡</sup>, Vincent A. Voelz<sup>†</sup>

<sup>†</sup>Department of Chemistry, Temple University, Philadelphia, PA 19122, USA

<sup>‡</sup>Department of Chemistry - BMC, Uppsala University, SE-75123 Uppsala, Sweden

### Abstract

Understanding mechanisms of protein folding and binding is crucial to designing their molecular function. Molecular dynamics (MD) simulations and Markov State Model (MSM) approaches provide a powerful way to understand complex conformational change that occurs over long timescales. Such dynamics are important for the design of therapeutic peptidomimetic ligands, whose affinity and binding mechanism are dictated by a combination of folding and binding. To examine the role of preorganization in peptide binding to protein targets, we performed massively parallel explicit-solvent MD simulations of cyclic  $\beta$ -hairpin ligands designed to mimic the p53 transactivation domain and competitively bind MDM2. Disrupting the MDM2-p53 interaction is a therapeutic strategy to prevent degradation of the p53 tumor suppressor in cancer cells. MSM analysis of over 3 ms of aggregate trajectory data enabled us to build a detailed mechanistic model of coupled folding and binding of four cyclic peptides which we compare to experimental binding affinities and rates. The results show a striking relationship between the relative preorganization of each ligand in solution and its affinity for MDM2. Specifically, changes in peptide conformational populations predicted by the MSMs suggest that entropy loss upon binding is the main factor influencing affinity. The MSMs also enable detailed examination of non-native interactions

voelz@temple.edu .

#### ASSOCIATED CONTENT

##### Supporting Information Available

Supporting information (simulation details, distribution of distances between ligands and MDM2, experimental and computed NOE distances, distribution of trajectory lengths, implied timescale plots, generalized matrix Rayleigh quotient, tICA plots of binding/ bound-state/umbrella simulations, contact maps of large computed Bayes Factors, comparison of experimental and simulated NOE distances) is available free of charge via the Internet at <http://pubs.acs.org>.

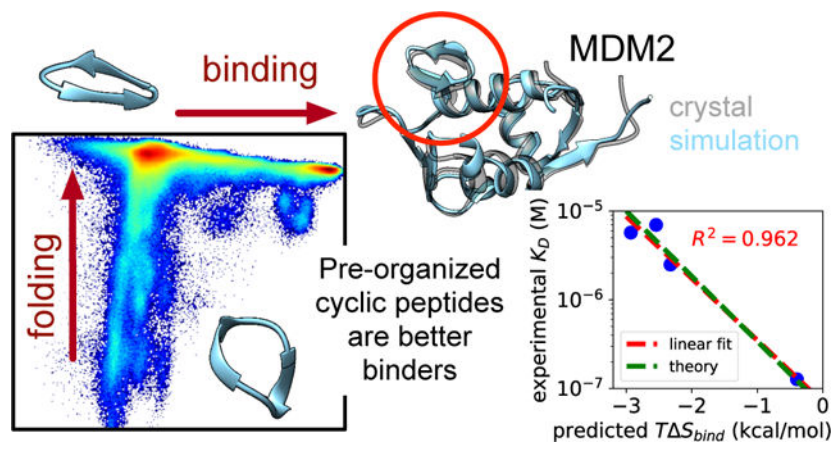
##### Data and Software Availability

**Datasets.** All input files, setup and subset of simulated trajectory data (used for analysis) are available at: <https://doi.org/10.5281/zenodo.3777274> (solution-state simulation trajectories of all peptides, <https://doi.org/10.5281/zenodo.3780463> (bound-state trajectory data peptides 1, 2 and 3), <https://doi.org/10.5281/zenodo.3782629> (bound-state trajectory data for peptide 4), <https://doi.org/10.5281/zenodo.3780467> (ligand-binding trajectories of all peptides), <https://doi.org/10.5281/zenodo.3814057> (umbrella-sampling trajectory data).

**Software.** Simulations were performed with the open-source molecular dynamics package GROMACS, which can be downloaded at <http://www.gromacs.org>. The MSMBuild3 software package was used to construct MSMs and is freely distributed at <https://github.com/msmbuilder/msmbuilder>. The PyEMMA software package was used to construct Multiensemble Markov models and can be freely downloaded at <https://github.com/markovmodel/PyEMMA>. Monte Carlo conformational searches were performed using the software MacroModel (v.9.1) as implemented in the Schrödinger package (<https://www.schrodinger.com/products/maestro>), with the parameters being given in reference.<sup>40</sup> NMR spectra for the NAMFIS analyses were recorded on a 900 MHz Bruker Avance III HD NMR spectrometer using the software Topspin v4 (<https://www.bruker.com/service/support-upgrades/software-downloads/nmr.html>), and were processed using the software MestReNova version 11.0 (<https://mestrelab.com/download/mnova/>). The NAMFIS analysis was performed using the NAMFIS algorithm,<sup>41</sup> which is available upon request from the authors.

which lead to misfolded states, and comparison of structural ensembles with experimental NMR measurements. In contrast to an MSM study of p53 TAD binding to MDM2, MSMs of cyclic  $\beta$ -hairpin binding show a conformational selection mechanism. Finally, we make progress towards predicting accurate off-rates of cyclic peptides using multiensemble Markov models (MEMMs) constructed from unbiased and biased simulated trajectories.

## Graphical Abstract



## INTRODUCTION

Understanding the conformational dynamics that underlies protein folding and binding is crucial to designing their molecular function. Molecular dynamics (MD) simulations have steadily become a powerful tool to gain mechanistic understanding of complex processes at the atomic level. Thanks to advances in computer hardware and enhanced sampling strategies, MD simulations can now routinely model conformational dynamics on the microsecond timescale, which is crucial for understanding the significant conformational rearrangements that peptides can undergo, and their association with protein targets.

Cyclic peptides have many beneficial properties for therapeutic development over their linear counterparts, including improved physical and chemical stability, binding affinity and membrane permeability. Much of these benefits are believed to arise through their intrinsic conformational preorganization in solution.<sup>1-6</sup>

A well-studied protein-peptide interaction targeted by cyclic peptides is between the E3 ubiquitin ligase MDM2 (mouse double minute 2 homolog) and the tumor suppressor protein p53, a potent transcription factor which plays a multifaceted role in the cell cycle.<sup>7</sup> The level and activity of wild-type p53 in the cell is controlled by MDM2, its negative regulator,<sup>8-11</sup> both by binding to the transactivation domain (TAD) of p53 and inhibiting its transcriptional activity,<sup>12</sup> and by targeting p53 for degradation in the proteasome.<sup>13</sup> In many cancers (about 50%), wild-type p53 is retained, but cannot mediate tumor suppression because of over-expression of MDM2. Therefore, many cancer therapeutic peptidomimetics have been developed to disrupt the MDM2/p53 interaction by competitively binding to MDM2.

The rational design of peptidomimetic ligands targeting MDM2 has benefited greatly from the existence of a crystal structure of p53 TAD (residues 15–29) bound to MDM2,<sup>12</sup> which shows the intrinsically disordered p53 TAD<sup>14</sup> bound to the hydrophobic cleft of MDM2 in a helical conformation, with p53 TAD residues Phe19, Trp23 and Leu26 filling the cleft. In addition to small molecule binders,<sup>15,16</sup> a number of linear and cyclic binders of MDM2 have been developed, including stapled peptides,<sup>17,18</sup>  $\beta$ -peptides,<sup>19</sup> cyclic  $\beta$ -hairpins<sup>20–23</sup> and D-peptides.<sup>24–26</sup> Cyclic  $\beta$ -hairpins ligands, pioneered by the Robinson group, represent a particularly attractive scaffold, stabilized by cyclization, intra-strand hydrogen-bonding and D-Pro/L-Pro capping groups that mimic  $\beta$ -turns.

A key design principle for cyclic peptides has been the idea that cyclization and crosslinking should enhance structural preorganization, which should in turn enhance binding affinity by decreasing the unfavorable conformational entropy change upon binding.<sup>6,21,27–33</sup> This idea has been borne out for many different helical peptides. Helical stapled-peptides based on p53 bind with higher affinity than linear designs.<sup>17,34–36</sup> Mutations of p53 TAD that enhance helicity also have higher affinity to MDM2 and affect signaling in cells.<sup>37</sup> A Markov State Model (MSM) of p53 TAD-MDM2 binding reaction, built from 831  $\mu$ s of aggregate trajectory data, reveals a mechanism in which binding of p53 TAD precedes folding, which predicts increased affinity as peptide helicity increases.<sup>38</sup> An MSM of MDM2 binding for a more helical, high-affinity designed peptide PMI<sup>24</sup> shows a mixture of induced-fit and conformational selection pathways.<sup>39</sup>

For cyclic  $\beta$ -hairpin binders of MDM2, the relationship between solution-state preorganization and binding properties is less clear. Unlike intrinsically disordered peptides, cyclic  $\beta$ -hairpins are partially folded in solution, and detailed characterization of the extent of their preorganization must come from careful comparison of biophysical measurements and simulated conformational ensembles. Recent work from Danelius et al. exploring the relationship between the rigidity of four cyclic  $\beta$ -hairpins and binding affinity to MDM2 offers insight into this question.<sup>40</sup> The first of these peptides (peptide 1) was originally designed by Fasan et al.<sup>23</sup> to mimic p53 TAD by displaying residues Phe, 6-chloro-Trp and Leu in a similar geometry (Figure 1). Peptides 2, 3 and 4 (Figure 1b) replace the D-Pro/L-Pro capping motif with a D-Pro-Gly  $\beta$ -turn type II' motif, substitute smaller polar residues (i.e. Glu  $\rightarrow$  Ser), and explore different halogenated aromatic substitutions. Danelius et al. used solution-state NMR, along with the NAMFIS algorithm,<sup>41</sup> to characterize the structural heterogeneity of each peptide, as well as the affinity of the four designs, measured using SPR and fluorescence polarization assays.

Here, to further explore the relationship between preorganization, binding affinity and binding mechanism, we perform massively parallel all-atom molecular simulations of the four peptides, both in solution, and in the presence of MDM2. By collecting and analyzing a large number of *ab initio* binding trajectories, we characterize both folding and binding mechanisms in microscopic detail.

## METHODS

### Molecular Simulation of Cyclic Hairpin Designs.

Cyclic peptides were constructed using the AmberTools *tleap* program. Topologies for non-natural amino acids were built using UCSF Chimera and parameterized using the Generalized AMBER force field (GAFF),<sup>42</sup> with partial charges computed using the AM1-BCC<sup>43</sup> methods. The ACPYPE program<sup>44</sup> was then used to convert the topology file format for use with the GROMACS. MD simulations were performed using GROMACS 4.5.<sup>45</sup> on the Folding@home distributed computing platform.<sup>46</sup> The AMBER ff99sb-ildn-NMR<sup>47</sup> force field was used in combination with the TIP3P explicit solvent model. Cubic periodic boxes were filled with solvated protein and counterions (0.1 M NaCl) to neutralize the system. A full list of particle numbers and box sizes can be found in Table S1. Simulations were minimized and equilibrated at 300 K at a constant pressure of 1 atm for 200 ps using a Berendsen thermostat and barostat with time constant 1 ps and compressibility  $4.5 \times 10^{-5}$  bar<sup>-1</sup>. Trajectory data was generated using constant-volume molecular dynamics at 300 K, a stochastic (Langevin) integration with a 2 fs time step, and friction constant 1 ps<sup>-1</sup>.

Simulations of MDM2-bound cyclic  $\beta$ -hairpin peptide 1 were initiated from the crystal structure pose (PDBid: 2axi). Simulations of peptides 2, 3 and 4 were initiated from homology models built from the crystal structure template. Simulations of *ab initio* peptide binding to MDM2 were initiated from ten different starting configurations where the cyclic peptide is randomly inserted in the simulation box, and MDM2 is placed the center of the box. The distance between the center of mass of MDM2 and the peptide ligand was chosen to be at least 3.0 nm, the distance beyond which we consider the peptide to be unbound (Figure S1). Over 3 ms of aggregate trajectory data was collected for analysis (Figure S2). The total simulation time and average trajectory length for each simulation can be found in Table S2–S3.

### Umbrella simulations of cyclic hairpins.

To enhance sampling of unbinding events, a series of umbrella simulations of peptide 1 were performed. A harmonic restraint with force constant  $200 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  was enforced for the distance between the  $\beta$ -carbon ( $C\beta$ ) of 6-Cl-Trp on peptide 1, and the  $\alpha$ -carbon ( $C\alpha$ ) of Gly58 on MDM2. Separate simulations were performed for a series of 27 different equilibrium distances ranging from 0.5 nm to 3.1 nm, in increments of 0.1 nm, resulting in ~ 56  $\mu\text{s}$  of aggregate trajectory data (Table S4).

### Markov State Model (MSM) Construction.

MSM approaches have been used with great success to model the conformational dynamics of protein folding and binding as a kinetic network of metastable conformational states, providing the atomistic details to help complement and interpret experimental measurements.<sup>48–50</sup> A summary of good practices for constructing MSMs in ligand binding mechanisms study can be found in this work.<sup>51</sup> A key advantage of MSMs is the ability to infer information about folding thermodynamics and kinetics from ensembles of short, nonequilibrium trajectories.

The MSMBuild<sup>352</sup> and PyEMMA<sup>53</sup> software packages were used to construct MSMs from the simulation trajectory data. This involves the following steps: (1) featurizing the trajectory data, (2) projecting the time-series of features to a lower-dimensional subspace, (3) conformational clustering in this subspace to define kinetically metastable discrete states, and (4) using the observed transitions counts between discrete states to estimate  $T(\tau)$ , the matrix of transition probabilities between discrete states at lag time  $\tau$ , from which all thermodynamic and kinetic properties can be inferred.

**MSMs of cyclic peptides in solution.**—To build MSMs of peptides in solution, we used the pairwise distances between  $C\alpha$  and  $C\beta$  atoms as features. Dihedral angle ( $\psi$ ,  $\phi$ ) features were also tested (resulting in poorer models, see Results section). Time-structure-based independent component analysis<sup>54,55</sup> (tICA) was used to project the distance features to a low-dimensional subspace for the purpose of conformational clustering. The outputs of tICA are the set of time-lagged independent components (tICs) that best capture a low-rank approximation of the time-lagged correlation matrix. Thus, the tICs represent the degrees of freedom along which the longest-timescale conformational dynamics occurs.

Conformational clustering was performed using the  $k$ -centers algorithm over four tICs (derived using a tICA lag time of 5 ns) to identify 50 metastable states for each cyclic peptide. This number of metastable states ( $N$ ) was chosen based on the generalized matrix Rayleigh quotient (GMRQ) variational cross-validation method.<sup>56</sup> We calculate the GMRQ score over %50/%50 split-shuffle partitions of the trajectory data, where half of the data is used for training, and the rest is used for testing (Figure S3). We can see for  $N = 50$  states, the GMRQ scores are all about the same.

A maximum-likelihood estimator (MLE) was used to estimate the MSM transition matrix from sliding-window counts. An MSM lag time of 20 ns was chosen for construction of the MSM model. Plots of implied time scales versus MSM lag time showed convergence near 20 ns, validating that dynamics was sufficiently Markovian (Figure S4). To compare differences in folding landscape between different peptides, we also construct what we call “joint MSMs” from the combined simulation data for multiple peptides. Previous work from our group showed the success of probing important differences in conformational dynamics by construction of joint MSMs with unified set of state definitions.<sup>57–59</sup>

**MSMs of cyclic peptide binding to MDM2.**—To build MSMs of the peptide binding to MDM2, we chose two different featurizations for comparison. The first set of features, which we will call F1, consisted of pairwise distances between  $C\alpha$  and  $C\beta$  atoms in the peptide and  $C\alpha$  and  $C\beta$  atoms for selected residues in MDM2: Lys51, Leu54, Phe55, Leu57, Gly58, Gln59, Ile61, Met62, Tyr67, Gly72, Val75, Ser92, Val93, His96, Ile99, Tyr100, Ile103. (see Figure S5). The second set of features, which we will call F2, consisted of pairwise distances between  $C\alpha$ ,  $C\beta$  and side chain heavy atoms in the peptide, and  $C\alpha$  and  $C\beta$  of selected residues in MDM2 (the same as listed above). Similar featurizations were used in our previous work, and found to be especially useful in analyzing peptide binding dynamics at high resolution.<sup>38</sup>

For peptide 1, all bound-state and *ab initio* binding trajectories (started from unbound states) were used as the input data to construct MSMs due to the large amount of data from binding simulations. The use of a combination of bound-state and binding trajectories in MSM construction has shown great success to understand binding mechanism at a higher resolution of motions.<sup>60</sup> To identify trajectories in which the peptide ligand successfully binds to MDM2, the center-of-mass distance between MDM2 and the cyclic peptide was computed and compared with the reference distance (1.2 nm) from crystal structure (PDBid:2axi).

The GMRQ method was used to determine optimal hyperparameters for constructing MSMs (Figure S6), such as the number of metastable states (200 states for the peptide 1 system). A tICA lag time of 5 ns and 4 tICA components were used for tICA analysis. A lag time of 25 ns was used for MSMs of peptide 1 binding to MDM2.

### Transition Path Theory analysis.

Transition Path Theory<sup>61–63</sup> (TPT) analysis was performed to estimate binding pathways, fluxes, and rates using the MSMs of cyclic peptides association. Briefly, TPT uses the transition probability matrix  $T^{(\tau)}$  to solve a set of self-consistent equations to obtain committor values  $q_i^+$  for every state  $i$ . The committor value  $q_i^+$  is the probability that a trajectory started from state  $i$  will reach a set of sink states (B) before reaching a set of source states (A). Once the committor values are determined, the reaction rate  $k_{AB}$  can be computed as  $k_{AB} = \frac{F}{\tau \sum_i \pi_i (1 - q_i^+)}$  where  $F = \sum_{i \in A} \sum_{j \notin A} \pi_i T_{ij}^{(\tau)} q_i^+$  is the total flux and  $\pi_i$  is the equilibrium population of state  $i$ . The rate  $k_{AB}$  is the expected number of observed  $A \rightarrow B$  transitions per unit time  $\tau$ , which can be converted to a predicted rate constant by  $k'_{AB} = k_{AB}/c$  where  $c$  is the simulated molar concentration.

To calculate rates of cyclic peptide binding to MDM2, unbound states used in TPT calculations are defined based on the relative distances between the ligand and MDM2 (Figure S7). Any conformational states with an ensemble averaged distance smaller than 0.8 nm and larger than 3.0 nm are defined as bound and unbound states, respectively (Figure S8). Uncertainty estimates were calculated using a bootstrap procedure, whereby ten different MSMs were constructed by sampling the input trajectories with replacement.

### Multiensemble Markov model construction.

The recently developed TRAM estimator by the Noé group (transition-based reweighting analysis method) combines the features of free energy estimators like MBAR with the features of transition count-based MSM rate estimators and gives results better than either alone.<sup>64,65</sup> In order to better estimate kinetics and thermodynamic properties, a MEMM was built from the unbiased MD and umbrella simulation trajectory data. To define MEMM states, we first used tICA to project the combined simulation data of both unbiased and umbrella sampling to tICs, using as features pairwise distances between  $C\alpha$  and  $C\beta$  atoms in ligands and selected residues in MDM2 (see Figure S5). We then confirmed both unbiased and umbrella simulation share the same slow motions (tIC<sub>1</sub> and tIC<sub>2</sub>) (Figure S9). Then we used the state definitions from our MSM of peptide binding to assign state indices



to each snapshot of the umbrella simulation trajectory data (based on their Euclidean distance to the cluster centers). In total, 200 microstates were used for MEMM construction. The trajectory data included both unbiased simulations starting from unbound state, and umbrella simulations starting from bound state. A lag time of 25 ns was used for TRAM estimation based on the convergence of the implied time scales (Figure S10). Uncertainty estimates were calculated using a bootstrap procedure, whereby ten different MEMMs were constructed by sampling the input trajectories from each ensemble with replacement.

### Bayes Factor Analysis.

A Bayes factor method was used to identify important contacts in each metastable state in MSM. Consider two sets of inter-residue contacts  $\{C_{ij}\}$  and  $\{C_{ij}^*\}$ . The variables  $C_{ij}$  are contact indicator variables. If a contact is present between residues  $i$  and  $j$ , then  $C_{ij} = 1$  otherwise,  $C_{ij} = 0$ . The Bayes factor  $BF$  is defined as the ratio of probabilities that the structure is in state  $k$  given the set of contacts  $\{C_{ij}\}$  versus the set of contact  $\{C_{ij}^*\}$ :

$$BF = \frac{P(k | \{C_{ij}\})}{P(k | \{C_{ij}^*\})} \quad (1)$$

Suppose the two sets of contacts differ only by a single contact,  $C_{mn}$  that is formed in the first set of contacts and not formed in the second set. If we assume that each contact is statistically independent, such that  $P(k | \{C_{ij}\}) = \prod_{ij} P(k | C_{ij})$ , then cancellation of terms and application of Bayes' Theorem yields:

$$BF_k(C_{mn}) = \frac{P(k | c_{mn} = 1)}{P(k | c_{mn} = 0)} = \frac{P(C_{mn} = 1 | k) P(C_{mn} = 0)}{P(C_{mn} = 0 | k) P(C_{mn} = 1)} \quad (2)$$

In practice, the Bayes factor  $BF_k(C_{mn})$  is estimated from the frequencies of contacts  $N$  observed in the simulations, where  $P(C_{mn} = 1) = N(C_{mn} = 1)/N_{total}$  and  $P(C_{mn} = 0) = 1 - P(C_{mn} = 1)$ ,  $P(C_{mn} = 1 | k) = N(C_{mn} = 1 | k)/N_{total}$  and  $P(C_{mn} = 0 | k) = (N_k - N(C_{mn} = 1 | k))/N_{total}$  where  $N_k$  is the number of snapshots assigned to microstate  $k$ . A pseudo-count  $N(C_{mn} = 1 | k) = 1$  is given to avoid a zero-valued denominator in the Bayes factors. Since  $N_{total}$  is a very large number ( $\sim 1M$ ) this approximation does not affect the results. In our analysis, we used the MDTraj library<sup>66</sup> to compute contacts separated by three or more residues and define a contact formed if any pair of non-hydrogen atoms between two residues is closer than 0.4 nm.

### NAMFIS analysis.

To identify the conformational families adopted in solution, we used the NMR Analysis of Molecular Flexibility In Solution (NAMFIS) algorithm,<sup>41</sup> which has previously been successfully applied to describe the solution ensembles of linear peptides<sup>67</sup> as well as of peptide,<sup>68–70</sup> natural product<sup>71,72</sup> and drug<sup>73,74</sup> macrocycles. It identifies solution conformers and their molar fraction by deconvoluting time-averaged NMR spectroscopic data using a computationally generated ensemble as an additional input. This input ensemble has to cover the full conformational space available for the molecule, and hence cannot

originate from a structure refinement that uses (experimental) restraints. Motivated by the poor,  $< 0.5$  mg/mL, aqueous solubility of the studied peptides, NOE build-up rates and scalar coupling constants were obtained by recording spectra for DMSO- $d_6$  solutions, similar to previous investigations of systems barely soluble in water,<sup>67,70,71,74</sup> at a 900 MHz NMR spectrometer, as described in detail by Danelius et al.<sup>40</sup>

We used two different sets of data as theoretical input ensembles for the NAMFIS algorithm: (i) one generated from an MSM constructed using unbiased MD simulations, or (ii) one computed using restraint-free Monte Carlo conformational search, as described earlier.<sup>40</sup> The latter theoretical input ensemble was built by combination of the output conformers from two independent conformational searches using the OPLS-2005 and the AMBER\* force fields, respectively, as implemented in the software Macromodel v9.1. Using only conformers within  $42 \text{ kJ mol}^{-1}$  from the global minimum, the ensembles were combined and redundant conformations were eliminated. The solution-state ensembles were then established by deconvolution of the experimentally obtained time-averaged interatomic distances, derived from NOE build up rates, and dihedral angles, obtained from scalar couplings, into conformational ensembles by varying the probabilities of each conformer of the priory predicted theoretical conformational pools. The conformers with non-zero probabilities were classified in terms of  $\beta$ -hairpin folding, with the overall folded populations being given in Table 1. We wish to emphasize that the NAMFIS algorithm is entirely driven by experimental data, and does not use any computed energies in the identification of conformers or their populations, and thereby it avoids the risks associated with the well-known method dependence of computed energies.<sup>75</sup> On the other hand, the reliability of its output vastly depends on the completeness of the theoretical input, which has to represent the entire theoretically available conformational space for the studied flexible compound. The quality of various input theoretical conformational ensembles may be evaluated by statistical algorithms for model comparison (see Results).

## RESULTS AND DISCUSSION

### MSMs suggest a highly preorganized solution-state conformational ensemble for peptide 1.

For each of the solution-state simulations of the four peptides, tICA analysis was performed as described in Methods. The projection of distance-feature trajectories to the first two tICs reveal long-lived kinetic intermediates across the folding landscape (Figure 2). Comparison of the tICA projections show that conformation sampling for peptide 1 is highly concentrated in the native-state basin, in marked contrast to peptides 2, 3 and 4, which populate a number of non-native basins.

Direct comparison of the tICA projections of the four peptides is problematic because the tICs are unique to each peptide's dynamics. To obtain a better comparison of how conformational dynamics varies across the four peptide designs, we constructed joint MSMs trained on the time-series of features for multiple peptides. In agreement with the individual tICA projections, joint tICA projections for pairs of peptides consistently show significant solution-state preorganization for peptide 1 (Figure S11–S13). Since peptide 1 has a D-Pro/L-Pro turn motif, while peptides 2, 3 and 4 have a D-Pro-Gly turn, the simulations suggest



that replacing the highly flexible glycine with L-proline can significantly rigidify the overall cyclic peptide structure. This is consistent with known ability of D-Pro/L-Pro turn mimics to robustly induce hairpin structure.<sup>76</sup>

To gain structural insight into the interactions that uniquely define non-native conformational basins, we computed Bayes factors to identify key inter-residue contacts as described in Methods (Figure S14–S16). For peptide 1, the most populated misfolded state is characterized by a non-native hydrogen-bond registration and interactions between Pro10 and aromatic side-chains. According to the tICA analysis, transitions between this state and the native state constitute the slowest (most time-correlated) dynamics. While this misfolded state is long-lived, few trajectories (1.13%) sample this conformational basin. An MSM model constructed from the trajectory data predicts a population of 11.8% for all misfolded states, which is much lower than the native state (88.2%, see Table 2).

Peptide 2 and 3 have a similar tICA landscape (Figure 2) due to the high similarity in sequences (only the halogen atom is different). They also have the lowest folded state population compared with other peptides (Table 2). The misfolded states are mainly induced by non-native hydrogen bonding between Ser side chains and Glu/Asp side chains (Figure S15). One of the design motivations for these peptide sequences, in which Glu residues of peptide 1 are replaced by Ser, is to provide higher ligand flexibility in solution.<sup>40</sup> However, our results would suggest that increased flexibility introduces unexpected interactions which further lower preorganization. In addition to distance features, we also constructed MSMs using backbone dihedral angle features, to determine whether it resulted in better models of molecular kinetics. By the variational principle of conformational dynamics,<sup>77</sup> the optimal eigenbasis to model molecular kinetics is the one that maximizes the eigenvalues of a time-correlation matrix, which means that slower implied timescales indicate a better model (assuming finite-sampling/over-training errors are negligible). We find that dihedral angle features lead to faster timescales than distance features (Figure S17), and therefore reject this approach in favor of pairwise distance featurization.

### **Binding affinities of cyclic peptides are highly correlated with solution-state stability.**

The highly similar binding interfaces of the four peptides allows for a unique opportunity to examine the relationship between solution-state pre-organization and binding affinity. While tICA analysis provides a qualitative analysis of the relative stability of peptides in solution, here we compare several more quantitative approaches.

Two metrics offer a relatively simple quantitative assessment of peptide preorganization in solution: (1) the native state population predicted by the MSMs, and (2) the  $\beta$ -strand content as computed by the Dictionary of Protein Secondary Structure (DSSP) algorithm. Both metrics provide identical rankings of the preorganization of the four peptides (Table 2). Comparison of predicted native state population with experimental binding affinities<sup>40</sup> reveals a remarkable correlation ( $R^2 = 0.950$ ) between preorganization and experimental binding affinity (Figure 3a). This result suggests an expanded view in which not only cyclization can stabilize peptides to enhance their affinity,<sup>17,78,79</sup> but sequence-based modulation of their intrinsic flexibility and folding properties.

To further explore the relationship between solution-state preorganization and binding affinity, we examined the applicability of two other models. First, we consider a conformational selection model in which the peptide must fold completely before binding to MDM2. In this case, the predicted dissociation constant is  $K_D = K_D^*(1 + K_u)$ , where  $K_D^*$  is the dissociation constant for a folded (i.e. completely preorganized) peptide, and  $K_u$  is the unfolding equilibrium constant (see SI for derivation). Using least-squares fitting of  $\ln K_D^*(1 + K_u)$  to the experimental  $\ln K_D$ , we found that  $K_D^* = 0.519 \mu\text{M}$  best matched the simulation results, with reasonable correlation ( $R^2 = 0.793$ ) between theory and experiment (Figure 3b).

Going beyond this simple two-state folding model, we next considered a model in which the free energy of binding  $G_{bind}$  is related to the change in the peptide's conformational entropy upon binding,  $S_{bind}$ . To see if this was the case, we used a featurization that only included pairwise distances of atoms in the *ligand* to construct joint MSMs from simulated trajectories of 1) the ligand in solution, 2) the ligand bound to MDM2, and 3) *ab initio* binding simulations. As expected, when bound with MDM2 (no unbinding events were observed), a very restricted set of conformational states are visited by peptide 1–4, and more conformational heterogeneity is observed when the peptide is unbound (Figure 4, S18–S20). The value of  $T S_{bind}$  was estimated as  $T(S_b - S_u)$  where  $S_u$  and  $S_b$  are peptide conformational entropies in the unbound and bound states, respectively, calculated from MSMs built from trajectory data of the ligand in solution, and ligand bound to MDM2, respectively. Entropies were estimated using the formula  $S = -R \sum_i p_i \ln p_i$  where  $p_i$  are the metastable state populations of the MSM.

If we assume that the different binding free energies of the four peptides are due entirely to differences in  $S_{bind}$ , we would expect that the experimental values of  $\ln K_D$  vs.  $T S_{bind}$  would obey a linear relationship, with a slope of  $-1/RT$ . Indeed, a scatter plot shows these values to be highly correlated ( $R^2 = 0.962$ ), with nearly exactly the expected slope (Figure 3c). This result suggests that solution-state preorganization completely determines the binding affinity of the four peptides. Moreover, it suggests that atomically-detailed simulation models may be essential for predicting this kind of information.

We note that such strong correlation with solution-state entropy is likely due in part to the high similarity of the four peptides' residues that form an interface with MDM2, and might not be observed for sequence variants which more directly perturb the binding interface. Other studies have also pointed out counterexamples where enhanced preorganization results in weaker binding, suggesting that greater flexibility may be necessary for some designed ligands.<sup>80,81</sup>

### Comparison of simulated ensembles with structural NMR measurements.

To validate the accuracy of our simulated ensembles, we compared simulated interproton distances to NOE measurements published by Danelius et al.<sup>40</sup> Simulated ensemble-averaged NOE distances  $d_i$  were computed as  $d_i = \sum_j \pi_j \langle r_i^{-6} \rangle_j^{-1/6}$  where  $\pi_j$  is the predicted equilibrium population of MSM state  $j$ ,  $r_i$  is distance for proton pair  $i$ , and the bracketed term represents the ensemble average over conformations belonging to state  $j$ .

A comparison of experimental and predicted NOE distances shows *rms* differences to be 1.55, 2.61, 2.68, and 1.35 Å for peptides 1–4, respectively (Figure S22). These differences are comparable to the *rms* differences of the back-calculated solution structures solved by Danelius et al. (1.18, 0.60, 0.82 and 0.43 Å for peptides 1–4, respectively), but slightly larger, which we attribute to our completely unrestrained simulations, performed in different solvent conditions than experiment. As is commonly found in NMR structural refinements, predicted NOE distances consistently overestimate the experimental values, in part because of the difficulty in simultaneously satisfying competing distance restraints.

To further evaluate the ability of the simulated ensembles to recapitulate the observed experimental NOEs, we constructed a representative collection of proton pairs that do *not* exhibit experimental NOEs, by randomly selecting pairs from the set of NOE protons, excluding those from adjacent residues. The predicted distributions of NOE distances from this ‘null’ set comprise distances whose average exceeds  $\sim 7.5$  Å. This indicates that simulated ensembles correctly predict proton pairs with experimentally observed NOEs to be in close contact, while proton pairs with no experimentally observed NOE are predicted to be far apart (Figure 5), suggesting the simulated ensembles agree well with the experimental NMR spectroscopy. Lists of specific interproton distances for each peptide, and their experimental and calculated NOE distances, along with a description of the ‘null’ set of chosen proton pairs, are described in Tables S5–S8.

A confusing discrepancy between our work and that of Danelius et al. was the apparent disagreement in the extent of folded structure reported for peptides 1–4. In Danelius et al., peptides 2 and 3 are reported as having the greatest extent of  $\beta$ -hairpin structure, while our simulations predict that peptide 1 has the greatest extent of folded structure. A deeper analysis, however, shows these disagreements are superficial. In Danelius et al., the NAMFIS algorithm<sup>41</sup> was used to determine an ensemble of structures compatible with NOE constraints, and their relative populations. These populations were in turn used to characterize the extent of  $\beta$ -hairpin population in solution, according to a metric where any structures with at least three hydrogen bonds (defined as hydrogen bond donor and acceptor distance within 0.3 nm) are considered as folded to a hairpin. When we input our MSM states and the ensemble averaged NOE distances to NAMFIS, we achieve excellent agreement with the results reported by Danelius et al. (Table 1).

Further comparison of the two input ensembles was performed using Akaike Information Criterion (AIC) analyses.<sup>82</sup> This statistical technique compares the ability of various models to describe reality, that is, the fit of the models to experimentally obtained observables. Model selection is important, as an under-fitted model may not capture the true conformational variability of a flexible molecule, whereas an over-fitted one lacks generality. The summary of the AIC analyses of the input models (i) and (ii) for the four peptides are given in Table 3.

The AIC analyses (Table 3) suggests that there is a decisive evidence that the Monte Carlo-derived input ensemble is the more robust model. This means that the NAMFIS output derived using the Monte Carlo conformational search better fits to the experimental data, as this input ensemble contains all, or at least more conformers that are present in the real-life

solution. There several possible explanations for this result, perhaps the most important being that the MD simulations (and subsequent MSMs) were performed in aqueous solution, whereas the NMR data was obtained for DMSO-solutions (made necessary by the low aqueous solubility of the studied peptides). Another explanation is provided by the nature of the conformational search methods. A Monte Carlo conformational search provides a highly diverse set of conformers, but an unreliable energetic ranking (not used in the NAMFIS analyses), whereas molecular dynamics is capable of reliably identifying the dominant solution conformers. Thus, MD methods may have difficulty discovering less-populated conformers, which in turn may complicate the ensemble fitting to experimental data. Finding the relevant conformations of macrocycles in solution has previously been shown to pose a significant challenge.<sup>84</sup>

SSD is the sum of square differences between the measured and modelled distances, as defined by Snyder et al.,<sup>83</sup> along with the error estimates given in the same article (NOE distance  $< 2.5 \text{ \AA}$ :  $\pm 0.1 \text{ \AA}$  error, NOE distance  $2.5 - 3.0 \text{ \AA}$ :  $\pm 0.2$ ;  $3.0 - 3.5 \text{ \AA}$ :  $\pm 0.3$ ; NOE distance  $3.5 - 6.0 \text{ \AA}$ :  $\pm 0.4$ ). AIC is the Akaike information criterion, AICc is the small sample size-corrected AIC that has a compensation for over-fitting for systems with small sample size,  $\Delta AIC_c$  is the difference in AIC of an individual model as compared to the best model,  $w_j$  is the Akaike weight of evidence in favor of a model being the actual best model for the given data,  $ER_j$  is the evidence ratio describing the relative likelihood of a pair of models, and  $LER_j$  is the logarithmic evidence ratio. The difference in  $LER_j$  expresses the degree by which the best proposed model is better than all other anticipated models. Accordingly, the model best fitting with reality has  $LER_j = 0$ , and an  $LER_j$  difference of  $0 - 0.5$  is interpreted as “weak”,  $0.5 - 1$  as “substantial”,  $1 - 2$  as “strong”, and  $> 2$  as “decisive” evidence.

That said, it is likely that the NAMFIS analysis does not agree with the MSM predictions of folded state populations (Table 2) because of the major differences in the information each algorithm uses to arrive at conformational population estimates. MSMs rely heavily on the simulation model, estimating conformational populations based on transitions observed between metastable states, with the only restraints being detailed balance. NAMFIS, as mentioned above, relies much more on restraints derived from experimental observables, estimating conformational populations from a constraint-based analysis of states with viable structures. A more thorough analysis comparing the performance of MSMs (which can be reconciled with experimental restraints using the BICePs algorithm<sup>85-87</sup>) and NAMFIS remains an interesting topic for future research.

### **Binding mechanisms of designed cyclic MDM2 ligands.**

A powerful advantage of MSMs is the ability to obtain information about both kinetics and thermodynamics. Towards this end, we performed MD simulations of MDM2 with cyclic peptide ligands initiated from bound and unbound states, which generated hundreds of *ab initio* binding events. The resulting trajectory data was combined and used to construct MSMs of the binding reaction, as described in Methods. Roughly equal numbers of binding events were observed for all ten starting positions of the ligand, which suggests that the simulated binding dynamics is not biased by the initial configuration (Figure S23).

For peptides 2, 3 and 4, we find that the tICA landscapes (i.e. the projection of the trajectory data to the first two tICs) are similar. As an example, we present the tICA landscape (constructed using feature set F1, see Methods) for peptide 4 (Figure 6). Results for peptides 2 and 3 are shown in Figure S24. For these peptides, the slowest motion (along tIC<sub>1</sub>) corresponds to peptide binding, while the next-slowest motion (along tIC<sub>2</sub>) corresponds to peptide folding. The tICA landscape shows the vast majority of binding pathways must first visit intermediates where the peptide is folded, but not yet bound to MDM2 in the correct binding pose. Thus, we conclude that designed cyclic peptides 2, 3 and 4 bind via a conformational selection mechanism.

In contrast to peptides 2–4, peptide 1 shows very little unfolding, resulting in a unique tICA landscape in which most metastable states have folded peptide conformations similar to the crystal pose (Figure 7). This finding is consistent with our simulations of cyclic peptides in solution, which show that peptide 1 has the highest extent of preorganization, and a more stable folded state than peptides 2–4. For this system, we found that feature set F2 (which includes heavy atoms of peptide side chains) produced superior models. The use of the F2 feature set resulted in slower implied timescales than F1, suggesting an MSM less prone to lumping errors, and better able to capture the relevant dynamics (Figure S25a–b). Moreover, the F2 featurization was able to discern binding to the correct pose on the 2D tICA landscape (Figure S25c–f).

Similar to the tICA landscapes of peptides 2, 3 and 4, dynamical motion along the largest tICA component (tIC<sub>1</sub>) corresponds to binding of the cyclic peptide to the crystallographic pose (Figure 7d). Motion along the next-slowest component (tIC<sub>2</sub>), however, represents binding of the peptide to two major off-pathway traps, in poses that differ from the crystal structure (Figure 7b, h). The center of the tICA landscape corresponds to the unbound state; therefore, most trajectories that visit these “trap” states must first unbind to correctly reach to the crystallographic bound-state pose.

Further inspection of the two trap states reveal near-bound poses in which hydrophobic residues designed to mimic the “hot spot” residues of Phe19, Trp23, and Leu26 of p53-TAD are mismatched with their preferred pocket in the MDM2 receptor. In peptide 1, 6-chloro-Trp (residue 3) is designed to mimic Trp23 in p53-TAD. In the two near-bound poses, we find instead that Phe8 (Figure 7a) and Phe1 (Figure 7g) occupy the pocket in MDM2 where Trp23 normally resides. These binding poses are further stabilized by  $\pi$ -stacking interactions between His74-Trp(Cl)3, Trp6-Tyr78 (Figure 7c) and Phe33-Trp(Cl)3, Trp6-Tyr45 (Figure 7i).

The significance of the trap states is intriguing to consider. Because the folding of peptide 1 is so robust, such trap states likely represent long-lived local free energy minima which may compete with the crystallographic bound state. Discovering “trap” binding modes in computational peptide design efforts may be useful in either penalizing off-pathway poses, or in discover alternative motifs that could be further optimized.

### Predicted binding rates of cyclic peptide 1 to MDM2.

To estimate the kinetics of cyclic peptides binding to MDM2, we computed mean-first-passage times (MFPTs) using TPT analysis based on constructed MSMs (see Methods). We present results only for peptide 1 (see Table 4), for which experimental measurements performed using surface plasmon resonance (SPR) have been published by Danelius et al.<sup>40</sup> The predicted binding rate,  $k_{\text{on}}$ , of peptide 1 agrees with the experimental measurement within an order of magnitude (Table 4). The estimated unbinding rate,  $k_{\text{off}}$ , and the bound-state population,  $\pi_{\text{bound}}$ , however, are poorly underestimated. Similar results were found for an MSM model of p53 TAD binding to MDM2,<sup>38</sup> and the reason for it is well-known: MSM estimators assume that input trajectory data is sampled at equilibrium, which is impractical for binding simulations of this kind. To realize this, consider a scenario in which a peptide has dissociation constant of  $K_D \sim 100$  nM. At our simulated concentration, we should expect only 3 of 100,000 samples to be in the unbound state. For MSM estimators that enforce detailed balance (as used in this study), input data that includes more than this fraction of unbound-state samples would tend to underestimate  $k_{\text{on}}$  and overestimate  $k_{\text{off}}$ . For MSM estimators that do not enforce detailed balance (such as a row-normalized counts estimator), finite-sampling error will affect estimates similarly.

To improve these estimates, we were interested in constructing multi-ensemble Markov models from umbrella simulations of peptide unbinding (see Methods) using an estimator called TRAM (Transition-based Reweighting Analysis Method),<sup>64</sup> which uses samples from multiple thermodynamic ensembles to estimate both free energies of metastable states and the transition rates between them. The approach has been used to much more accurately estimate slow dissociation rates of ligands, including for high-affinity designed peptide ligands of MDM2.<sup>65</sup>

We find that rate estimation using TRAM greatly improves the accuracy of both  $k_{\text{off}}$  and  $\pi_{\text{bound}}$  when compared to the conventional MSM (Table 4). Specifically, the TRAM estimation of  $\pi_{\text{bound}}$  (~96%) agrees better with the experimental value (99.997%). The TRAM estimate of  $k_{\text{off}}$  is similarly improved (i.e. smaller by a factor of 20) over the conventional MSM estimate, and is within two orders of magnitude from the experimental value. One might expect this estimate to improve further if  $k_{\text{on}}$  estimates weren't overestimated with respect to experiment by an order of magnitude. While rate predictions from TRAM are far from perfect, we note that the improvements are remarkable given that the total amount of biased trajectory data (from umbrella simulations, see Methods) used in the calculation is only ~ 4% of the unbiased data.

We also note that relatively few MD-based estimates of  $k_{\text{off}}$  for high-affinity peptide ligands have been reported in the literature to date, and that the accuracy of the TRAM results we achieve is comparable to the current state-of-the-art (around one or two orders of magnitude). For the designed PMI peptide, a nanomolar inhibitor of MDM2 with experimental  $k_{\text{off}}$  of  $3.73 [2.93, 4.05] \times 10^{-2} \text{ s}^{-1}$ , Paul et al. report estimates of  $1.14 [0.72, 2.08] \text{ s}^{-1}$  and  $1.25 [0.25, 6.7] \times 10^{-1} \text{ s}^{-1}$  using two different rate matrix estimators.<sup>65</sup>

These results show the utility of multi-ensemble Markov models in accurately estimating kinetics and thermodynamic of peptide binding, which bodes well for future studies using



greater amounts of biased trajectory data. Accurate prediction of  $k_{\text{off}}$  is highly desirable, as the dissociation rate is recognized as an important quantity in drug efficacy and safety in drug discovery.<sup>88–90</sup> Much pioneering work has recently been done to develop improved methods for predicting dissociation rates of protein-ligand interactions,<sup>91–94</sup> and we are eager to contribute efforts towards this goal.

## CONCLUSION

We have used molecular simulation approaches to understand the relationship between solution-state preorganization, binding mechanism, and binding affinity for four designed cyclic peptide ligands of MDM2. The microscopic view of conformational dynamics provided by Markov models reveals good agreement with previously published NMR studies, and a strong correlation between the loss of conformational entropy upon binding, and the experimental binding free energy. The best model explaining this correlation is one where solution-state preorganization completely determines the binding affinity of the four peptides, suggesting that simulation-based prediction of the solution-state properties of cyclic peptide binders may be key to discovering peptides with enhanced binding affinity.

Markov state models of the binding reaction for the four cyclic peptides show a conformational selection mechanism of binding. Even for the most preorganized design, peptide 1, non-native “trap” states are present, which could be designable features in future computational design studies. MSMs give reasonable agreement with experiment binding rates, which can be improved using the TRAM estimator alongside additional umbrella sampling simulations.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

This research was supported in part by the National Institutes of Health through grant 1S10OD020095-01. Y.G and S.Z was supported by National Institutes of Health grant 1R01GM123296. We thank the Swedish Research Council (2016-03602, 2020-03431) for financial support. The NAMFIS analyses were performed using resources provided by the Swedish National Infrastructure for Computing (SNIC) at NSC (project 2020-5-435). This research includes calculations carried out on Temple University’s HPC resources and thus was 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. We would like to thank the participants of Folding@home, without whom this work would not be possible. The author also appreciates the insightful discussions with conference attendees during the poster session of the Biophysical Society Annual Meeting in 2019.

## References

- (1). Malde AK; Hill TA; Iyer A; Fairlie DP Crystal Structures of Protein-Bound Cyclic Peptides. *Chem. Rev.* 2019, 119, 9861–9914. [PubMed: 31046237]
- (2). Nielsen DS; Shepherd NE; Xu W; Lucke AJ; Stoermer MJ; Fairlie DP Orally Absorbed Cyclic Peptides. *Chem. Rev.* 2017, 117, 8094–8128. [PubMed: 28541045]
- (3). Dougherty PG; Sahni A; Pei D Understanding Cell Penetration of Cyclic Peptides. *Chem. Rev.* 2019, 119, 10241–10287. [PubMed: 31083977]

- (4). Qian Z; Dougherty PG; Pei D ScienceDirect Targeting intracellular protein–protein interactions with cell-permeable cyclic peptides. *Curr. Opin. Chem. Biol.* 2017, 38, 80–86. [PubMed: 28388463]
- (5). Wakefield AE; Wuest WM; Voelz VA Molecular simulation of conformational pre-organization in cyclic RGD peptides. *J. Chem. Inf. Model.* 2015, 55, 806–813. [PubMed: 25741627]
- (6). Demmer O; Frank AO; Hagn F; Schottelius M; Marinelli L; Cosconati S; Brack-Werner R; Kremb S; Wester H-J; Kessler H A Conformationally Frozen Peptoid Boosts CXCR4 Affinity and Anti-HIV Activity. *Angew. Chem., Int. Ed.* 2012, 51, 8110–8113.
- (7). Joerger AC; Fersht AR The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harbor Perspect. Biol.* 2010, 2, a000919–a000919.
- (8). Kubbutat MH; Jones SN; Vousden KH Regulation of p53 stability by Mdm2. *Nature* 1997, 387, 299–303. [PubMed: 9153396]
- (9). Haupt Y; Maya R; Kazaz A; Oren M Mdm2 promotes the rapid degradation of p53. *Nature* 1997, 387, 296–299. [PubMed: 9153395]
- (10). Momand J; Zambetti GP; Olson DC; George D; Levine AJ The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* 1992, 69, 1237–1245. [PubMed: 1535557]
- (11). Tao W; Levine AJ Nucleocytoplasmic shuttling of oncoprotein Hdm2 is required for Hdm2-mediated degradation of p53. *Proc. Natl. Acad. Sci. U. S. A.* 1999, 96, 3077–3080. [PubMed: 10077639]
- (12). Kussie PH; Gorina S; Marechal V; Elenbaas B; Moreau J; Levine AJ; Pavletich NP Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain. *Science* 1996, 274, 948–953. [PubMed: 8875929]
- (13). Honda R; Tanaka H; Yasuda H Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Lett.* 1997, 420, 25–27. [PubMed: 9450543]
- (14). Wells M; Tidow H; Rutherford TJ; Markwick P; Jensen MR; Mylonas E; Svergun DI; Blackledge M; Fersht AR Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105, 5762–5767. [PubMed: 18391200]
- (15). Rew Y; Sun D; Gonzalez-Lopez De Turiso F; Bartberger MD; Beck HP; Canon J; Chen A; Chow D; Deignan J; Fox BM; Gustin D; Huang X; Jiang M; Jiao X; Jin L; Kayser F; Kopecky DJ; Li Y; Lo M-C; Long AM; Michelsen K; Oliner JD; Osgood T; Ragains M; Saiki AY; Schneider S; Toteva M; Yakowec P; Yan X; Ye Q; Yu D; Zhao X; Zhou J; Medina JC; Olson SH Structure-Based Design of Novel Inhibitors of the MDM2–p53 Interaction. *J. Med. Chem.* 2012, 55, 4936–4954. [PubMed: 22524527]
- (16). Estrada-Ortiz N; Neochoritis CG; Dömling A How To Design a Successful p53-MDM2/X Interaction Inhibitor: A Thorough Overview Based on Crystal Structures. *ChemMedChem* 2015, 11, 757–772. [PubMed: 26676832]
- (17). Baek S; Kutchukian PS; Verdine GL; Huber R; Holak TA; Lee KW; Popowicz GM Structure of the Stapled p53 Peptide Bound to Mdm2. *J. Am. Chem. Soc.* 2011, 134, 103–106. [PubMed: 22148351]
- (18). Chang YS; Graves B; Guerlavais V; Tovar C; Packman K; To K-H; Olson KA; Kesavan K; Gangurde P; Mukherjee A; Baker T; Darlak K; Elkin C; Filipovic Z; Qureshi FZ; Cai H; Berry P; Feyfant E; Shi XE; Horstick J; Annis DA; Manning AM; Fotouhi N; Nash H; Vassilev LT; Sawyer TK Stapled  $\alpha$ helical peptide drug development: A potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc. Natl. Acad. Sci. U. S. A.* 2013, 110, E3445–E3454. [PubMed: 23946421]
- (19). Kritzer JA; Lear JD; Hodsdon ME; Schepartz A Helical  $\beta$ -Peptide Inhibitors of the p53-hDM2 Interaction. *J. Am. Chem. Soc.* 2004, 126, 9468–9469. [PubMed: 15291512]
- (20). 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. [PubMed: 18725320]
- (21). Robinson J a.  $\beta$ -Hairpin Peptidomimetics: Design, Structures and Biological Activities. *Acc. Chem. Res.* 2008, 41, 1278–1288. [PubMed: 18412373]

- (22). Fasan R; Dias RLA; Moehle K; Zerbe O; Vrijbloed JW; Obrecht D; Robinson J a. Using a  $\beta$ -Hairpin To Mimic an  $\alpha$ -Helix: Cyclic Peptidomimetic Inhibitors of the p53–HDM2 Protein–Protein Interaction. *Angew. Chem., Int. Ed.* 2004, 116, 2161–2164.
- (23). Fasan R; Dias RLA; Moehle K; Zerbe O; Obrecht D; Mittl PRE; Grütter MG; Robinson J a. Structure-Activity Studies in a Family of  $\beta$ -Hairpin Protein Epitope Mimetic Inhibitors of the p53-HDM2 Protein-Protein Interaction. *ChemBioChem* 2006, 7, 515–526. [PubMed: 16511824]
- (24). Pazgier M; Liu M; Zou G; Yuan W; Li C; Li C; Li J; Monbo J; Zella D; Tarasov SG; Lu W Structural basis for high-affinity peptide inhibition of p53 interactions with MDM2 and MDMX. *Proc. Natl. Acad. Sci. U. S. A.* 2009, 106, 4665–4670. [PubMed: 19255450]
- (25). Liu M; Pazgier M; Li C; Yuan W; Li C; Lu W A Left-Handed Solution to Peptide Inhibition of the p53-MDM2 Interaction. *Angew. Chem., Int. Ed.* 2010, 49, 3649–3652.
- (26). Zhan C; Zhao L; Wei X; Wu X; Chen X; Yuan W; Lu W-Y; Pazgier M; Lu W An Ultrahigh Affinity d-Peptide Antagonist Of MDM2. *J. Med. Chem.* 2012, 55, 6237–6241. [PubMed: 22694121]
- (27). Mas-Moruno C; Rechenmacher F; Kessler H Cilengitide: the first anti-angiogenic small molecule drug candidate design, synthesis and clinical evaluation. *Anti-Cancer Agents Med. Chem.* 2010, 10, 753–768.
- (28). Syud FA; Stanger HE; Gellman SH Interstrand Side Chain Side Chain Interactions in a Designed  $\beta$ -Hairpin: Significance of Both Lateral and Diagonal Pairings. *J. Am. Chem. Soc.* 2001, 123, 8667–8677. [PubMed: 11535071]
- (29). Martín-Gago P; Ramón R; Aragón E; Fernández-Carneado J; Martín-Malpartida P; Verdaguer X; López-Ruiz P; Colás B; Cortes MA; Ponsati B; Macias MJ; Riera A A tetradecapeptide somatostatin dicarba-analog: Synthesis, structural impact and biological activity. *Bioorg. Med. Chem. Lett.* 2014, 24, 103–107. [PubMed: 24342240]
- (30). Lama D; Quah ST; Verma CS; Lakshminarayanan R; Beuerman RW; Lane DP; Brown CJ Rational Optimization of Conformational Effects Induced By Hydrocarbon Staples in Peptides and their Binding Interfaces. *Sci. Rep.* 2013, 3, 277–10.
- (31). Saglam AS; Wang DW; Zwier MC; Chong LT Flexibility vs Preorganization: Direct Comparison of Binding Kinetics for a Disordered Peptide and Its Exact Preorganized Analogues. *J. Phys. Chem. B* 2017, 121, 10046–10054. [PubMed: 28992700]
- (32). Chatterjee D; Boyd CD; O’Toole GA; Sondermann H Structural Characterization of a Conserved, Calcium-Dependent Periplasmic Protease from *Legionella pneumophila*. *J. Bacteriol.* 2012, 194, 4415–4425. [PubMed: 22707706]
- (33). Brown CJ; Quah ST; Jong J; Goh AM; Chiam PC; Khoo KH; Choong ML; Lee MA; Yurlova L; Zolghadr K; Joseph TL; Verma CS; Lane DP Stapled Peptides with Improved Potency and Specificity That Activate p53. *ACS Chem. Biol.* 2013, 8, 506–512. [PubMed: 23214419]
- (34). Schafmeister CE; Po J; Verdine GL An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. *J. Am. Chem. Soc.* 2000, 122, 5891–5892.
- (35). Walensky LD; Bird GH Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress. *J. Med. Chem.* 2014, 57, 6275–6288. [PubMed: 24601557]
- (36). Edwards AL; Gavathiotis E; LaBelle JL; Braun CR; Opoku-Nsiah KA; Bird GH; Walensky LD Multimodal interaction with BCL-2 family proteins underlies the proapoptotic activity of PUMA BH3. *Chem. Biol.* 2013, 20, 888–902. [PubMed: 23890007]
- (37). Borchers W; Theillet F-X; Katzer A; Finzel A; Mishall KM; Powell AT; Wu H; Manieri W; Dieterich C; Selenko P; Loewer A; Daughdrill GW Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nat. Chem. Biol.* 2014, 10, 1000–1002. [PubMed: 25362358]
- (38). Zhou G; Pantelopulos GA; Mukherjee S; Voelz VA Bridging Microscopic and Macroscopic Mechanisms of p53-MDM2 Binding with Kinetic Network Models. *Biophys. J.* 2017, 113, 785–793. [PubMed: 28834715]
- (39). Paul F; Noé F; Weikl TR Identifying Conformational-Selection and Induced-Fit Aspects in the Binding-Induced Folding of PMI from Markov State Modeling of Atomistic Simulations. *J. Phys. Chem. B* 2018, 122, 5649–5656. [PubMed: 29522679]

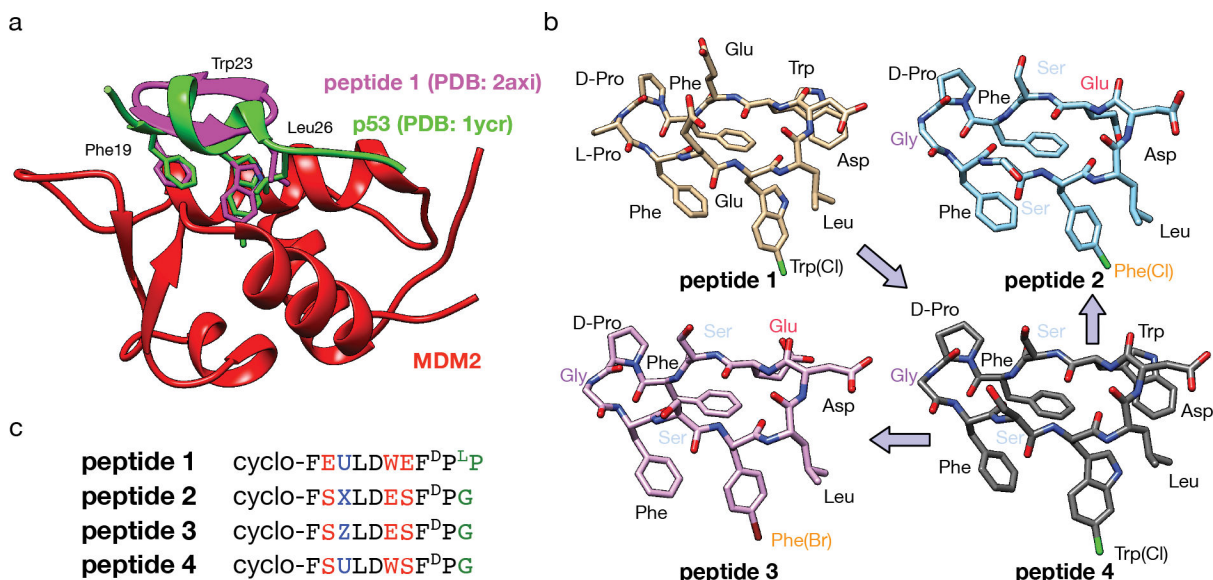
- (40). Danelius E; Pettersson M; Bred M; Min J; Waddell MB; Guy RK; Grøtli M; Erdelyi M Flexibility is important for inhibition of the MDM2/p53 protein–protein interaction by cyclic  $\beta$ -hairpins. *Org. Biomol. Chem.* 2016, 14, 10386–10393. [PubMed: 27731454]
- (41). Cicero DO; Barbato G; Bazzo R NMR Analysis of Molecular Flexibility in Solution: A New Method for the Study of Complex Distributions of Rapidly Exchanging Conformations. Application to a 13-Residue Peptide with an 8-Residue Loop. *J. Am. Chem. Soc.* 1995, 117, 1027–1033.
- (42). Wang J; Wolf RM; Caldwell JW; Kollman PA; Case D a. Development and testing of a general amber force field. *J. Comput. Chem.* 2004, 25, 1157–1174. [PubMed: 15116359]
- (43). Jakalian A; Jack DB; Bayly CI Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* 2002, 23, 1623–1641. [PubMed: 12395429]
- (44). Sousa da Silva AW; Vranken WF ACPYPE - AnteChamber PYthon Parser interface. *BMC Res. Notes* 2012, 5, 367. [PubMed: 22824207]
- (45). Pronk S; Páll S; Schulz R; Larsson P; Bjelkmar P; Apostolov R; Shirts MR; Smith JC; Kasson PM; van der Spoel D; Hess B; Lindahl E GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 2013, 29, 845–854. [PubMed: 23407358]
- (46). Shirts M; Pande VS COMPUTING: Screen Savers of the World Unite! *Science* 2000, 290, 1903–1904. [PubMed: 17742054]
- (47). Li D-W; Brüschweiler R NMR-Based Protein Potentials. *Angew. Chem., Int. Ed.* 2010, 49, 6778–6780.
- (48). Beauchamp KA; McGibbon R; Lin Y-S; Pande VS Simple few-state models reveal hidden complexity in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 2012, 109, 17807–17813. [PubMed: 22778442]
- (49). Acharyya A; Ge Y; Wu H; DeGrado WF; Voelz VA; Gai F Exposing the Nucleation Site in  $\alpha$ -Helix Folding: A Joint Experimental and Simulation Study. *J. Phys. Chem. B* 2019, 123, 1797–1807. [PubMed: 30694671]
- (50). Silva D-A; Bowman GR; Sosa-Peinado A; Huang X A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein. *PLoS Comput. Biol.* 2011, 7, 1–11.
- (51). Ge Y; Voelz VA Markov State Models to Elucidate Ligand Binding Mechanism. In *Protein-Ligand Interactions and Drug Design*; Ballante F, Ed.; Springer US: New York, NY, 2021; pp 239–259.
- (52). Harrigan MP; Sultan MM; Hernández CX; Husic BE; Eastman P; Schwantes CR; Beauchamp KA; McGibbon RT; Pande VS MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophys. J.* 2017, 112, 10–15. [PubMed: 28076801]
- (53). Scherer MK; 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. *J. Chem. Theory Comput.* 2015, 11, 5525–5542. [PubMed: 26574340]
- (54). Pérez-Hernández G; Paul F; Giorgino T; De Fabritiis G; Noé F Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* 2013, 139, 015102–14. [PubMed: 23822324]
- (55). Schwantes CR; Pande VS Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* 2013, 9, 2000–2009. [PubMed: 23750122]
- (56). McGibbon RT; Pande VS Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.* 2015, 142, 124105–13. [PubMed: 25833563]
- (57). Razavi AM; Wuest WM; Voelz VA Computational Screening and Selection of Cyclic Peptide Hairpin Mimetics by Molecular Simulation and Kinetic Network Models. *J. Chem. Inf. Model.* 2014, 54, 1425–1432. [PubMed: 24754484]
- (58). Razavi AM; Voelz VA Kinetic Network Models of Tryptophan Mutations in  $\beta$ -Hairpins Reveal the Importance of Non-Native Interactions. *J. Chem. Theory Comput.* 2015, 11, 2801–2812. [PubMed: 26575573]

- (59). Ge Y; Kier BL; Andersen NH; Voelz VA Computational and Experimental Evaluation of Designed  $\beta$ -Cap Hairpins Using Molecular Simulations and Kinetic Network Models. *J. Chem. Inf. Model.* 2017, 57, 1609–1620. [PubMed: 28614661]
- (60). Ge Y; Borne E; Stewart S; Hansen MR; Arturo EC; Jaffe EK; Voelz VA Simulations of the regulatory ACT domain of human phenylalanine hydroxylase (PAH) unveil its mechanism of phenylalanine binding. *J. Biol. Chem.* 2018, 293, 19532–19543. [PubMed: 30287685]
- (61). Metzner P; Schütte C; Vanden-Eijnden E Transition Path Theory for Markov Jump Processes. *Multiscale Model. Simul.* 2009, 7, 1192–1219.
- (62). Noé F; Schütte C; Vanden-Eijnden E; Reich L; Weikl TR Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. U. S. A.* 2009, 106, 19011–19016. [PubMed: 19887634]
- (63). Berezhkovskii A; Hummer G; Szabo A Reactive flux and folding pathways in network models of coarse-grained protein dynamics. *J. Chem. Phys.* 2009, 130, 205102–6. [PubMed: 19485483]
- (64). Wu H; Paul F; Wehmeyer C; Noé F Multiensemble Markov models of molecular thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U. S. A.* 2016, 113, E3221–30. [PubMed: 27226302]
- (65). Paul F; Wehmeyer C; Abualrous ET; Wu H; Crabtree MD; Schöneberg J; Clarke J; Freund C; Weikl TR; Noé F Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. *Nat. Commun.* 2017, 8, 1095. [PubMed: 29062047]
- (66). McGibbon RT; Beauchamp KA; Harrigan MP; Klein C; Swails JM; Hernández CX; Schwantes CR; Wang L-P; Lane TJ; Pande VS MD-Traj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* 2015, 109, 1528–1532. [PubMed: 26488642]
- (67). Koivisto JJ; Kumpulainen ETT; Koskinen AMP Conformational ensembles of flexible beta-turn mimetics in DMSO-d<sub>6</sub>. *Org. Biomol. Chem.* 2010, 8, 2103–2116. [PubMed: 20401387]
- (68). Thepchatri P; Eliseo T; Cicero DO; Myles D; Snyder JP Relationship Among Ligand Conformations in Solution, in the Solid State, and at the Hsp90 Binding Site: Geldanamycin and Radicicol. *J. Am. Chem. Soc.* 2007, 129, 3127–3134. [PubMed: 17323946]
- (69). Danelius E; G OR; Mukandlia Mulumba AH; Ong H; S C; Erdelyi M; Lubell WD. Dynamic Chirality in the Mechanism of Action of Allosteric CD36 Modulators of Macrophage-Driven Inflammation. *J. Med. Chem.* 2019, 62, 11071–11079. [PubMed: 31774287]
- (70). E D; Danelius E; Mitchell SA; F HD; Erdelyi M; T TA. A Chemical Biology Approach to Understanding Molecular Recognition of Lipid II by Nisin(1–12): Synthesis and NMR Ensemble Analysis of Nisin(1–12) and Analogues. *Chem Eur. J.* 2019, 25, 14572–14582. [PubMed: 31599485]
- (71). Thepchatri P; Cicero DO; Monteagudo E; Ghosh AK; Cornett B; Weeks ER; Snyder JP Conformations of Laulimalide in DMSO-d<sub>6</sub>. *J. Am. Chem. Soc.* 2005, 127, 12838–12846. [PubMed: 16159277]
- (72). Erdelyi M; B P; Haunstein K; Forher J; Gertsch J; Altman K-H; Carlomagno T. Conformational Preferences of Natural and C3-Modified Epothilones in Aqueous Solution. *J. Med. Chem.* 2008, 51, 1469–1478. [PubMed: 18271516]
- (73). Danelius E; Poongavanam V; Peintner S; Wieske LHE; Erdelyi M; J. K. Solution Conformations Explain the Chameleonic Behaviour of Macrocyclic Drugs. *Chem. - Eur. J.* 2020, 26, 5231–5244. [PubMed: 32027758]
- (74). Peng C; Y A; amd Xu Z WJ; Poongavanam V; Shi J; Kihlberg J; W Z; Erdelyi M. Conformation of the Macrocyclic Drug Lorlatinib in Polar and Nonpolar Environments: A MD Simulation and NMR Study. *ACS Omega* 2019, 4, 22245–22250. [PubMed: 31891108]
- (75). Lakdawala A; Wang M; Nevins N; Liotta DC; Rusinska-Rozsak D; Lozynski M; Snyder JP Calculated conformer energies for organic molecules with multiple polar functionalities are method dependent: Taxol (case study). *BMC Chem. Biol.* 2001, 1, 2. [PubMed: 11710970]
- (76). Favre M; Moehle K; Jiang L; Pfeiffer B; Robinson JA Structural mimicry of canonical conformations in antibody hypervariable loops using cyclic peptides containing a heterochiral diproline template. *J. Am. Chem. Soc.* 1999, 121, 2679–2685.
- (77). Nüske F; Keller BG; Pérez-Hernández G; Mey ASJS; Noé F Variational Approach to Molecular Kinetics. *J. Chem. Theory Comput.* 2014, 10, 1739–1752. [PubMed: 26580382]

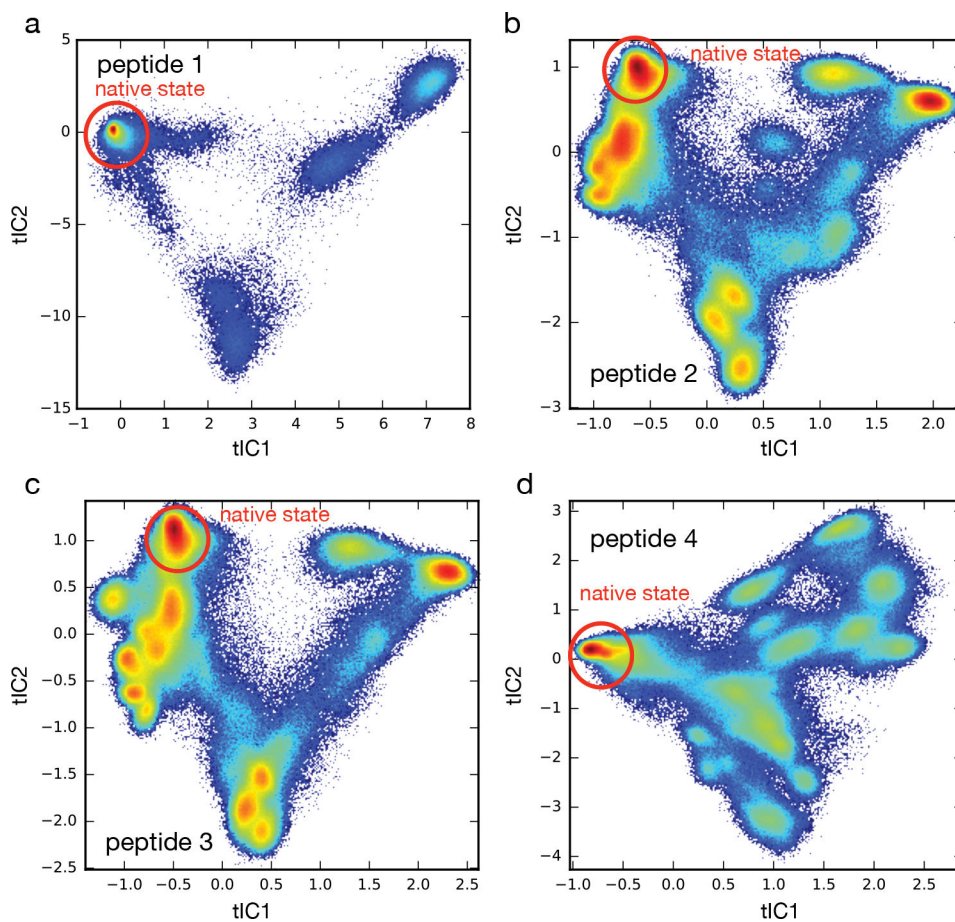


- (78). Bernal F; Wade M; Godes M; Davis TN; Whitehead DG; Kung AL; Wahl GM; Walensky LD A Stapled p53 Helix Overcomes HDMX-Mediated Suppression of p53. *Cancer Cell* 2010, 18, 411–422. [PubMed: 21075307]
- (79). Guo Z; Streu K; Krilov G; Mohanty U Probing the Origin of Structural Stability of Single and Double Stapled p53 Peptide Analogs Bound to MDM2. *Chem. Biol. Drug Des.* 2014, 83, 631–642. [PubMed: 24418072]
- (80). Doak BC; Zheng J; Dobritzsch D; Kihlberg J How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets. *J. Med. Chem.* 2016, 59, 2312–2327. [PubMed: 26457449]
- (81). Kumar EA; Chen Q; Kizhake S; Kolar C; Kang M; Chang C.-e. A.; Borgstahl GEO; Natarajan A. The paradox of conformational constraint in the design of Cbl(TKB)-binding peptides. *Sci. Rep.* 2013, 3, 677–7.
- (82). Akaike H In Information Theory and an Extension of the Maximum Likelihood Principle. In *Proceedings of the 2nd International Symposium on Information Theory*; Akademiai Kiado, Budapest, 1973; pp 267–281.
- (83). Nevins N; Cicero D; Snyder JP A Test of the Single-Conformation Hypothesis in the Analysis of NMR Data for Small Polar Molecules: A Force Field Comparison. *J. Org. Chem.* 1999, 64, 3979–3986.
- (84). Poongavanam V; Danelius E; Peintner S; Alcaraz L; Caron G; Cummings MD; Wlodek S; Erdelyi M; Hawkins PCD; Ermondi G; Kihlberg J Conformational Sampling of Macrocyclic Drugs in Different Environments: Can We Find the Relevant Conformations? *ACS Omega* 2018, 3, 11742–11757. [PubMed: 30320271]
- (85). Voelz VA; Zhou G Bayesian inference of conformational state populations from computational models and sparse experimental observables. *J. Comput. Chem.* 2014, 35, 2215–2224. [PubMed: 25250719]
- (86). Ge Y; Voelz VA Model Selection Using BICePs: A Bayesian Approach for Force Field Validation and Parameterization. *J. Phys. Chem. B* 2018, 122, 5610–5622. [PubMed: 29518328]
- (87). Wan H; Ge Y; Razavi A; Voelz VA Reconciling Simulated Ensembles of Apomyoglobin with Experimental Hydrogen/Deuterium Exchange Data Using Bayesian Inference and Multiensemble Markov State Models. *J. Chem. Theory Comput.* 2020, 16, 1333–1348. [PubMed: 31917926]
- (88). Copeland RA The drug-target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discovery* 2016, 15, 87–95. [PubMed: 26678621]
- (89). Copeland RA; Pompliano DL; Meek TD Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discovery* 2006, 5, 730–739. [PubMed: 16888652]
- (90). Núñez S; Venhorst J; Kruse CG Target–drug interactions: first principles and their application to drug discovery. *Drug Discovery Today* 2012, 17, 10–22. [PubMed: 21777691]
- (91). Lotz SD; Dickson A Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. *J. Am. Chem. Soc.* 2018, 140, 618–628. [PubMed: 29303257]
- (92). Tiwary P; Limongelli V; Salvalaglio M; Parrinello M Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U. S. A.* 2015, 112, E386–E391. [PubMed: 25605901]
- (93). Bruce NJ; Ganotra GK; Kokh DB; Sadiq SK; Wade RC New approaches for computing ligand-receptor binding kinetics. *Curr. Opin. Struct. Biol.* 2018, 49, 1–10. [PubMed: 29132080]
- (94). Gobbo D; Piretti V; Di Martino RMC; Tripathi SK; Giabbai B; Storici P; Demitri N; Giroto S; Decherchi S; Cavalli A Investigating Drug-Target Residence Time in Kinases through Enhanced Sampling Simulations. *J. Chem. Theory Comput.* 2019, 15, 4646–4659. [PubMed: 31246463]

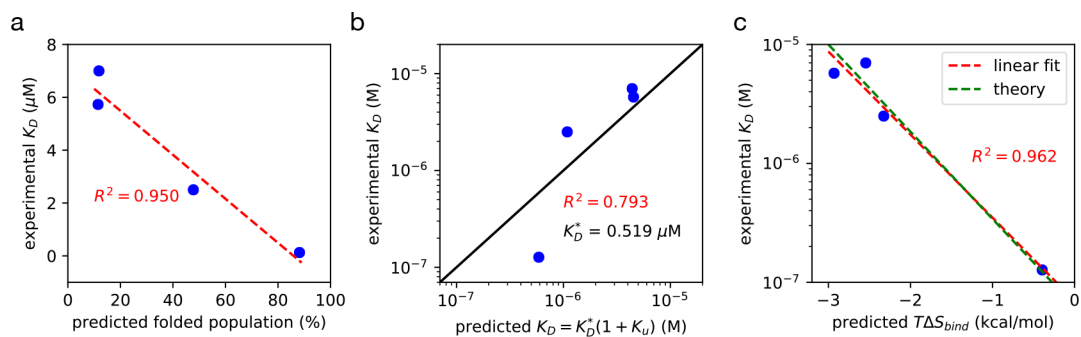


**Figure 1:**

(a) Superposition of the crystal structure of peptide 1 bound to MDM2 (PDBid: 2axi, magenta) and the crystal structure of the p53 transactivation domain peptide (green) bound to MDM2 (red, PDBid: 1ycr). (b) Structures of the four designed cyclic  $\beta$ -hairpins examined in this study. Peptide 1 was originally designed by Fasan et al.<sup>23</sup> Peptide 4 is obtained by substitution of Glu and L-Pro with Ser and Gly, respectively. Peptide 2 is obtained by further substitution of Trp and 6-chloro-(L)-Trp with Glu and 4-chloro-(L)-Phe. Peptide 3 is obtained after further replacement of 4-chloro-(L)-Phe with 4-bromo-(L)-Phe. (c) Sequences of the four peptides examined in this study. Single-letter residue codes  $U$  = 6-chloro-(L)-tryptophan;  $X$  = 4-chloro-(L)-phenylalanine;  $Z$  = 4-bromo-(L)-phenylalanine.

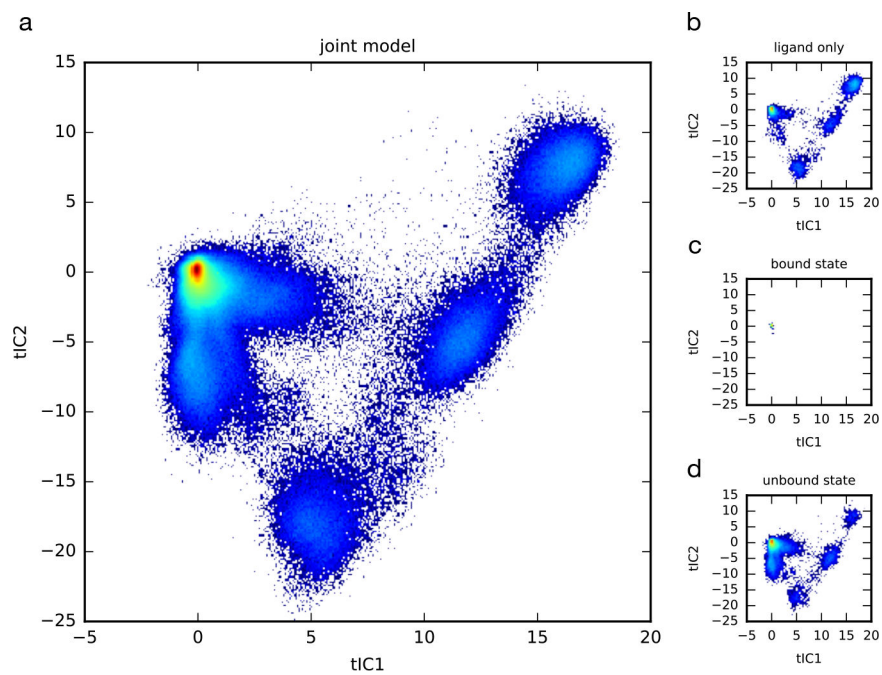


**Figure 2:** A heat map of ligand-only simulation trajectory data projected to the 2D tICA landscape of (a) peptide 1, (b) peptide 2, (c) peptide 3, (d) peptide 4. The native folded states are circled in red.

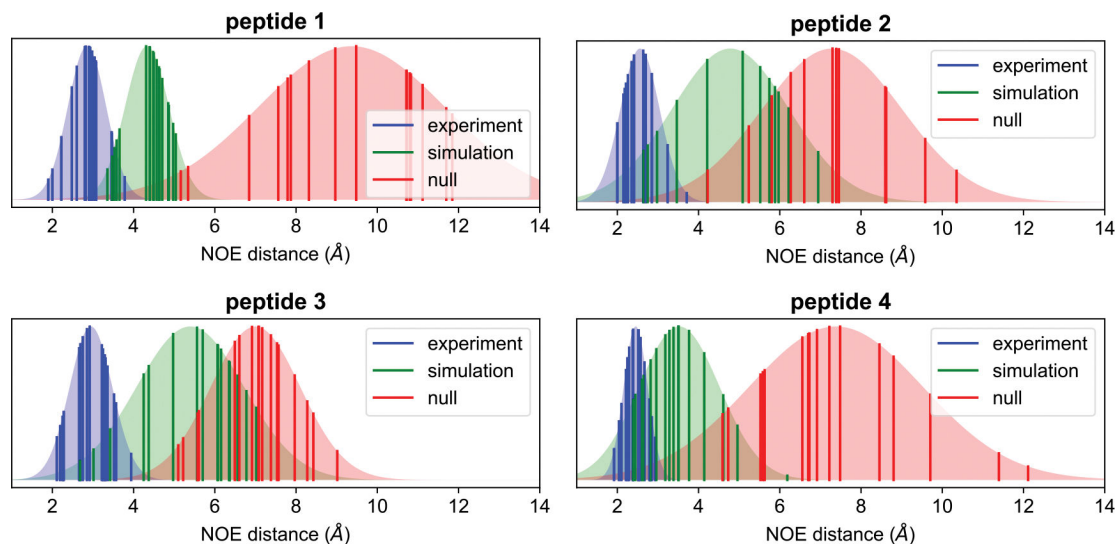


**Figure 3:**

(a) Experimentally measured dissociation constants versus folded state population predicted by MSMs. (b) Experimentally measured dissociation constants versus predictions from a three-state conformational selection model. (c) Experimentally measured dissociation constants versus  $-T \mathcal{S}_{\text{bind}}$ .

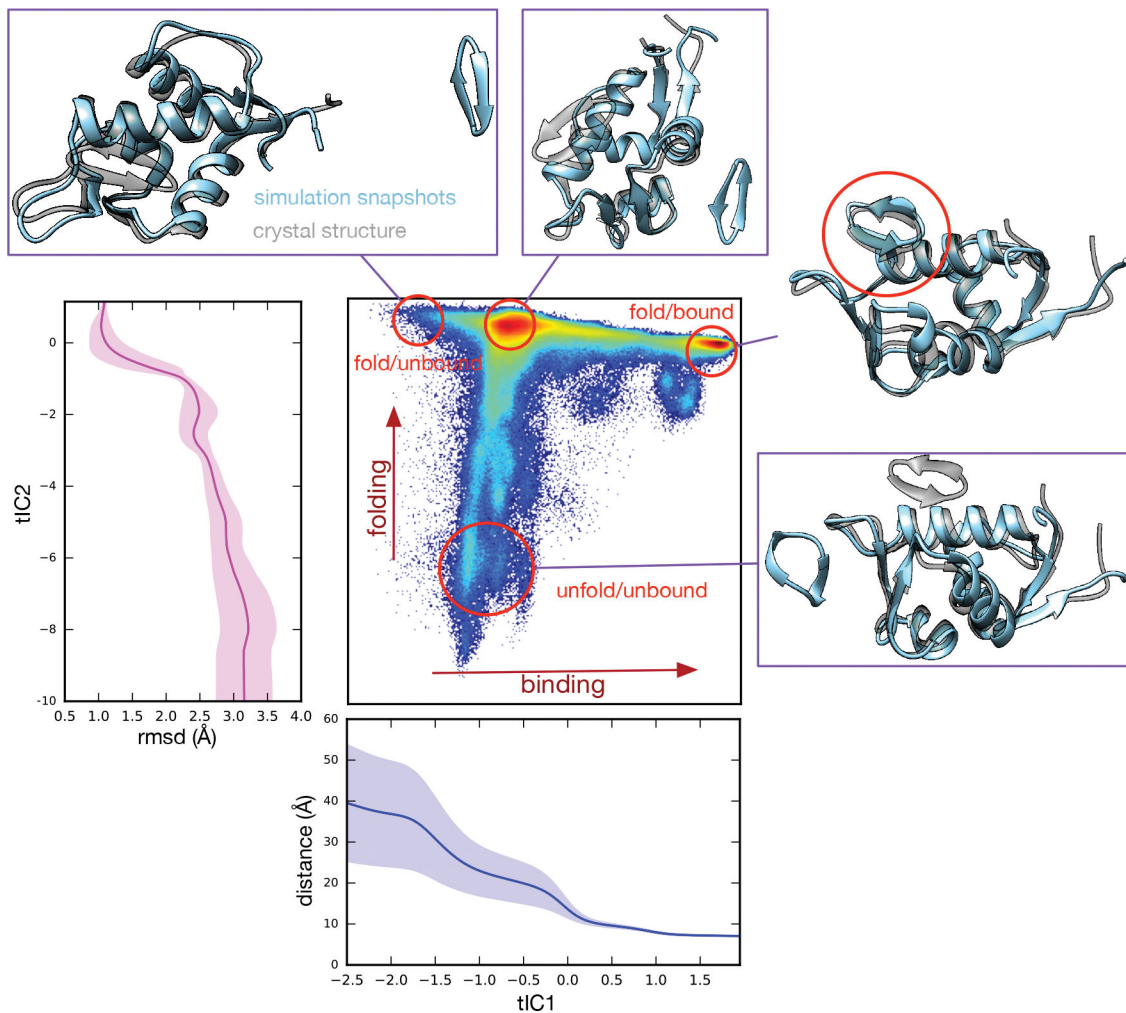


**Figure 4:**  
(a) Combined trajectory data of peptide 1 projected to the 2D tICA landscape. (b-d) Transformation of simulation trajectory data using tICA model generated from combined trajectory data for ligand only, bound state and unbound state.



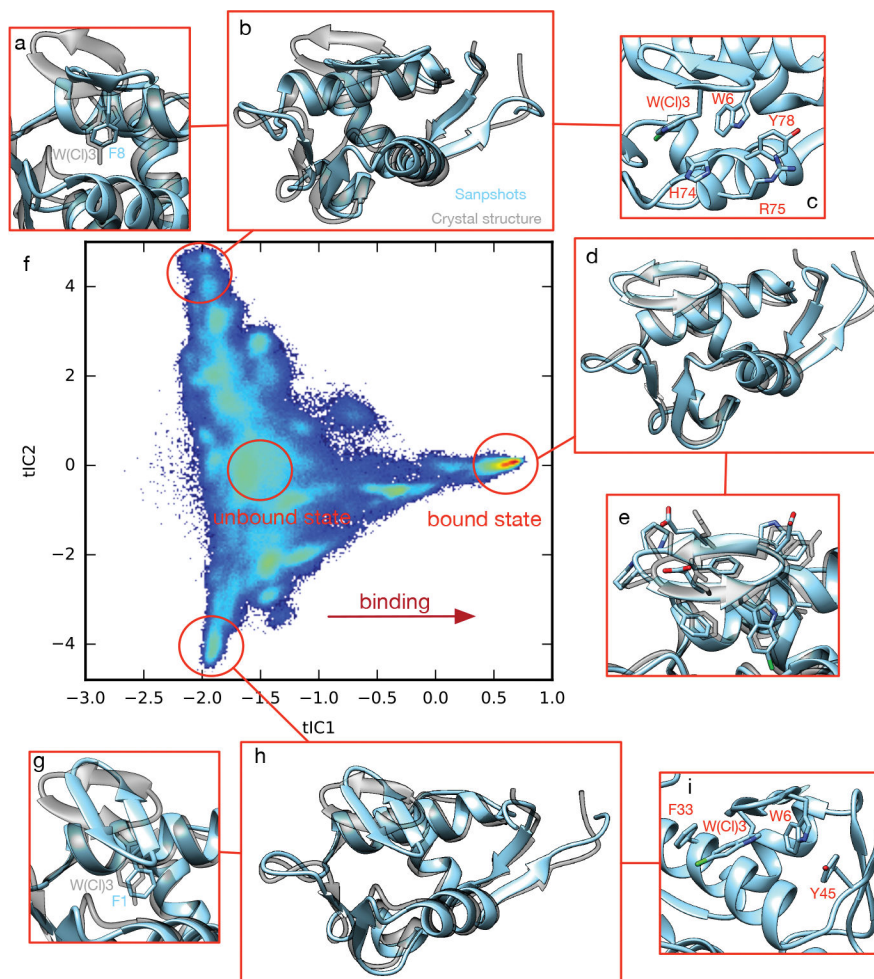
**Figure 5:**

Simulated ensembles correctly predict proton pairs with experimentally observed NOEs to be in close contact, while proton pairs with no experimentally observed NOEs are predicted to be far apart. Experimental distances (blue) and simulated distances (green) for proton pairs with experimentally observed NOEs are compared to an equal number of randomly chosen 'null' distances (red) for proton pairs without experimentally observed NOEs. To guide the eye, Gaussian distributions with the sample mean and variance of each set of distances are shown. (The full list of proton pairs are described in Tables S5–S8.)



**Figure 6:** The slowest dynamics (along  $tIC_1$ ) corresponds to the binding process whereas the next-slowest dynamics corresponds to the ligand folding process. A conformational selection binding mechanism is revealed from the tICA landscape. Averaged distances between selected residues on MDM2 and the ligand (see Figure S6) is shown in blue and rmsd change along  $tIC_2$  is shown in magenta.





**Figure 7:** The slowest dynamics (along  $tIC_1$ ) corresponds to the binding process whereas the next-slowest dynamics corresponds to the ligand binding with different poses other than the crystal structure. Snapshot from simulation trajectories with key interactions observed are shown for selected metastable states.

**Table 1:**

NAMFIS predicted folded state population.

Peptide	Previous published NAMFIS analysis <sup>40</sup> (%)	NAMFIS predicted folded state population from MSMs using a H-bonding metric (%)
1	24	20
2	61	63
3	61	56
4	39	57

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:**

Stability predicted from simulations vs. binding affinities measured from experiments. The uncertainty estimates of predictions are done by sampling the input trajectories with replacement (10 times).

Peptide	$\beta$ -strand content in solution (DSSP) (%)	folded state population in solution (MSMs) (%)	Binding affinities $K_D$ ( $\mu$ M)
1	$33.66 \pm 0.29$	$88.49 \pm 4.31$	$0.127 \pm 0.001$
2	$0.73 \pm 0.07$	$11.49 \pm 1.90$	$7.00 \pm 0.10$
3	$0.52 \pm 0.05$	$10.19 \pm 3.82$	$5.73 \pm 0.09$
4	$19.70 \pm 1.02$	$47.94 \pm 3.79$	$2.50 \pm 0.02$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3:**

Summary of AIC results for the theoretical models used as input for the estimation of the  $\beta$ -hairpin populations of peptides 1–4.

Model	SSD	AIC	AICc	AICc	$W_i$	$ER_i$	$LER_i$
Peptide 1							
Monte Carlo	576.5	594.5	595.9	0.0	1.0	1.0	0
MD sim.	815.1	851.1	873.1	277.2	0.0	1.6e60	60.2
Peptide 2							
Monte Carlo	150.9	164.9	166.1	0.0	1.0	1.0	0
MD sim.	459.8	471.8	473.8	307.7	0.0	6.6e51	66.8
Peptide 3							
Monte Carlo	227.7	241.7	242.9	0.0	1.0	1.0	0.0
MD sim.	509.5	515.5	516.0	273.1	0.0	2.0e59	59.3
Peptide 4							
Monte Carlo	147.5	165.5	167.4	0.0	1.0	1.0	0.0
MD sim.	357.9	367.9	369.2	201.8	0.0	6.8e43	43.8

SSD is the sum of square differences between the measured and modelled distances, as defined by Snyder et al.,<sup>83</sup> along with the error estimates given in the same article (NOE distance  $< 2.5 \text{ \AA}$ :  $\pm 0.1 \text{ \AA}$  error, NOE distance  $2.5 - 3.0 \text{ \AA}$ :  $\pm 0.2$ ;  $3.0 - 3.5 \text{ \AA}$ :  $\pm 0.3$ ; NOE distance  $3.5 - 6.0 \text{ \AA}$ :  $\pm 0.4$ ). AIC is the Akaike information criterion, AICc is the small sample size-corrected AIC that has a compensation for over-fitting for systems with small sample size,  $\Delta AICc$  is the difference in AIC of an individual model as compared to the best model,  $w_i$  is the Akaike weight of evidence in favor of a model being the actual best model for the given data,  $ER_j$  is the evidence ratio describing the relative likelihood of a pair of models, and  $LER_j$  is the logarithmic evidence ratio. The difference in  $LER_j$  expresses the degree by which the best proposed model is better than all other anticipated models. Accordingly, the model best fitting with reality has  $LER_j 0$ , and an  $LER_j$  difference of  $0 - 0.5$  is interpreted as “weak”,  $0.5 - 1$  as “substantial”,  $1 - 2$  as “strong”, and  $> 2$  as “decisive” evidence.

**Table 4:**

Estimated kinetics and thermodynamics properties using different estimators. Uncertainties are derived from 10-fold bootstrapping of input trajectories with replacement. Uncertainties in estimated rates  $k$  are reported as confidence intervals (brackets), calculated from  $\pm\sigma_{\ln k}$ .

Estimator	$\pi_{\text{bound}}$ (%)	$k_{\text{on}}$ ( $\text{s}^{-1}\text{M}^{-1}$ )	$k_{\text{off}}$ ( $\text{s}^{-1}$ )
MSM	$56.51 \pm 10.73$	$8.20 [5.63, 10.74] \times 10^6$	$9.98 [6.53, 13.44] \times 10^2$
TRAM	$96.12 \pm 4.25$	$8.03 [4.62, 13.96] \times 10^6$	$3.72 [0.30, 45.80] \times 10^1$
Experiment	$99.997^*$	$8.30 (\pm 0.02) \times 10^5$	$1.074 (\pm 0.001) \times 10^{-1}$

\* derived from experimental measured  $K_D = 129.3 \pm 0.3$  nM (see SI).