

Article

Binding Ensembles of *p53*-MDM2 Peptide Inhibitors by Combining Bayesian Inference and Atomistic Simulations

Lijun Lang  and Alberto Perez * 

Chemistry Department, University of Florida, Gainesville, FL 32611, USA; lijunlang@chem.ufl.edu

* Correspondence: perez@chem.ufl.edu; Tel.: +1-352 3927009

Abstract: Designing peptide inhibitors of the *p53*-MDM2 interaction against cancer is of wide interest. Computational modeling and virtual screening are a well established step in the rational design of small molecules. But they face challenges for binding flexible peptide molecules that fold upon binding. We look at the ability of five different peptides, three of which are intrinsically disordered, to bind to MDM2 with a new Bayesian inference approach (MELD×MD). The method is able to capture the folding upon binding mechanism and differentiate binding preferences between the five peptides. Processing the ensembles with statistical mechanics tools depicts the most likely bound conformations and hints at differences in the binding mechanism. Finally, the study shows the importance of capturing two driving forces to binding in this system: the ability of peptides to adopt bound conformations ($\Delta G_{conformation}$) and the interaction between interface residues ($\Delta G_{interaction}$).

Keywords: IDP 1; binding 2; molecular dynamics 3; MELD×MD 4; advanced sampling 5; *p53* 6; MDM2 7



Citation: Lang, L.; Perez, A. Binding Ensembles of *p53*-MDM2 Peptide Inhibitors by Combining Bayesian Inference and Atomistic Simulations. *Molecules* **2021**, *26*, 198. <https://doi.org/10.3390/molecules26010198>

Academic Editor: Marilisa Leone
Received: 13 November 2020
Accepted: 28 December 2020
Published: 2 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Peptide molecule inhibitors have the potential to bind to proteins classified as “undruggable” by small molecules thanks to their flexibility and complementary nature to proteins [1,2]. Rational drug design of small molecules via computational tools (e.g., docking of virtual libraries) is a common practice in the drug discovery process. However, these tools are not well suited to handle the flexible nature of peptide molecules, many of which are intrinsically disordered and only adopt stable structures in the presence of their binding partners [3].

Modeling the binding of flexible molecules continues to be a grand challenge in computational structure prediction. In recent years, with the increase of peptide therapeutics in the market there has been a continuous development and adaptation of docking tools to capture protein-peptide interactions [3,4]. Docking programs address the flexibility of peptides by two main routes: (1) using homology models, PDB (Protein Data Bank) structural motifs, or other sources of structures for docking [5–8]; and (2) provide peptide flexibility for folding upon binding [9–14]. Initial peptide conformations for docking could come from computationally expensive molecular dynamics (MD) simulations of the free peptide. However, many such peptides are intrinsically disordered (IDP), limiting their use [7]. Full exploration of folding upon binding through standard molecular dynamics becomes too computationally demanding [15], requiring advanced sampling strategies to efficiently sample the energy landscape.

In this work, we take a look at binding and free-peptide ensembles (simulating the peptide in isolation) for different peptides to better understand the nature of the *p53*-MDM2 interaction. *p53* is called the guardian of the genome, triggering programmed death (apoptosis) when cells misbehave. MDM2 down-regulates *p53* limiting its tumor suppressor activity. Thus, inhibitors of the *p53*-MDM2 and the closely related MDMX interaction have long been a cancer drug target [16–18]. Multiple studies of the native

interaction [19–22] and the ability to design inhibitors that simultaneously block MDM2 and MDMX [23–25] provide a wealth of data to assess new computational tools. Since binding simulations are more computationally demanding than free peptide simulations, our goal is to identify peptide properties that might make the peptide a better binder—leading to faster computational screening of peptide therapeutics.

The *p53*-MDM2 interaction is characterized by three hydrophobic residues (Phe19, Trp23 and Leu26) from the peptide which anchor into a deep cavity in MDM2. In order for the three hydrophobic residues to align with the pocket, the *p53* epitope adopts a helical conformation. This is in contrast with the IDP nature of the peptide in isolation. We use noisy information to guide binding using our previously developed Bayesian inference approach (MELD×MD [26]) to identify the subset of data that is most compatible with the force field and the resulting bound conformations (see Figure 1). To further test the methodology, we simulated five different peptides, including the peptide epitope from *p53*, two inhibitors, and two alanine-based peptides that we do not expect to be good binders, as control. The work highlights the ability of molecular dynamics tools to capture the two driving forces behind binding: preferences of the peptides to adopt bound-like conformations and the use of binding simulations to differentiate binding preferences.

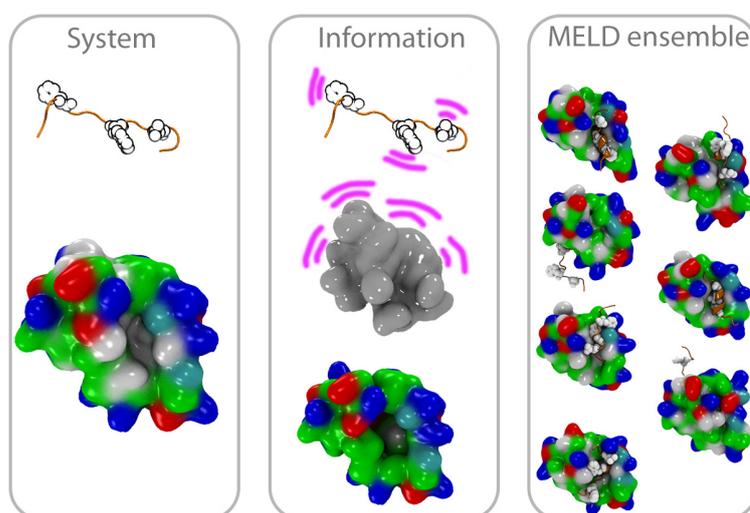


Figure 1. Outline of the MELD×MD setup. We start with the peptide far from MDM2 (system). We use noisy information to favor sampling of binding/unbinding events (middle panel). A statistical mechanics of the posterior distribution coming from the MELD ensemble identifies conformations that are most consistent with the force field and a subset of the data, and we compare these to the experimental structure.

2. Results

2.1. Free Peptide Simulation Ensembles Show the IDP Nature of *p53*

We simulated five peptides in their free form (see Methods and Table 1), capturing their intrinsic degree of disorder. All peptides are able to visit multiple states with short life times. A 2D-RMSD clustering of the ensemble reveals many clusters with low populations for the *p53* and two control peptides, consistent with their intrinsically disordered nature (see unrestrained Molecular Dynamics (MD) column in Table 2). The peptide *pdiq* adopts stable helix conformations for a significant amount of time, while *ATSP-7041* is an outlier in this analysis, adopting very stable helical conformations due to the presence of a chemical staple. We used these simulations to define a common reference frame to compare simulations for all peptides in their free and binding simulations (see Methods). Each peptide ensemble was projected onto the corresponding eigenvectors that showed a good separation between helical and non-helical states—as those are the states required for binding (see left panels in Figures 2 and A1–A4). Clustering on the space defined by the top 14 eigenvectors shows

that only *ATSP-7041* and, to a lesser extent, *pdiq* adopt stable helical structures—consistent with the IDP nature of the other three peptides.

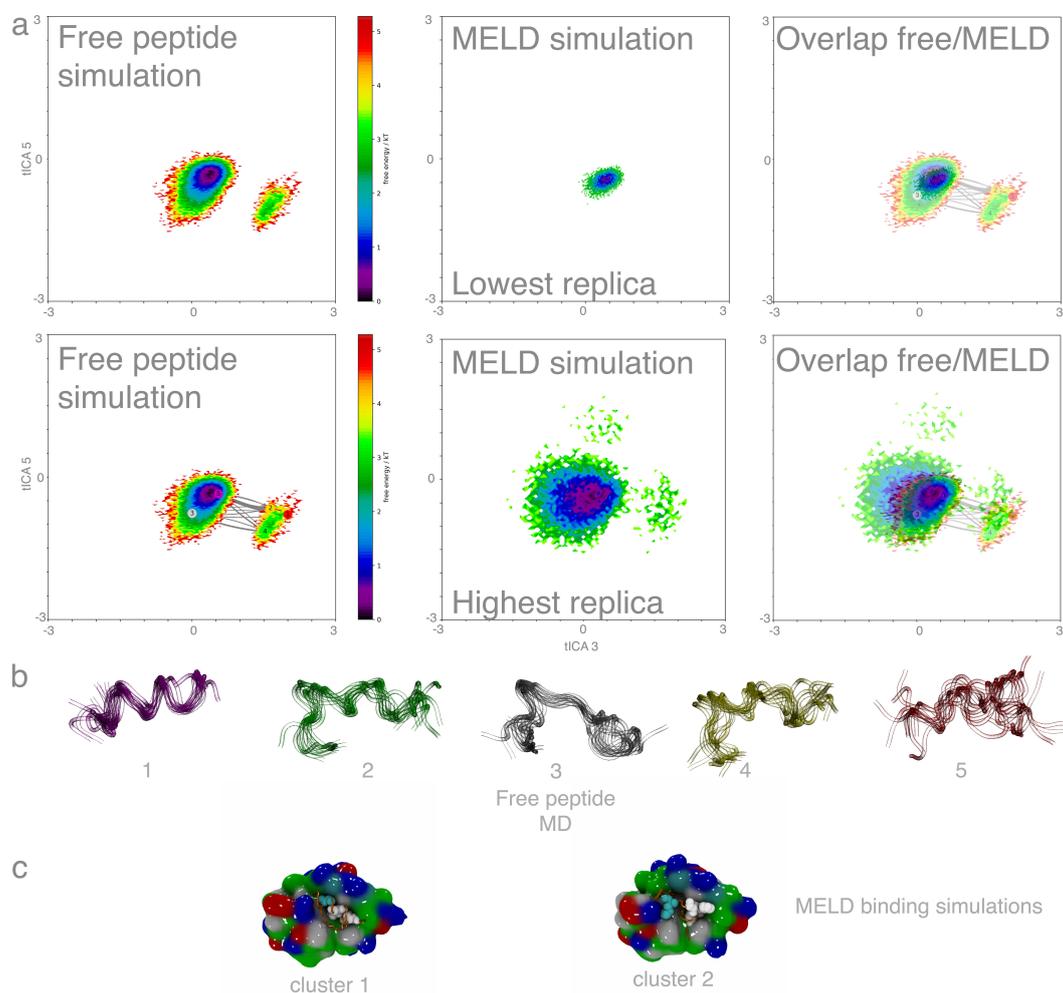


Figure 2. Comparison of the conformational space for free peptide versus binding simulations for the *ATSP-7041* peptide. (a) The peptide ensembles are projected onto the third and fifth tICA eigenvectors common to all five peptides. (b) The metastable states sampled for the free peptide. (c) Top clusters by population from MELD×MD binding simulations.

Table 1. Peptides used in the current work. Bold letters represent the anchoring residues.

Name	Sequence																
<i>p53</i>	S	Q	E	T	F	S	D	L	W	K	L	L	P	E	N		
<i>pdiq</i>			E	T	F	E	H	W	W	S	Q	L	L	S			
<i>Ala1</i>			A	A	F	A	A	A	W	A	A	L	A	A			
<i>Ala2</i>			A	A	A	A	A	A	A	A	A	A	A	A			
<i>ATSP-7041</i>		ACE	L	T	F	R8	E	Y	W	A	Q	Cba	S5	S	A	A	NHE

Table 2. Populations for peptides in free and MELD binding simulations. Clustering is done on the lowest temperature replica using hierarchical clustering with $\epsilon = 1.5$.

Name	Peptide Population (% Top Cluster)		
	Unrestrained MD	MELD×MD (Peptide Align)	MELD×MD (Protein Align)
<i>p53</i>	0.6	70.6	46.1
<i>pdiq</i>	24.0	97.6	95.3
<i>Ala1</i>	1.4	54.7	16.0
<i>Ala2</i>	0.2	31.3	17.5
ATSP-7041	69.5	97.8	91.6

2.2. MELD×MD Simulations Balance Exploration and Exploitation of the Binding Energy Landscape

Figure 3 provides a visual outlook on the binding process explored by the MELD×MD replica exchange procedure in terms of the relative position of the peptide with respect to the protein and the peptide's intrinsic conformational preferences. At high replica indexes, the force constants for the restraints are set to zero and the temperature is high (see methods). In these conditions, the peptide samples conformations far away from the active site, distributed uniformly around the protein. During the binding process, the MDM2 flexibility allows for the opening and closing of the binding cavity (see right panel in Figure A5). As the replica index decreases, the temperature decreases and the biasing restraints towards the protein become active, producing a frustrated energy landscape. Under these conditions, the peptide samples conformations on the surface of the protein, identifying early on the MDM2 hydrophobic pocket as the most likely region for binding. Sampling is concentrated in the binding pocket at the lowest replica. Thus, at the highest replica, the protocol favors full exploration of the energy landscape, while, at the lowest replica, it favors full exploitation by sampling around a particular binding region near the protein. The nature of MELD×MD enhances binding/unbinding events by allowing replicas to explore different Hamiltonian and temperature conditions, leading to a different balance of exploration and exploitation [27,28].

2.3. Peptides Become More Structured in Proximity to MDM2

MELD×MD binding simulations show a higher fraction of helical conformations for all peptides with respect to their free simulations (see middle column in Table 2). However, the increase in helical content for the peptide is not always associated with binding at the correct binding site (right column in Table 2). A 2D-RMSD clustering calculation on all replicas (see Figures 4 and A6) reveals the funneled nature of binding for three peptides. We can identify three broadly defined regions in the funneling plots based on the RMSD distribution: between 0–5 Å (high accuracy binding), 5–15 Å (pre-bound), ~15–30 Å (misbound), and a fourth region for unbound conformations sampled by higher replicas (see Figure A5). All five peptides identify the binding pocket as the binding site, but the two control sequences bind through multiple backbone conformations with little structural preference. ATSP-7041 exhibits the most funneled behavior, rapidly converging onto a large high accuracy native-like cluster. Both *pdiq* and *p53* exhibit a similar behavior, in which all three regions are explored even at the lower replicas, with funneling to one major state. For *p53*, the native configuration is sampled, but is not identified as the most populated cluster. The observed binding mode introduces a kink in the backbone between the helical and non-helical region that is not observed in the experimental structure. The control *Ala1* sequence contains the three anchoring residues present in *p53*, but exhibits a binding profile more similar to control *Ala2*, which lacks the anchoring residues. Thus, the control sequences show that the MELD×MD setup is not over-constraining the peptide to bind

in the binding pocket or in the binding conformation, and large cluster populations are reflective of significant binding.

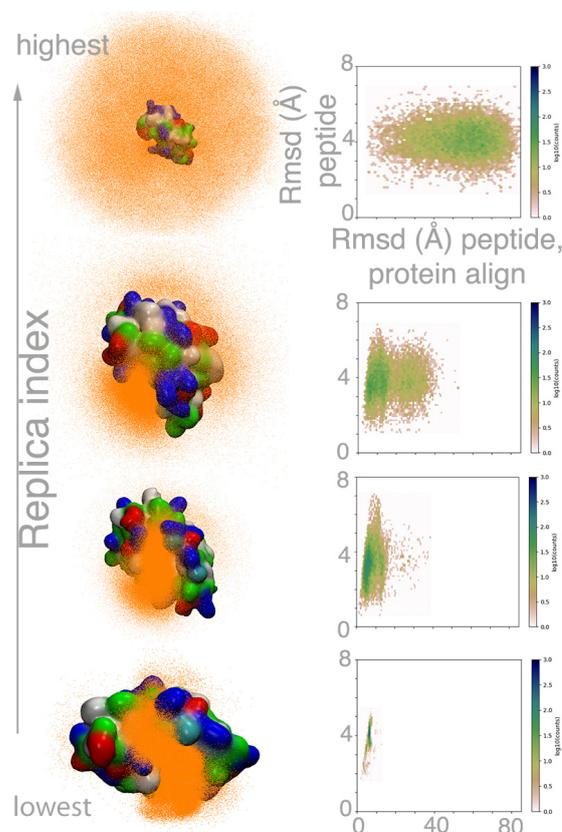


Figure 3. MELD×MD simulations explore unbound states, and different possible binding regions of *p53* on the surface of the protein. The left panel shows a superposition of all peptide conformations (heavy atoms of each conformation are drawn as orange dots) sampled at different replicas. The right panel shows the internal backbone RMSD of the peptide with respect the experimental conformation versus the RMSD of the peptide when aligning to the protein.

We compare all peptide binding ensembles on equal footing by projecting them on the same eigenvector space as the free peptides. Figure 2 compares the free peptide ensemble with those produced from MELD×MD at the lowest/highest replica index (bound/unbound) for *ATSP-7041*. The figure also shows the clusters arising from the free peptide ensemble, as well as the highest population clusters, from the binding simulations. The preferred conformation for *ATSP-7041* in its free peptide is the same conformation needed for binding, resulting in significant binding observed throughout the simulations. A similar behavior is observed for *pdiq*, where the ensemble of the free peptide is larger due to the absence of the chemical staple (see Figure A1). For the three IDP peptides (see Figures A2–A4), the ensembles are even broader than for *pdiq* resulting in a larger number of clusters. The free peptide clusters for these IDP peptides are low in population and lack agreement with the preferred binding mode. However, in binding simulations at low temperature, *p53* explores a narrow conformational ensemble similar to *pdiq* and very different from the broader ensembles sampled by the control sequences. For the two control sequences, the minima of the free ensemble distribution is displaced with respect to the three other peptides, disfavoring bound-like conformations, which result in broader ensembles for the two control peptides during MELD×MD binding simulations.

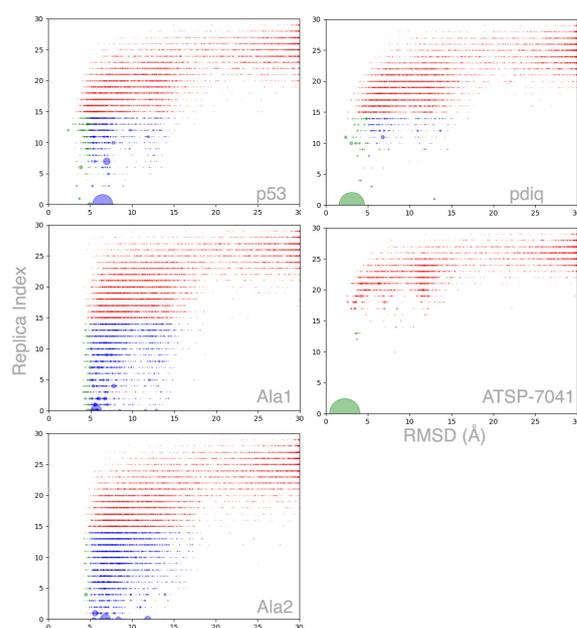


Figure 4. Funneling binding plot for the five peptides. Each dot corresponds to a cluster center from a 2D-RMSD based on all replicas. The larger the circle the larger the population of the cluster. Each circle is plotted at the average RMSD inside that cluster with respect to the native conformation and the mode of the index replica in that cluster. The color code is green ($RMSD < 5$) or blue ($RMSD > 5$) when the mode of the replica index is lower than 15, and red otherwise.

Complementary knowledge for the binding process emerges from looking at the internal structure of the peptide (radius of gyration) with respect to the position of the peptide to MDM2 (RMSD, see right column in Figures 3 and A7). At high replica index, all peptides sample conformations far from the protein, with large fluctuations in the radius of gyration (between 5 and 12 Å except for the ATSP-7041 peptide, where the chemical staple prevents conformations with a radius of gyration above 9 Å). When binding in the MDM2 hydrophobic pocket, the peptide adopts compact conformations with a radius of gyration around 7 Å. This happens early in the binding process (higher replica index) for the *pdiq* and ATSP-7041 peptides and is not observed for the *Ala2* control due to the lack of anchoring residues.

2.4. Helical Propensities Show Different Binding Patterns

The binding ensembles produce a higher helical content with respect to the free-peptide simulations. Figure A8 shows the three anchoring residues to be predominantly in helical conformations for *pdiq* and ATSP-7041. *p53* is well known to make a helix in the N-terminal region of the peptide which our simulations reproduce. The *Ala1* control sequence, which has three anchoring residues, is also able to adopt helical conformations, to a lesser extent. The spacing of the anchoring residues in the sequence (residues i , $i + 4$ and $i + 7$) and the size of the binding site favor helical conformations for the simultaneous interaction inside the active site. However, the *Ala2* control, adopts very small amounts of helical conformations for two of the three anchoring regions, consistent with the lack of anchoring residues to stabilize those conformations. Not surprisingly, the peptides with larger helical content also have narrower conformational ensembles at low replica index (ATSP-7041 and *pdiq*), whereas *p53*, *Ala1*, and *Ala2* have progressively larger ensembles at the lowest temperature replica.

2.5. MDM2 Exhibits a sMall Conformational Change upon Binding

There is a small conformational rearrangement of the backbone (1.9 Å RMSD) between the apo (unbound) and holo (bound) crystal structures (1z1m [29] and 1ycr [30], respec-

tively), which opens up the cavity for binding. Sidechain rearrangement of the surface residues happens on a faster timescale, changing the surface accessibility to the binding cavity. In MELD×MD binding, the conformational freedom of MDM2 by using flat-bottom harmonic restraints on the C_α around the holo structure to prevent unfolding in the replica exchange ladder (see methods). At high temperatures, we find that the protein is sampling conformations between 2–3.5 Å backbone RMSD from the holo structure and a similar range (2.5–4.0 Å) with respect to the apo MDM2 structure (see Figure A9). At low temperatures, the thermal ensemble is narrower, with an RMSD in between 1 to 2 Å from the holo structure and 2–4 Å from the apo structure. The presence of the peptide binding for a significant amount of time to the active site further shifts the RMSD to lower values (see *pdiq* and *ATSP-7041* in Figure A9). There are no restraints on the sidechains, in which fluctuations determine the open/closed state of the cavity. Reorientation of these sidechains is fast and adapts to the presence of the peptide near the active site.

3. Discussion

The debate between conformational selection and induced fit mechanisms of binding is being reconciled into a mixture of the two [7], with different balance of each depending of the system. In the MDM2/*p53* system, the protein undergoes a small conformational change from its apo (unbound) to holo (bound) structure, whereas the IDP peptide folds upon binding to the active site. MD simulations of the free *p53* sequence shows its intrinsically disordered nature, with little propensity for helical conformations. Querying the *p53* binding motif in the PDB returns twelve structures, covering four different protein targets: MDM2 [30], MDMX [20], p300 [31], and the CREB-binding protein [32]. In all cases, the *p53* N-terminal domain adopts a helical conformation, but with different sidechain rotameric states [32]. Our binding simulations reproduce these trend, with the size of the conformational ensemble considerably shrinking upon binding (see Figures A2 and A7). Our MELD×MD simulations lose the kinetic information of binding, but, following a particular replica as it goes up and down the replica ladder, we can observe the series of events that lead to binding. In general, we observe a higher helical content for *p53* near the protein surface, consistent with excluded volume effects [33]. The spacing of the anchoring residues (i , $i + 4$ and $i + 7$), combined with the size of the hydrophobic pocket favors binding in helical conformations. The funneling towards the binding site driven by hydrophobic interactions in MELD responds to our knowledge that the hydrophobic anchoring residues were vital for binding. Hydrophobic residues on the surface of the protein are more likely in the active site, hence accelerating binding towards it. A different choice of information (e.g., using polar residues) would have resulted in less directed sampling, as polar and charged residues are frequent in the protein surface. Visual inspection shows binding through different modes, with one of the hydrophobic residues anchoring in the cavity and driving the rest; although the peptide does not bind as a helix, it quickly adopts a partial helical structure (see Figure A8) consistent with experiments. The tryptophan seems to be critical for adopting the correct experimental binding conformations: we observe many instances of the peptide bound in the cavity as a helix with the bulky tryptophan interacting with MDM2 hydrophobic sidechains not in the binding pocket, leading to kinks in the backbone structure (see cluster 1 in Figure A2). These incorrect bindings have a long life time in our simulations and require at least partial unbinding before correctly binding according to the experimental motif, which, in our simulations, is sampled but is not the predominant state.

The *pdiq* inhibitor binds experimentally with longer helical motifs covering all amino acids. Analyzing PDB codes 1ycr and 3jzs reveals differences in the secondary structure (using STRIDE [34] as incorporated in VMD (Visual Molecular Dynamics) [35]). The difference in helicity affects the last anchoring residue (leucine), which is in a coil rotameric state for *p53* and in a helical state for *pdiq*. Our simulations show that *pdiq* forms a significant amount of helix in its free form, which favors binding significantly in our binding simulations (see Figures A1 and A8). For *pdiq*, we observe pre-binding in helical confor-

mations, shifted with respect to the experimental binding site, and fast rearrangement of the peptide, sometimes involving rapid helix unfolding and refolding in the binding site leading to the experimental structure. The helical behavior is further accentuated for the ATSP-7041 inhibitor, where all binding takes place through helical conformations thanks to the chemical staple. Rearrangements in the active site involve displacements of the alpha helix to allow better interaction of the alpha helices; this is done through partial unbinding of the helix without losing the helical character. Both control sequences access the binding site unfolded and explore many possible conformations. *Ala1* can sample the helical conformations which favor strong binding as seen from the top clusters (see Figure A3), but they have significantly lower population than the other three binding peptides (see Figure 4). On the contrary, *Ala2* has no anchoring residues and is rarely observed adopting helical conformations in the binding site (see Figure A4). These observations are supported by looking at the narrow conformational binding ensembles sampled at the lowest temperature replica for the peptides that bind (see Figures 2, A1, and A2), in contrast with the broader ensembles of the control peptides (see Figures A3 and A4). Taken together, the results show that the anchoring residues are necessary to adopt the helical conformations associated with good binding to MDM2 but not enough on their own to promote this helical state.

The ensembles at different replica index depict the nature of the binding/unbinding dynamics. The control peptides rapidly lose any memory of the bound conformation as the replica index increases (see Figures A7 and A10). For the other three peptides, we observe a bimodal distribution of states: for *p53* centered at 2/4 Å and at 1/2 Å for *pdiq* and ATSP-7041. As we increase the temperature in the replica ladder, the higher RMSD state becomes more prevalent. By replica 20, all peptides are sampling broad distributions with very low populations of the 1 RMSD state for the peptide, except for ATSP-7041, which, due to the chemical staple, even at high temperatures can sample conformations close to the bound conformation. However, at the highest replica, all peptides have lost memory of the bound conformation. Overall, we can distinguish three regions: an unbound conformation in which the peptide explores conformations far from its bound conformation, a pre-bound conformation, and a bound conformation. Both pre-bound and bound conformation lock the protein close to its holo conformation. For the poly-ala peptide, the bound state is rarely seen, while, for the poly-ala with binding side-chains, it is easier but not the predominant state. The pre-bound state for *pdiq* and ATSP-7041 is very close to the bound conformations, while, for *p53*, it is further away.

The binding free energy can be separated into a contribution coming from the conformational preferences of the peptide and protein systems, as well as an interaction contribution ($\Delta G_{bind} = \Delta G_{conformation} + \Delta G_{interaction}$), where $\Delta G_{conformation}$ can further be separated into the protein and peptide contributions ($\Delta G_{conformation} = \Delta G_{conformation}^{peptide} + \Delta G_{conformation}^{protein}$). $\Delta G_{interaction}$ is dictated by the specific interactions between the protein and peptide, which, in this case, arises from the three anchoring residues highlighted in Table 1, as shown by alanine scan mutagenesis studies [25]. Given a force field, MELD×MD samples multiple binding/unbinding events, effectively capturing both free energy contributions, even if it cannot decompose the contribution of each. Using the same protocol for all peptides allows us to identify differences in their binding preferences and peptide conformations. The main advantage is that, in this process, the peptide is completely flexible, free to adopt conformations in response to the environment. We observe the active site changing in response to the presence—and conformation—of the peptide.

Our studies hint that the binding mechanisms for *pdiq* and ATSP-7041 both favor initial binding as helices, with different mechanisms for rearrangement. Since kinetics are lost in our replica exchange ladder, testing this hypothesis will require future work in which the states discovered from our ensembles can be used for seeding unbiased simulations to construct markov models that show the binding pathways [36–38]. The chemical staple successfully increases the helical content, but it also plays a role in reducing side chain rotamer freedom through the steric volume it occupies (see Figure A11). Thus, ATSP-7041 is

predisposed to make helical conformations, and to establish the right interactions. Figure 2 shows only two clusters: a major cluster binding as a helix with the three anchoring residues in the active site and a minor one with the staple in the active site. For *pdiq*, we see a higher number of minor clusters (see Figure A1) exhibiting helical conformations, in which at least one anchoring residue is not in the active site.

Thus, for accurate modeling of the *p53*-MDM2 interaction, we need to capture: (1) the intrinsic peptide propensity to helical conformations and (2) type and alignment of the anchoring residues inside the binding cavity. Peptides that, in their free form, favor helices seem to favor binding (given the same interface residues) by reducing the $\Delta G_{conformation}^{peptide}$. However, even when shifting the helical propensities, binding simulations are needed as the binding mode can change (as we see for *ATSP-7041* and *pdiq*).

4. Materials and Methods

4.1. Choice of Peptide Systems

We chose a set of five peptides for this study: the sequence from the *p53* binding epitope, two high affinity inhibitors (*pdiq* [39] and *ATSP-7041*) [25] and two control peptides, based on the poly-Ala sequence (*Ala1* and *Ala2*; see Table 1). Of the two control sequences, *Ala1* sequence conserves the set of hydrophobic residues that allow binding, and *Ala2* does not. *ATSP-7041* is a stapled peptide using three non-standard amino acids, where one of the three anchoring residues (Leucine) is substituted by a non-canonical amino acid.

For *p53* and *pdiq*, we used crystal structures of the peptides binding to MDM2 (PDB codes 1ycr [30] and 3jzs [39]). For *ATSP-7041*, we used the structure bound to MDMX (PDB code 4n5t [25]) and superposed the active site onto MDM2 to have the reference structure of the peptide on the active site of MDM2. For the two control peptides based on poly-ALA, there is no native structure. We compare it to the *p53*-MDM2 conformation for those two peptides. Parameters for the *ATSP-7041* peptide are derived from the general amber force field (GAFF) [40], deriving charges based on the AM1 model [41].

4.2. Free Peptide Simulations

We used the ff14SB force field for amino acid sidechains [42] and the ff99SB force field for backbone parameters [43], using the GBneck2 implicit solvent model (igb = 8) [44] to improve sampling efficiency. We ran the simulations for 2 μ s using hydrogen mass repartitioning [45] with a 4fs timestep using the Amber molecular dynamics package [46]. A concern with implicit solvents is the bias towards some secondary structure [47]. However, this combination of force field with implicit solvent has shown to be reliable in reproducing the folding of peptide and protein systems [44,48,49].

4.3. MELD \times MD Binding Simulations

We ran 1 μ s-long H,T-REMD simulations using OpenMM [50] with the MELD plugin [26]. MELD allows us to incorporate noisy information to increase the sampling in regions of interest [51,52]. In this case, our interest was in observing the peptide-protein association. We required that there were at least five heavy-atom contacts between the three anchoring residues in the peptide (F, W, and L in *p53*) and any other hydrophobic residue in MDM2, and the pool of possible contacts was selected from the combinatorics of both sets. The restraints were imposed using flat-bottom harmonic restraints. The flat region was defined as a pair of residues closer than 5 Å from each other, the restraints increased quadratically up to 7 Å and linearly beyond, with a force constant of 250 J/K/mol. At every timestep, all possible restraints are evaluated, sorted by energy, and only the lowest 5 in restraint energy are used until the next timestep. In this way, no information is lost as the simulation progresses.

The H,T-REMD protocol includes 30 replicas, where the change in Hamiltonian affects the force constant of the restraints. The 30 replicas are mapped to a value of alpha (α) between 0 (lowest replica) and 1 (highest replica). The Hamiltonian and temperature have defined values of the restraint force constant and temperature as a function of alpha.

The temperature increases geometrically from 300 K ($\alpha = 0$) to 500 K ($\alpha = 0.5$) and is kept at this temperature for higher values of alpha. The force constants for the restraints is set to 0 J/K/nm² at $\alpha = 1$ and is gradually increased to the value of 250 J/K/nm² for $\alpha \leq 0.6$. Exchanges between active restraints are more likely at higher index replicas.

4.4. Clustering Analysis

We use 2D-RMSD hierarchical clustering using a single linkage scheme within cpptraj [53] and report the centroid structure of each cluster and its population as representative of the clusters. We used the last 500ns of each replica, aligning on the protein (C_α and C_β atoms) and clustering on the overlapping peptide residues (C_α and C_β atoms). For Figure 2, the lowest temperature replica and all replicas for the clustering of the funnel plots, both with $\epsilon = 1.5$. For those in Figures 2 and A1–A4, we increase ϵ to 2.0, to depict more diverse clusters.

4.5. Projections Onto a Common Feature Space

We used pyEMMA [54] to featurize our system according to phi and psi dihedrals by choosing a common set of residues on all peptide systems resulting in 22 dihedrals, and we used dihedral shifting to reduce discontinuities in the distribution rather than using sine and cosines on the dihedrals [55]. The ensembles from free *p53*, *pdiq*, and *Ala1* were chosen as a common ensemble before dimensionality reduction of the system by using time-independent coordinate analysis [56] with a lag time of 10 ns, from which we extracted the top 14 eigenvectors that account for 95% of the variance. We then projected each peptide ensemble (from free and bound simulations) into the top eigenvectors. Finally, we performed clustering of the free peptide ensembles in the space defined by the top 14 eigenvectors to produce Figures 2 and A1–A4. Since the vectors were calculated for intrinsically disordered ensembles of the free form of the peptides, they are not representative of the slowest transitions during the binding process, which we cannot extract from the MELD-biased ensembles. Nonetheless, they provide a common set of vectors to represent all free and bound peptide systems studied. For these plots, we decided to project onto the third and fifth eigenvectors since these offered the best separation between clusters for the relevant states during binding.

5. Conclusions

Predicting bound structures for IDP peptides that fold upon binding is a computational grand challenge. We have shown that possible peptide inhibitors do not necessarily bind with the same binding mode, requiring modeling approaches that allow identification of the correct binding pose. The method successfully reproduces the binding of the two inhibitors and the *p53* epitope, while showing that the two control peptides are unsuccessful binders. We further show that, by changing the intrinsic properties (e.g., helical propensity, in this case), we can identify better binders; this simplifies the design of peptide inhibitors into two distinct tasks: optimizing interface residues and optimize structural propensities. The first task requires knowing the binding mode, and the second one can be assessed by MD simulations on the free peptide, at a lower computational cost than the binding simulations. Finally, we have shown that MELD×MD is a useful tool to handle flexible binding and helps to ensure that the designed binders indeed bind and what their preferred binding mode is.

Author Contributions: Conceptualization, L.L. and A.P.; methodology, A.P.; software, L.L., A.P.; validation, L.L. and A.P.; formal analysis, L.L. and A.P.; investigation, L.L. and A.P.; resources, A.P.; data curation, L.L. and A.P.; writing—original draft preparation, A.P.; writing—review and editing, L.L., A.P.; visualization, L.L., A.P.; supervision, A.P.; project administration, A.P.; funding acquisition, A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The MELD code used to run binding simulations is available to download from github: <https://github.com/maccallumlab/meld>.

Acknowledgments: This research was supported by startup funds from the Chemistry department at the University of Florida.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

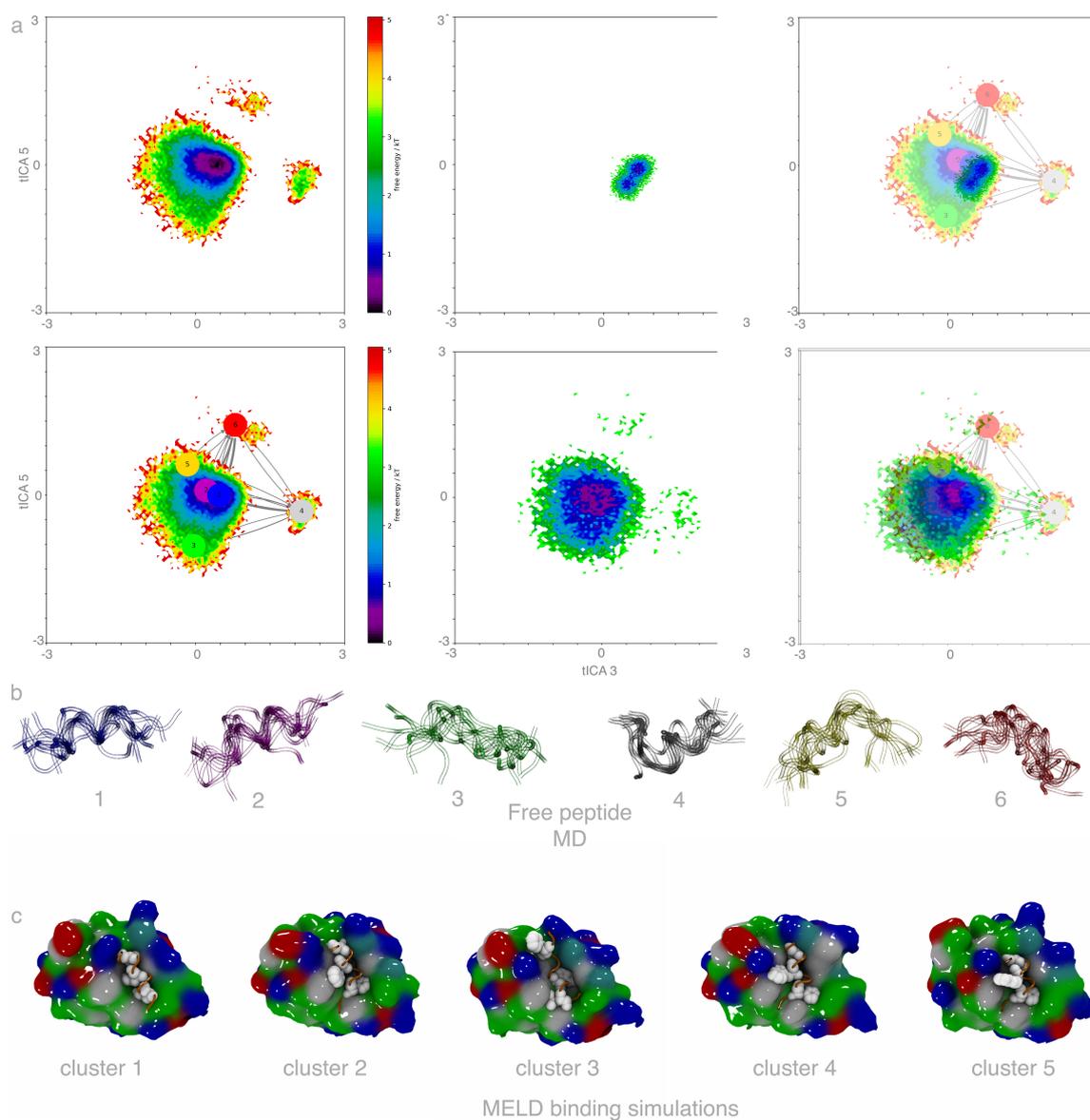


Figure A1. Comparison of the conformational space for free peptide versus binding simulations for the *pd*iq peptide. (a) The peptide ensembles are projected onto two tICA eigenvectors common to all five peptides. (b) The metastable states sampled for the free peptide. (c) Top clusters by population from MELD×MD binding simulations.

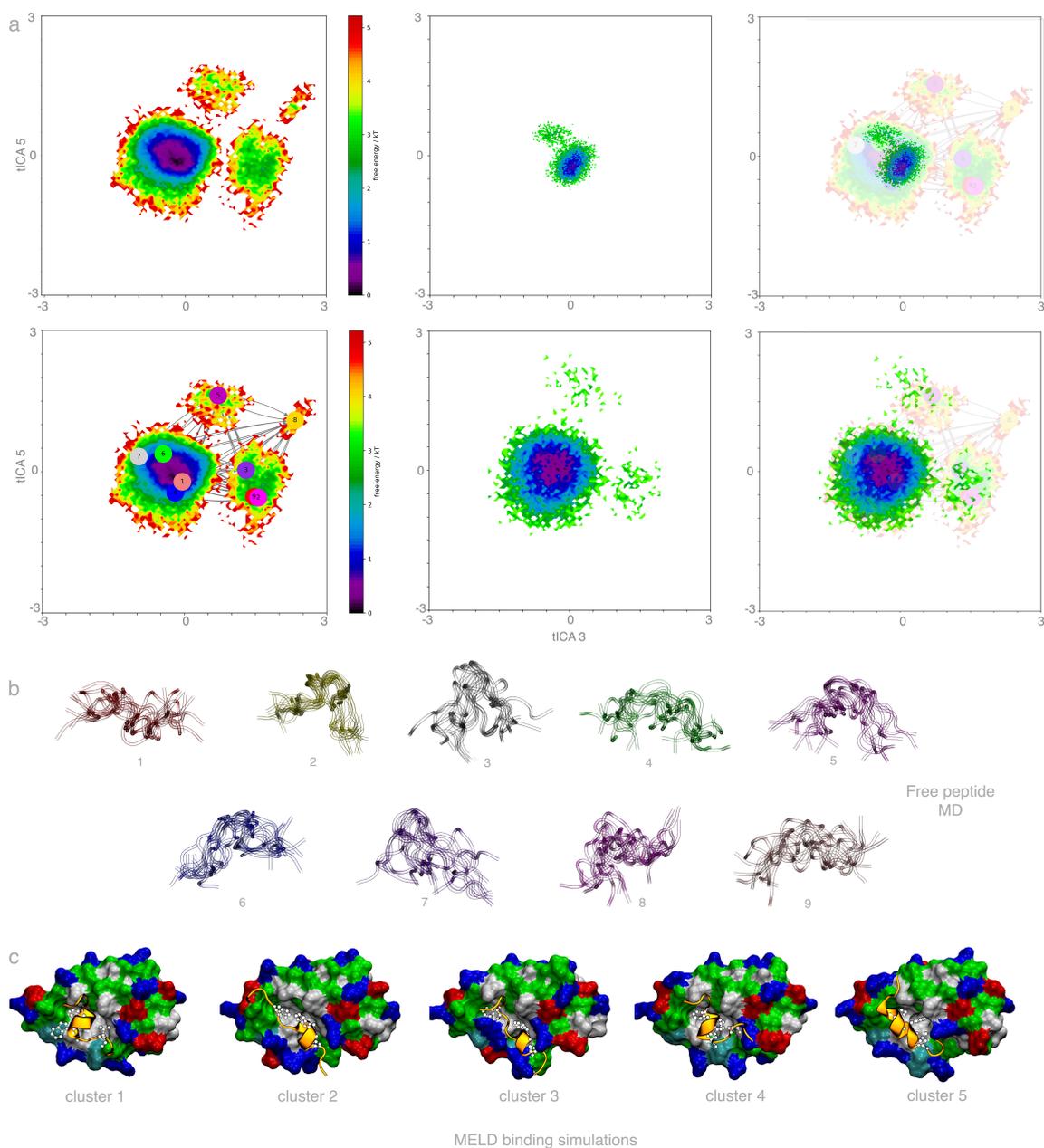


Figure A2. Comparison of the conformational space for free peptide versus binding simulations for the *p53* epitope. **(a)** The peptide ensembles are projected onto two tICA eigenvectors common to all five peptides. **(b)** The metastable states sampled for the free peptide. **(c)** Top clusters by population from MELD×MD binding simulations.

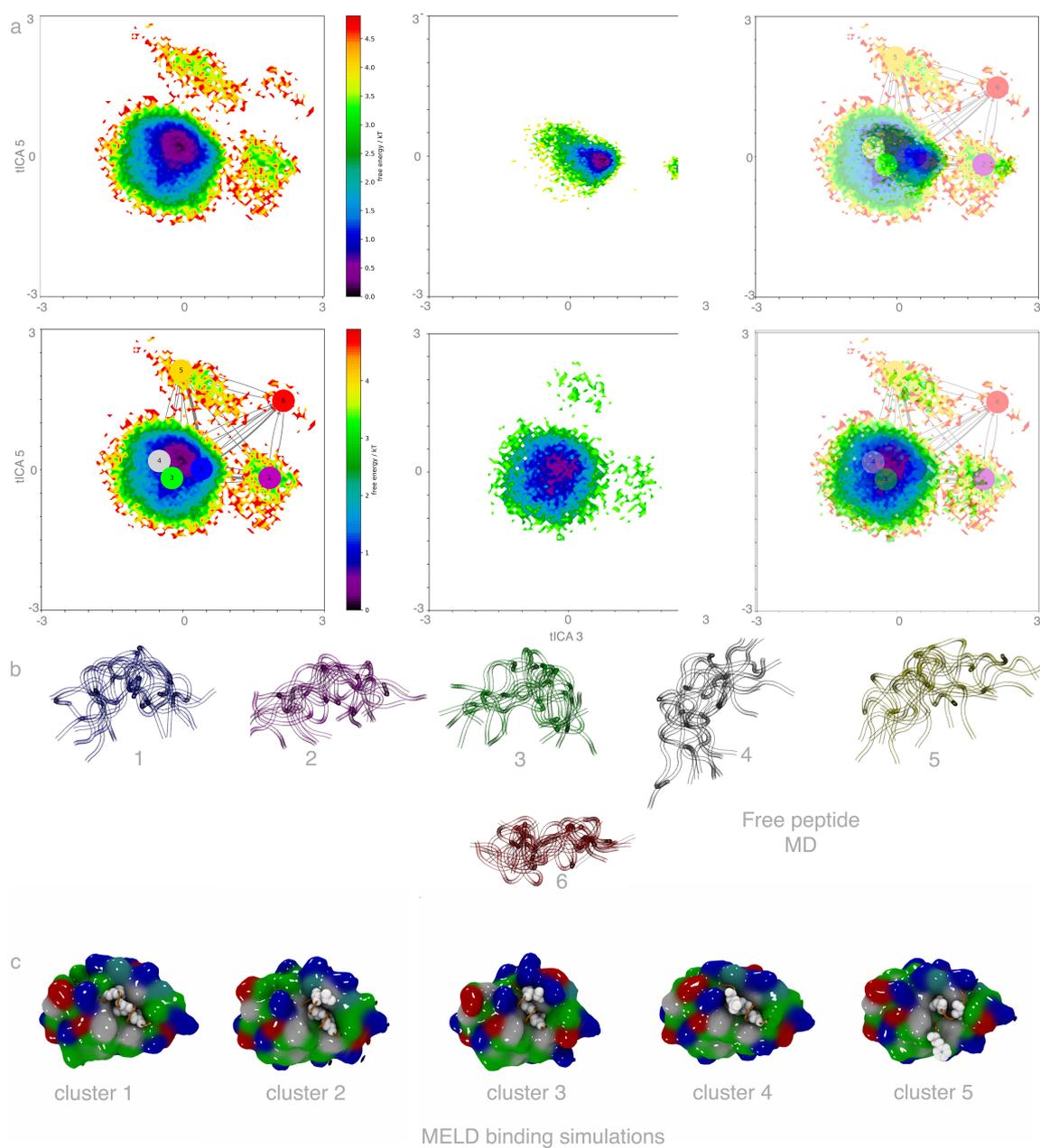


Figure A3. Comparison of the conformational space for free peptide versus binding simulations for *Ala1*. **(a)** The peptide ensembles are projected onto two tICA eigenvectors common to all five peptides. **(b)** The metastable states sampled for the free peptide. **(c)** Top clusters by population from MELD×MD binding simulations.

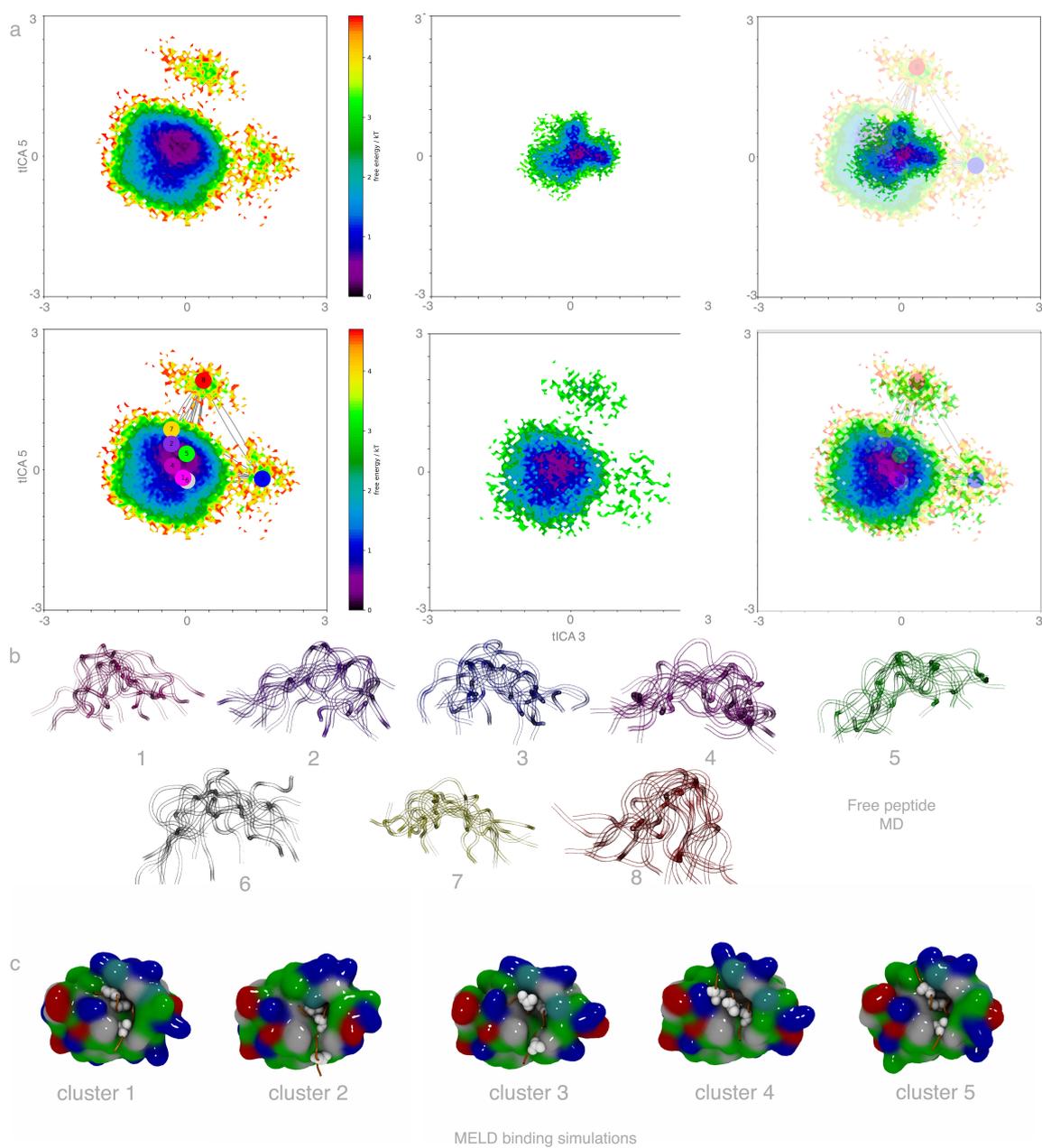


Figure A4. Comparison of the conformational space for free peptide versus binding simulations for *Ala2*. **(a)** The peptide ensembles are projected onto two tICA eigenvectors common to all five peptides. **(b)** The metastable states sampled for the free peptide. **(c)** Top clusters by population from MELD×MD binding simulations.

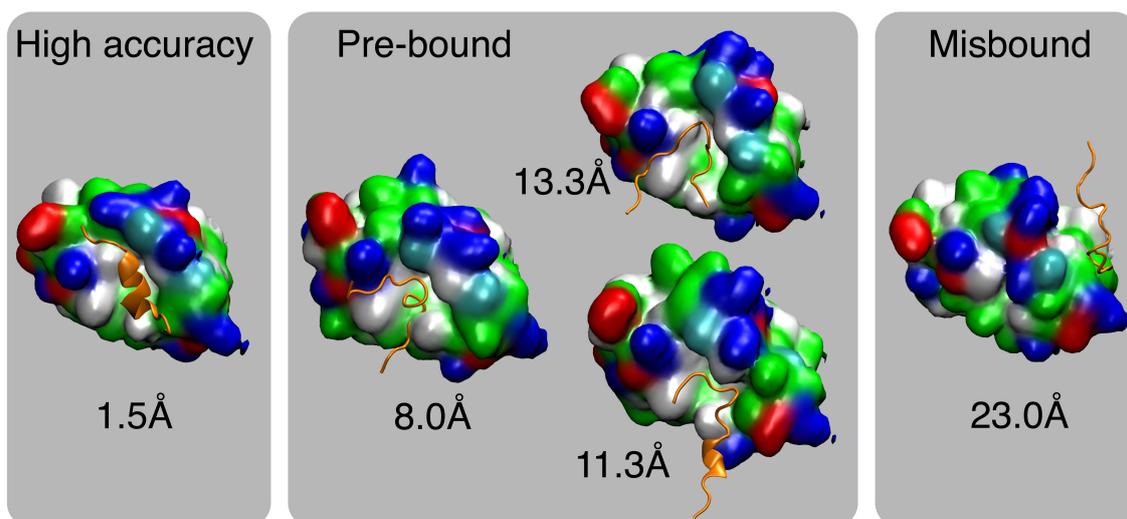


Figure A5. Clustering MELD×MD ensembles correctly identifies the MDM2 pocket as the binding site. Clustering is done by aligning on the MDM2 protein and using that alignment and the RMSD of the peptide between conformations as a distance metric. We define high accuracy binding (left) as the peptide binding in the right pocket with the right conformation (RMSD < 5 Å). For many clusters, the peptide is at least partially occupying the experimental binding site, with incorrect peptide conformations. Our approach samples these conformations at higher replica index, with a few progressing to the experimentally bound conformation. Finally, we find some conformations interacting with MDM2 at different sites in the protein (right panel).

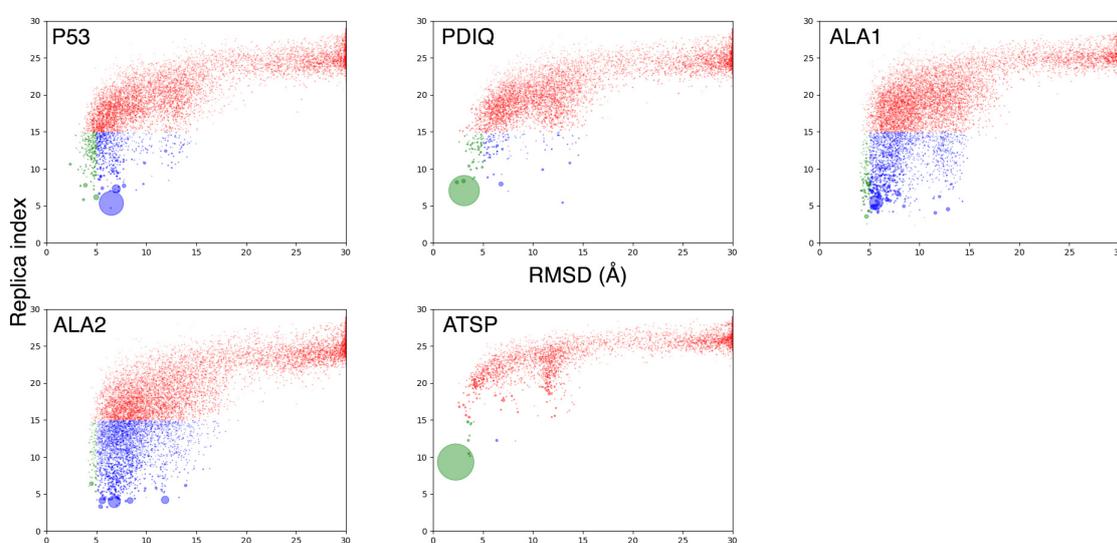


Figure A6. Funneling binding plot for the five peptides. Each dot corresponds to a cluster center from a 2D-RMSD based on all replicas. The larger the circle the larger the population of the cluster. Each circle is plotted at the average RMSD inside that cluster with respect to the native conformation and the average of the index replica in that cluster. The color code is green (RMSD < 5) or blue (RMSD > 5) when the average of the replica index is lower than 15, and red otherwise.

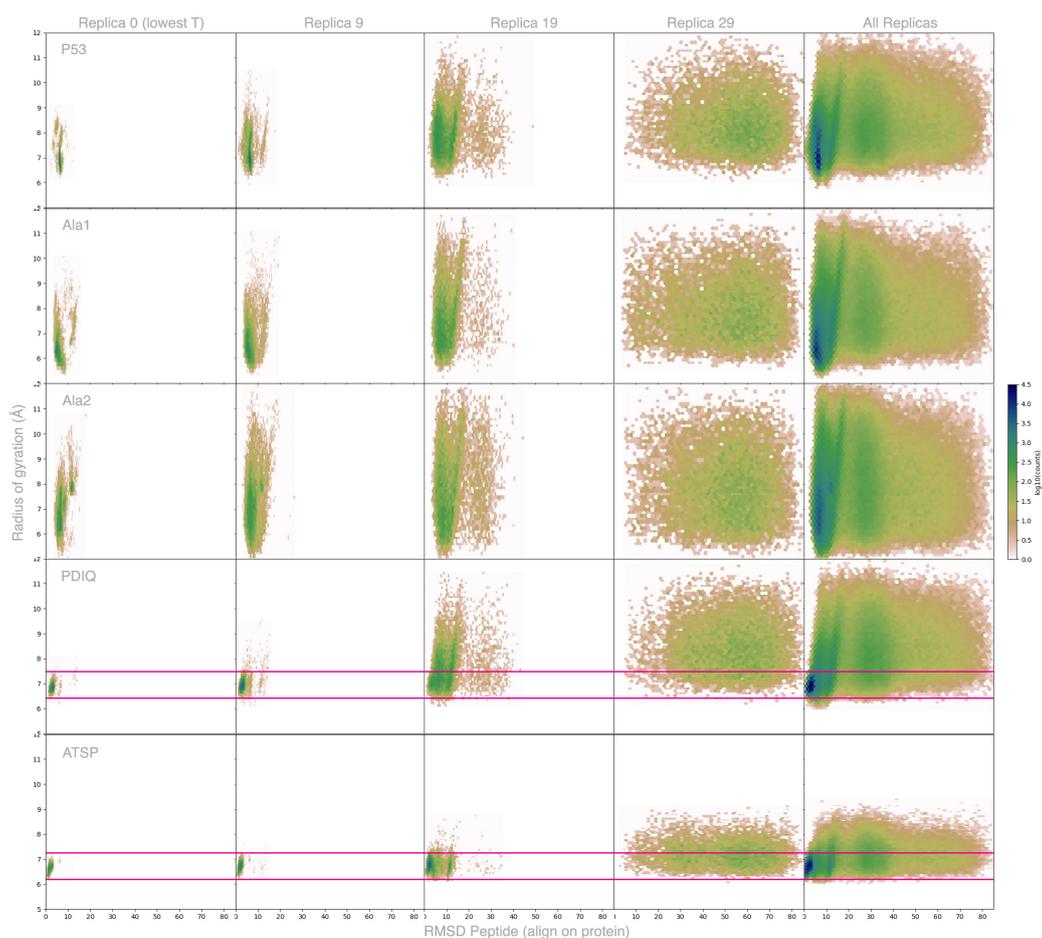


Figure A7. Radius of gyration versus RMSD of the peptide (protein align) for different replicas in MELD \times MD for all peptides studied. The red lines for peptides *pdiq* and *ATSP-7041* depict the region of the radius of gyration explored at the lowest temperature replica.

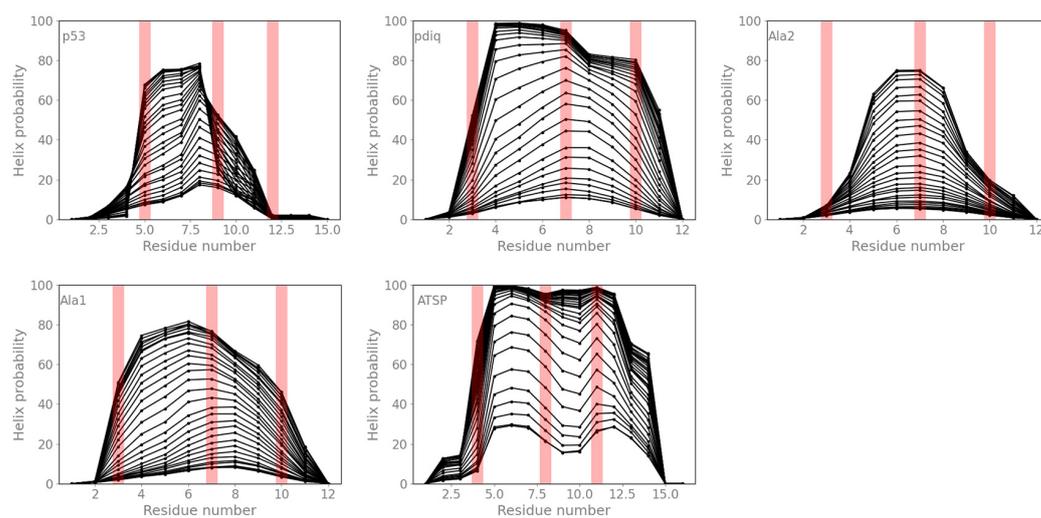


Figure A8. Secondary structure profiles for each peptide in MELD \times MD runs. Each line represents a different replica (different H and T). Replicas with higher percentage of secondary structure are sampling at low temperature and ambiguous restraints guiding the peptide to the protein at full strength. Red bars represent the location of the three anchoring residues for each peptide.

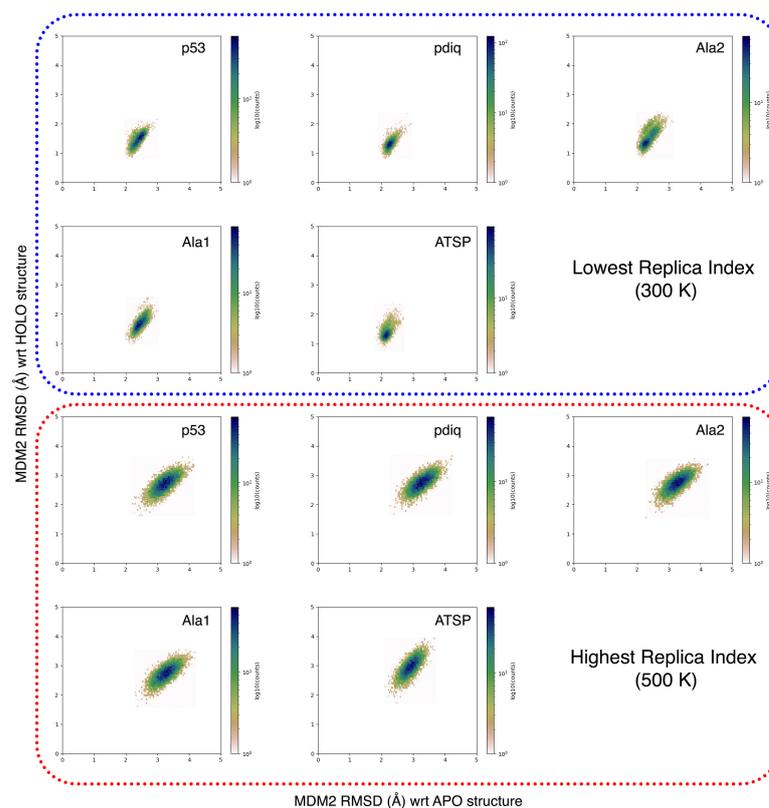


Figure A9. MDM2 protein ensembles with respect to holo and apo experimental structures at the lowest and highest index replicas. Simulations used a flat-bottom harmonic restraint on C_{α} positions with a 3.5 flat-bottom region.

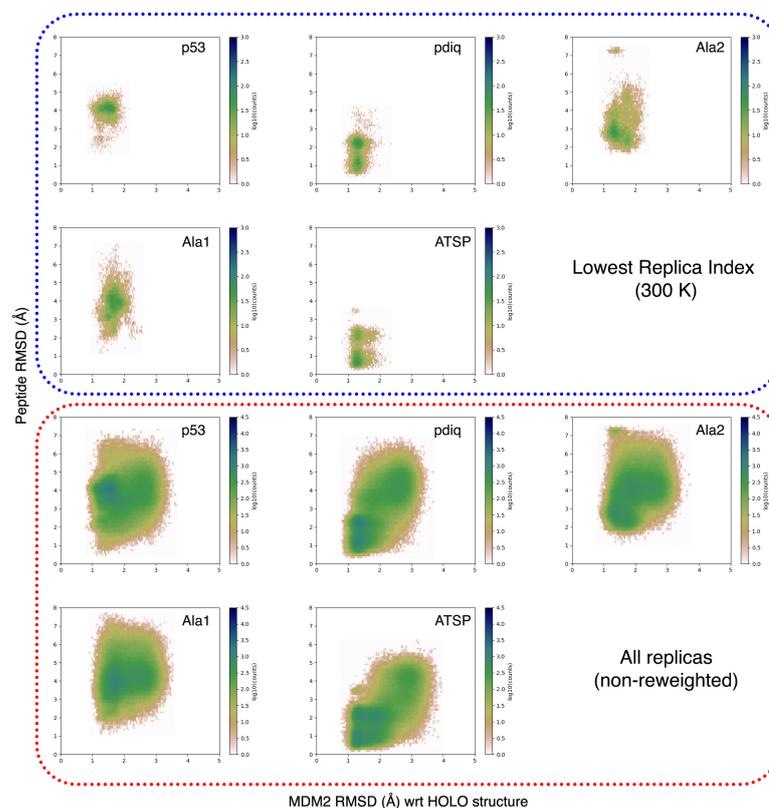


Figure A10. MDM2 versus peptide RMSD ensembles for the five peptides.

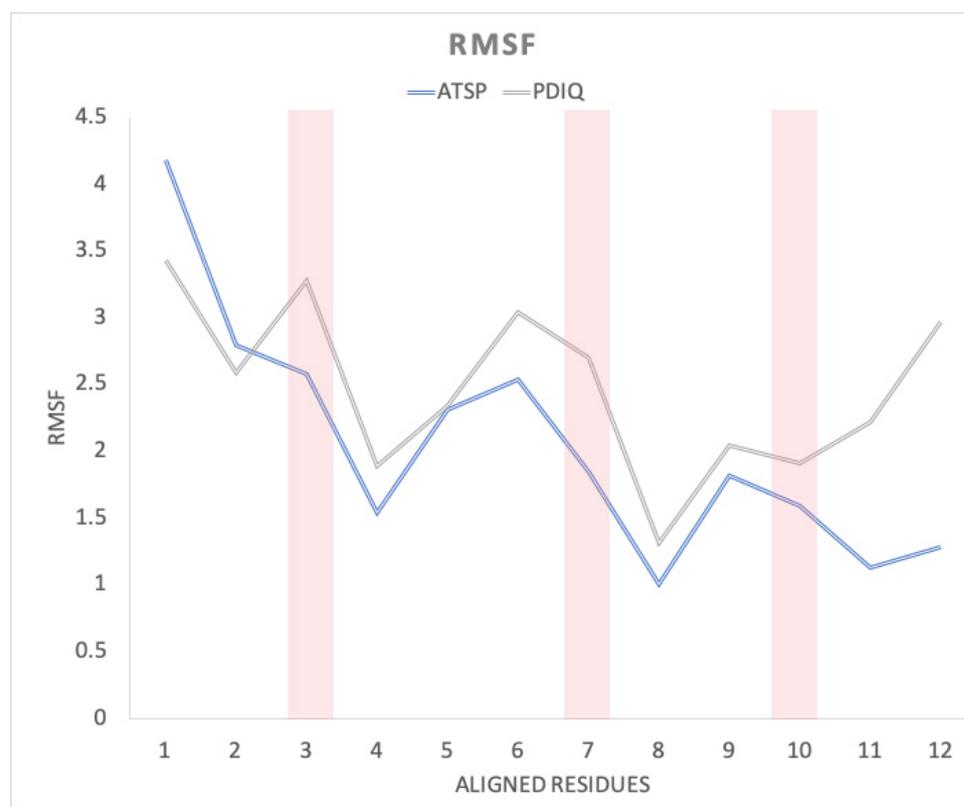


Figure A11. RMSF (Root Mean Square Fluctuation) comparison of the two best peptide binders. The red bars indicate the locations of the anchoring residues. Both peptides have been aligned to match the anchoring residues and end residues removed from ATSP-7041 for this analysis.

References

- Verdine, G.L.; Walensky, L.D. The Challenge of Drugging Undruggable Targets in Cancer: Lessons Learned from Targeting BCL-2 Family Members. *Clin. Cancer Res.* **2007**, *13*, 7264–7270. [[CrossRef](#)] [[PubMed](#)]
- Uhlig, T.; Kyprianou, T.; Martinelli, F.G.; Oppici, C.A.; Heiligers, D.; Hills, D.; Calvo, X.R.; Verhaert, P. The emergence of peptides in the pharmaceutical business: From exploration to exploitation. *EuPA Open Proteom.* **2014**, *4*, 58–69. [[CrossRef](#)]
- Weng, G.; Gao, J.; Wang, Z.; Wang, E.; Hu, X.; Yao, X.; Cao, D.; Hou, T. Comprehensive Evaluation of Fourteen Docking Programs on Protein–Peptide Complexes. *J. Chem. Theory Comput.* **2020**, *16*, 3959–3969. [[CrossRef](#)] [[PubMed](#)]
- Ciemny, M.; Kurcinski, M.; Kamel, K.; Kolinski, A.; Alam, N.; Schueler-Furman, O.; Kmiecik, S. Protein–peptide docking: Opportunities and challenges. *Drug Discov. Today* **2018**, *23*, 1530–1537. [[CrossRef](#)]
- Lee, H.; Heo, L.; Lee, M.S.; Seok, C. GalaxyPepDock: A protein–peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res.* **2015**, *43*, W431–W435. [[CrossRef](#)]
- Porter, K.A.; Xia, B.; Beglov, D.; Bohnuud, T.; Alam, N.; Schueler-Furman, O.; Kozakov, D. ClusPro PeptiDock: Efficient global docking of peptide recognition motifs using FFT. *Bioinform.* **2017**, *33*, 3299–3301. [[CrossRef](#)]
- Trellet, M.; Melquiond, A.S.J.; Bonvin, A.M.J.J. A Unified Conformational Selection and Induced Fit Approach to Protein–Peptide Docking. *PLoS ONE* **2013**, *8*, e58769. [[CrossRef](#)]
- Antunes, D.A.; Moll, M.; Devaurs, D.; Jackson, K.R.; Lizée, G.; Kavraki, L.E. DINC 2.0: A New Protein–Peptide Docking Webserver Using an Incremental Approach. *Cancer Res.* **2017**, *77*, e55–e57. [[CrossRef](#)]
- Kurcinski, M.; Jamroz, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res.* **2015**, *43*, W419–W424. [[CrossRef](#)]
- Alam, N.; Goldstein, O.; Xia, B.; Porter, K.A.; Kozakov, D.; Schueler-Furman, O. High-resolution global peptide–protein docking using fragments-based PIPER-FlexPepDock. *PLoS Comput. Biol.* **2017**, *13*, e1005905. [[CrossRef](#)]
- De Vries, S.J.; Rey, J.; Schindler, C.; Zacharias, M.; Tuffery, P. The pepATTRACT web server for blind, large-scale peptide–protein docking. *Nucleic Acids Res.* **2017**, *45*, gkx335. [[CrossRef](#)] [[PubMed](#)]
- Xu, X.; Yan, C.; Zou, X. MDockPeP: An ab-initio protein–peptide docking server. *J. Comput. Chem.* **2018**, *39*, 2409–2413. [[CrossRef](#)] [[PubMed](#)]
- Zhou, P.; Jin, B.; Li, H.; Huang, S.Y. HPEPDOCK: A web server for blind peptide–protein docking based on a hierarchical algorithm. *Nucleic Acids Res.* **2018**, *46*, gky357. [[CrossRef](#)] [[PubMed](#)]

14. Zhang, Y.; Sanner, M.F. AutoDock CrankPep: Combining folding and docking to predict protein–peptide complexes. *Bioinformatics* **2019**, *35*, 5121–5127. [[CrossRef](#)] [[PubMed](#)]
15. Shan, Y.; Kim, E.T.; Eastwood, M.P.; Dror, R.O.; Seeliger, M.A.; Shaw, D.E. How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* **2011**, *133*, 9181–3. [[CrossRef](#)]
16. Migliorini, D.; Denchi, E.L.; Danovi, D.; Jochemsen, A.; Capillo, M.; Gobbi, A.; Helin, K.; Pelicci, P.G.; Marine, J.C. Mdm4 (Mdmx) Regulates p53-Induced Growth Arrest and Neuronal Cell Death during Early Embryonic Mouse Development. *Mol. Cell Biol.* **2002**, *22*, 5527–5538. [[CrossRef](#)]
17. Marine, J.C. Chapter 3 MDM2 and MDMX in Cancer and Development. *Curr. Top. Dev. Biol.* **2011**, *94*, 45–75. [[CrossRef](#)]
18. Wade, M.; Wahl, G.M. Targeting Mdm2 and Mdmx in Cancer Therapy: Better Living through Medicinal Chemistry? *Mol. Cancer Res.* **2008**, *7*, 1–11. [[CrossRef](#)]
19. Schon, O.; Friedler, A.; Bycroft, M.; Freund, S.M.V.; Fersht, A.R. Molecular mechanism of the interaction between MDM2 and p53. *J. Mol. Biol.* **2002**, *323*, 491–501. [[CrossRef](#)]
20. Popowicz, G.M.; Czarna, A.; Rothweiler, U.; Szwagierczak, A.; Krajewski, M.; Weber, L.; Holak, T.A. Molecular basis for the inhibition of p53 by Mdmx. *Cell Cycle (Georget. Tex.)* **2007**, *6*, 2386–2392. [[CrossRef](#)]
21. Moll, U.M.; Petrenko, O. The MDM2-p53 interaction. *Mol. Cancer Res.* **2003**, *1*, 1001–1008. [[PubMed](#)]
22. Chen, H.F.; Luo, R. Binding induced folding in p53-MDM2 complex. *J. Am. Chem. Soc.* **2007**, *129*, 2930–2937. [[CrossRef](#)] [[PubMed](#)]
23. Carvajal, L.A.; Neria, D.B.; Senecal, A.; Benard, L.; Thiruthuvanathan, V.; Yatsenko, T.; Narayanagari, S.R.; Wheat, J.C.; Todorova, T.I.; Mitchell, K.; et al. Dual inhibition of MDMX and MDM2 as a therapeutic strategy in leukemia. *Sci. Transl. Med.* **2018**, *10*, eaao3003. [[CrossRef](#)] [[PubMed](#)]
24. Bernal, F.; Wade, M.; Godes, M.; Davis, T.N.; Whitehead, D.G.; Kung, A.L.; Wahl, G.M.; Walensky, L.D. A Stapled p53 Helix Overcomes HDMX-Mediated Suppression of p53. *Cancer Cell* **2010**, *18*, 411–422. [[CrossRef](#)]
25. Chang, Y.S.; Graves, B.; Guerlavais, V.; Tovar, C.; Packman, K.; To, K.H.; Olson, K.A.; Kesavan, K.; Gangurde, P.; Mukherjee, A.; et al. Stapled α -helical peptide drug development: A potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E3445–E3454. [[CrossRef](#)]
26. MacCallum, J.L.; Perez, A.; Dill, K. Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6985–6990. [[CrossRef](#)]
27. Zimmerman, M.I.; Bowman, G.R. FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *J. Chem. Theory Comput.* **2015**, *11*, 5747–5757. [[CrossRef](#)]
28. Berger-Tal, O.; Nathan, J.; Meron, E.; Saltz, D. The exploration-exploitation dilemma: A multidisciplinary framework. *PLoS ONE* **2014**, *9*, e95693. [[CrossRef](#)]
29. Uhrinova, S.; Uhrin, D.; Powers, H.; Watt, K.; Zheleva, D.; Fischer, P.; McInnes, C.; Barlow, P.N. Structure of free MDM2 N-terminal domain reveals conformational adjustments that accompany p53-binding. *J. Mol. Biol.* **2005**, *350*, 587–598. [[CrossRef](#)]
30. Kussie, P.H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A.J.; Pavletich, N.P. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **1996**, *274*, 948. [[CrossRef](#)]
31. Feng, H.; Jenkins, L.M.M.; Durell, S.R.; Hayashi, R.; Mazur, S.J.; Cherry, S.; Tropea, J.E.; Miller, M.; Wlodawer, A.; Appella, E.; et al. Structural Basis for p300 Taz2-p53 TAD1 Binding and Modulation by Phosphorylation. *Structure* **2009**, *17*, 202–210. [[CrossRef](#)] [[PubMed](#)]
32. Krois, A.S.; Ferreone, J.C.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E. Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E1853–E1862. [[CrossRef](#)] [[PubMed](#)]
33. Zhou, H.X.; Rivas, G.; Minton, A.P. Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. *Annu. Rev. Biophys.* **2008**, *37*, 375–397. [[CrossRef](#)] [[PubMed](#)]
34. Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Genet.* **1995**, *23*, 566–579. [[CrossRef](#)]
35. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
36. Plattner, N.; Doerr, S.; Fabritiis, G.D.; Noé, F. Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem.* **2017**, *9*, 1005–1011. [[CrossRef](#)] [[PubMed](#)]
37. Ferruz, N.; Fabritiis, G.D. Binding Kinetics in Drug Discovery. *Mol. Inform.* **2016**, *35*, 216–226. [[CrossRef](#)]
38. Perez, A.; Sittel, F.; Stock, G.; Dill, K. MELD-Path Efficiently Computes Conformational Transitions, Including Multiple and Diverse Paths. *J. Chem. Theory Comput.* **2018**, *14*, 2109–2116. [[CrossRef](#)]
39. Phan, J.; Li, Z.; Kasprzak, A.; Li, B.; Sebt, S.; Guida, W.; Schönbrunn, E.; Chen, J. Structure-based design of high affinity peptides inhibiting the interaction of p53 with MDM2 and MDMX. *J. Biol. Chem.* **2010**, *285*, 2174–2183. [[CrossRef](#)]
40. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [[CrossRef](#)]
41. Jakalian, A.; Jack, D.B.; Bayly, C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641. [[CrossRef](#)] [[PubMed](#)]
42. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [[CrossRef](#)] [[PubMed](#)]

43. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725. [[CrossRef](#)] [[PubMed](#)]
44. Nguyen, H.; Roe, D.R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034. [[CrossRef](#)]
45. Hopkins, C.W.; Grand, S.L.; Walker, R.C.; Roitberg, A.E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874. [[CrossRef](#)]
46. Case, D.; Belfon, K.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T.I.; Cruzeiro, V.; Darden, T.; Duke, R.; Giambasu, G.; et al. *AMBER 2020*; University of California: San Francisco, CA, USA, 2020.
47. Perez, A.; MacCallum, J.L.; Brini, E.; Simmerling, C.; Dill, K. Grid-based backbone correction to the ff12SB protein force field for implicit-solvent simulations. *J. Chem. Theory Comput.* **2015**, *11*, 4770–4779. [[CrossRef](#)]
48. Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *J. Am. Chem. Soc.* **2014**, *136*, 13959–13962. [[CrossRef](#)]
49. Robertson, J.C.; Perez, A.; Dill, K. MELD × MD Folds Nonthreadables, Giving Native Structures and Populations. *J. Chem. Theory Comput.* **2018**, *14*, 6734–6740. [[CrossRef](#)]
50. Eastman, P.; Swails, J.; Chodera, J.D.; McGibbon, R.T.; Zhao, Y.; Beauchamp, K.A.; Wang, L.P.; Simmonett, A.C.; Harrigan, M.P.; Stern, C.D.; Wiewiora, R.P.; Brooks, B.R.; Pande, V.S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659. [[CrossRef](#)]
51. Perez, A.; MacCallum, J.L.; Dill, K.A. Accelerating molecular simulations of proteins using Bayesian inference on weak information. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 11846–11851. [[CrossRef](#)]
52. Morrone, J.A.; Perez, A.; MacCallum, J.; Dill, K. Computed Binding of Peptides to Proteins with MELD-Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2017**, *13*, 870–876. [[CrossRef](#)] [[PubMed](#)]
53. Roe, D.R.; Cheatham, T.E., III. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095. [[CrossRef](#)] [[PubMed](#)]
54. Scherer, M.K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525–5542. [[CrossRef](#)] [[PubMed](#)]
55. Sittel, F.; Filk, T.; Stock, G. Principal component analysis on a torus: Theory and application to protein dynamics. *J. Chem. Phys.* **2017**, *147*, 244101. [[CrossRef](#)]
56. Pérez-Hernández, G.; Paul, F.; Giorgino, T.; Fabritiis, G.D.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139*, 015102. [[CrossRef](#)]