

Markov models of the *apo*-MDM2 lid region reveal diffuse yet two-state binding dynamics and receptor poses for computational docking

Sudipto Mukherjee^{1,+}, George A. Pantelopulos^{1,+,‡}, and Vincent A. Voelz^{1,*}

¹Department of Chemistry, Temple University, Philadelphia, PA

*voelz@temple.edu

+these authors contributed equally to this work

‡Current address: Department of Chemistry, Boston University, Boston, MA

ABSTRACT

MDM2 is a negative regulator of p53 activity and an important target for cancer therapeutics. The N-terminal lid region of MDM2 modulates interactions with p53 via competition for its binding cleft, exchanging slowly between docked and undocked conformations in the absence of p53. To better understand these dynamics, we constructed Markov State Models (MSMs) from large collections of unbiased simulation trajectories of *apo*-MDM2, and find strong evidence for diffuse, yet two-state folding and binding of the N-terminal region to the p53 receptor site. The MSM also identifies *holo*-like receptor conformations highly suitable for computational docking, despite initiating trajectories from closed-cleft receptor structures unsuitable for docking. Fixed-anchor docking studies using a test set of high-affinity small molecules and peptides show simulated receptor ensembles achieve docking successes comparable to cross-docking studies using crystal structures of receptors bound by alternative ligands. For p53, the best-scoring receptor structures have the N-terminal region lid region bound in a helical conformation mimicking the bound structure of p53, suggesting lid region association induces receptor conformations suitable for binding. These results suggest that MD+MSM approaches can sample binding-competent receptor conformations suitable for computational peptidomimetic design, and that inclusion of disordered regions may be essential to capturing the correct receptor dynamics.

Introduction

Under normal cellular conditions, the tumor suppressor protein p53 is kept at a low basal level in part due to downregulation by MDM2 (mouse double minute 2 homolog), an E3 ubiquitin ligase that recruits p53 for degradation via direct interaction with the p53 transactivation domain (TAD). Since many tumor cells still retain wild-type p53, a promising avenue of cancer treatment is to restore p53 activity by blocking the MDM2-p53 interaction with high-affinity MDM2-binding ligands.

A high-resolution x-ray crystal structure of p53 TAD bound to MDM2 in a helical conformation has been available for some time, and has spurred widespread effort towards developing inhibitors that potently disrupt p53-MDM2 binding.¹ In addition to small molecules,^{2,3} peptidomimetics have been designed to mimic the p53 helix, such as stapled peptides,⁴ beta-peptides,⁵ spirologomers,⁶ high-affinity D-peptides,⁷⁻⁹ arylamides, terphenyls, hydrogen-bond surrogates¹⁰ and oligoioxopiperazines,¹¹ many of which were developed as a result of—or in concert with—computational modeling and design.^{3,11-16}

Aside from its therapeutic interest, the p53-MDM2 interaction has served as a valuable model system for understanding protein-protein interactions, especially for intrinsically disordered proteins such as the p53 TAD that fold upon binding.¹⁷ Underscoring the importance of this work is recent evidence that residual helicity in the p53 TAD directly alters cell signaling in vivo.¹⁸ Similarly, consideration of intrinsic disorder is important to understanding MDM2, as it contains an unstructured N-terminal lid region (residues 1-25) which competes with p53 for the binding cleft. In the absence of p53, quantitative NMR spectroscopy has shown transient structuring and binding of the lid region to the p53 cleft on slow (>10 ms) exchange timescales, consistent with the structuring of a short, well-ordered helix in residues 19-24 (Figure 1).¹⁹ Recent NMR and X-ray co-crystal structures have revealed that small-molecule inhibitors can induce structuring of the lid region through specific favorable interactions,²⁰ suggesting that computational prediction lid region structure and dynamics could be very useful for computational design.

Here, to better understand the structure and dynamics of the N-terminal lid region of *apo*-MDM2, we perform extensive simulation studies to characterize the mechanism of association with the p53 binding cleft, and explore the possible role of such computational studies in drug discovery. From many independent trajectories of MDM2 starting from the *apo* state obtained by parallel distributed simulation, we construct a Markov State Model (MSM) of N-terminus dynamics that predicts two-state

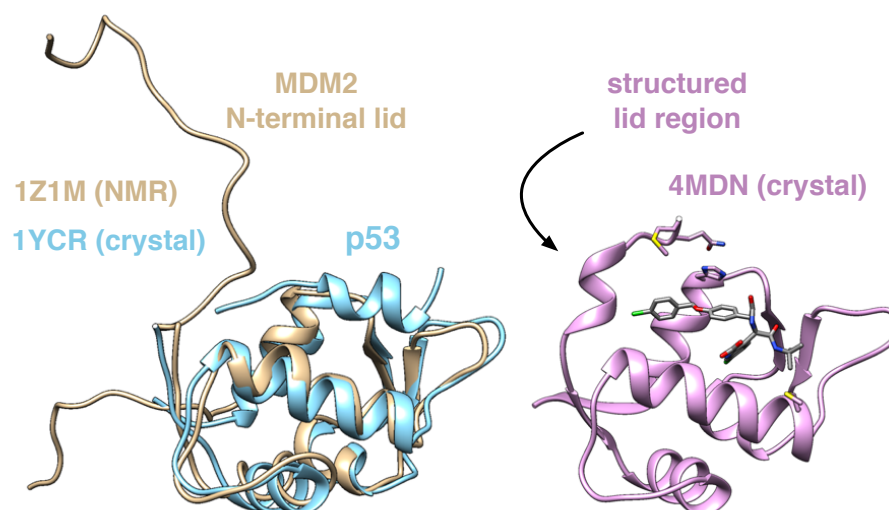


Figure 1. *Apo* and *holo* structures of MDM2. (left) The *apo* form of MDM2 (tan) has an unstructured N-terminal lid (residues 1-25) that associates with the cleft. The binding of p53 displaces the lid region from the binding cleft. Quantitative NMR of *apo*-MDM2¹⁹ has shown that a portion of the lid region (residues 19-24) slowly converts between an unstructured and a structured state. A recent co-crystal structure²⁰ with a small-molecule inhibitor (pink) shows a structured form of the lid region.

binding to the p53 cleft, in agreement with experimental findings. We then explore the utility of the MSM for *in silico* drug discovery by performing computational docking studies to kinetic metastable states of the MDM2 receptor. Remarkably, our findings suggest that the ensemble of metastable receptor conformations identified in the MSM can be used to achieve docking results similar to or better than cross-docking studies of crystal structures, and moreover, that inclusion of the N-terminus is essential in selecting open-cleft receptor conformations suitable for docking.

Results

Markov State Model (MSM) analysis of simulated *apo*-MDM2 dynamics predicts two-state binding of the lid region to the p53 cleft.

MSMs describe conformational dynamics as a network of transitions between kinetically metastable states.²¹ To construct an MSM of N-terminal dynamics from simulation data, trajectory snapshots are first assigned to metastable conformational states. To identify these metastable states, we first used tICA^{22,23} to find a low-dimensional subspace reflecting the slowest conformational motions of the N-terminal region (residues 1-25) and residues in the binding site. Projections to the two largest components (tIC1 and tIC2) were subsequently used for conformational clustering into 2000 metastable microstates, and for visualizing the folding/binding landscape.

Next, observed transitions between states are used to infer a transition matrix $\mathbf{T}^{(\tau)}$, whose elements T_{ji} contain the probability of transitioning from state i to state j within time τ . The right (ϕ_n) and left (ψ_n) eigenvectors of the transition matrix yield a complete description of state population dynamics, via the chemical master equation, $d\mathbf{p}/dt = \mathbf{K}\mathbf{p}$, where $\mathbf{T} = \exp(\tau\mathbf{K})$, whose solution is

$$\mathbf{p}(t) = \sum_n \langle \psi_n | \mathbf{p}(0) \rangle \phi_n \exp(-t/\tau_n) \quad (1)$$

Here, $\mathbf{p}(0)$ is a vector of initial state populations at time $t = 0$, and the implied timescales $\tau_n = -\tau/\ln \mu_n$ associated with each eigenmode n are related to the eigenvalues μ_n of \mathbf{T} . We define the sign structure of each eigenvector such that $\langle \psi_n | \mathbf{1} \rangle$ is positive, so that dynamics (starting from a hypothetical uniform distribution) can be described as a superposition of positive-amplitude eigenmodes ϕ_n , each decaying at time scale τ_n . The stationary eigenvector (i.e. the equilibrium state populations) is ϕ_0 , for which $\tau_n = \infty$. The resulting MSM clearly shows a two-state mechanism for lid region binding to the cleft. The sign structure of the slowest relaxation eigenmode ϕ_1 shows population flux from unbound to bound states of the lid regions, indicated by two diffuse basins aligned with tIC₁, the degree of freedom representing the slowest conformational motions (Figure 2a). Interestingly, compared to the tICA landscapes reported for many protein folding systems,^{24–26} the lid

landscape is remarkably diffuse, even in the secondary eigenmodes (Figure S4), reflecting the lack of residual structure. Similar landscapes have been found in other MSM studies of disordered proteins.^{27,28} Implied timescales computed at lag times ranging from 100 ps to 100 ns show a clear gap between the slowest and next-slowest implied timescale, indicative of apparent two-state dynamics (Figure 2b). The slowest implied timescale is close to 1 μ s, comparable to the molecular on-rate of a peptide at high effective concentration. This timescale is over four orders of magnitude faster than the slow (>10 ms) conformational exchange of residues 19-24 reported by Showalter et al., which suggests that our simulation trajectories, each shorter than 1 μ s, do not capture rare unstructuring events expected in this region. Nevertheless, the simulations show good agreement with experimentally measured chemical shifts in this region for the *apo* state, which is estimated to have $\sim 90\%$ of the lid population in an associated state.

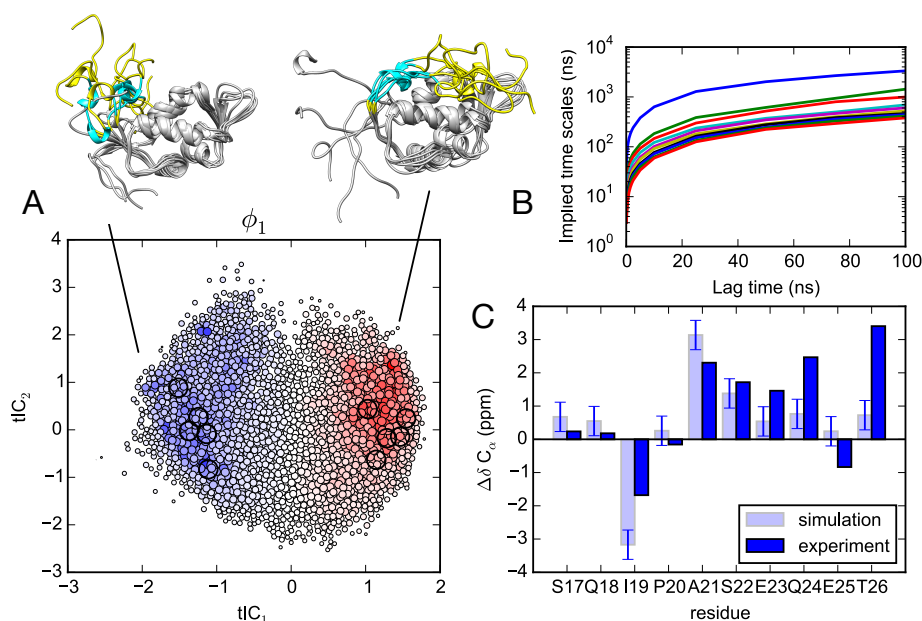


Figure 2. (A) Projection of the 2000 MSM microstates (filled circles) to tIC_1 and tIC_2 coordinates. The size of each circle is proportional to the equilibrium population, and is colored according to the slowest relaxation eigenmode, ϕ_1 . Population flux along this mode is from blue to red, representing a transition from unbound to bound states of the lid region, which we visualize using five representative structures from each basin. (B) Implied timescales versus MSM lag time show a clear gap indicating apparent two-state kinetics. (C) Simulation predictions of C_α chemical shift deviations from random-coil for the lid region (residues 17-26, cyan ribbon in panel A) calculated by SHIFTX2²⁹ agree with experimental values.¹⁹

The next-slowest eigenmode relaxation, ϕ_2 , reflects conformational dynamics of the lid region along the tIC_2 component, and it is similarly diffuse (Figure S4). To gain structural insight into these motions, we performed secondary structure analysis and Bayes Factor analysis²⁵ of interresidue contacts formed along different quadrants of the tICA projection (SI Text, Figures S4 and S5). While the slowest relaxation (along tIC_1) corresponds to disassociation of the N-terminus from the C-terminus, structuring of the lid region into a helix, and association with the binding cleft helix α_2 , the next-slowest relaxation (along tIC_2) largely reflects an increase in average self-association of the lid region, with an increase in sheet structure.

Computational docking of known MDM2 ligands to simulated receptor ensembles achieves success comparable to crystal structure cross-docking.

Virtual screening studies rely heavily on the availability of high-resolution crystal structures. Since the MDM2 trajectories were initiated from an *apo* NMR structure (PDB: 1Z1M) with a closed binding cleft unsuitable for computational docking, our work presents an excellent opportunity to test how successfully an MD+MSM approach can be used as a refinement procedure to achieve high-quality receptor structures for docking.

To evaluate the quality of simulated receptor structures, we used the DOCK6 algorithm to perform computational docking of a test set of 10 ligands to the 2000 MSM microstate structures (with the lid region removed). Our test set consisted of eight small-molecule ligands and two peptide ligands, all with high-resolution crystal structures (Table 1). The small-molecule ligands include, among others, the best-in-class inhibitor nutlin, and similar compounds. The peptide ligands include the native

p53 fragment,³⁰ and a high-affinity designed inhibitor sequence, PMI N8A.³¹ Several modifications were made to standard docking procedures to facilitate the efficient docking of peptide sequences, most notably: fixing backbone atoms in their helical conformation via an artificial cyclization bond between terminal alpha-carbons, while retaining side chain rotamer search (see Methods).

Molecule	PDB	Ligand	Rot. bonds	Affinity (nM)	Reference
1	4j7d	Nutlin RO5045331	6	20000	Fry, et al. ACS Med Chem Lett 2013
2	3jzk	chromeno-triazolo-pyrimidine	2	1230	Allen, et al. J Med Chem 2009
3	3lbn	WK23	3	916	Popowic, et al. Cell Cycle 2010
4	4j3e	Nutlin 3a	8	88	Vu, et al. ACS Med Chem Lett 2013
5	4j74	nutlin RO0503918	2	26000	Fry, et al. ACS Med Chem Lett 2013
6	4jve	morpholinone 16	8	86	de Turiso, et al. J Med Chem 2013
7	4lwt	spiroindolinone RO5027344	4	3900	Zhang, et al. Bioorg Med Chem 2014
8	4lwu	spiroindolinone RO5499252	8	5	Zhang, et al. Bioorg Med Chem 2014
9*	1ycr	p53 (ETFSDLWKLLPE)	29	440	Kussie, et al. Science 1996
10*	3lnz	N8A PMI (TSFAEYWALLSP)	21	0.49	Li et al. JMB 2010

* peptides

Table 1. Test set of small-molecule and peptide ligands of MDM2 used for computational docking studies.

To establish the baseline accuracy of the DOCK algorithm for this system, ligands were re-docked to their own co-crystal structures, and cross-docked to all other receptor structures in the test set (Figure 3).

In all cases, the best re-docking scores corresponded to a correctly docked pose, which we define as having an rmsd of 2.0 Å or less to the crystal pose, thus validating the accuracy of DOCK. Cross-docking results show the inherent variability of docking to a target receptor structure, and show that some MDM2 receptor crystal structures are more likely to produce false positives or outright failures when non-native ligands are docked. Cross-docking is the least successful for small-molecule docking to peptide-bound receptor structures, and vice versa. We also cross-docked all the ligands in the test set to the *apo*-MDM2 receptor structure (PDB:1Z1M, with the lid region removed), which confirmed its unsuitability for docking; best-scoring poses for all ligands showed rmsd values > 5.4 Å.

By comparison, computational docking to the ensemble of 2000 MSM microstates is much more successful. Plots of the DOCK score versus ligand pose rmsd show a funnel-like correlation, indicating that low scores indeed predict good ligand poses (Figure 4). Because of this, a significant enrichment in correct docking predictions is achieved. If only the five best-scoring receptor poses were considered (the top 0.25%), half of the ligands would be correctly docked; 80% are correctly docked if only the top 20 (1%) receptor poses are considered.

A potential caveat of these results is that the DOCK energy function is designed for the inexpensive evaluation of very large screening sets, at the potential cost of accuracy. For the PMI N8A peptide ligand, the lowest-energy DOCK score consistently predicts a non-native pose in which the key tryptophan and phenylalanine are placed correctly in the binding site, but with non-native sidechain rotamers, tilting the PMI helix $\sim 30^\circ$ in the binding cleft. We explored several alternative protocols designed to test whether this was due to our artificial cyclization scheme used to fix the backbone, or other search parameters; based on similar results in all cases, we conclude that scoring function accuracy is responsible.

Simulation of functional lid motions is key to successful computational docking.

Since our simulations started from an *apo*-MDM2 structure with a closed binding cleft not amenable to computational docking, we were curious to see how the functional lid motions identified in the MSM might be related to the generation of docking-competent receptor structures. A projection of the DOCK scores to the tICA landscape reveals that a significant clustering of low-scoring poses are found on the far right edge of the landscape, corresponding to states where the lid region is associated with the binding cleft (Figure 5a). This feature is more pronounced for the peptide docking results, but can also be seen clearly for the small-molecule docking results (Figure S6). In previous work, we performed a number of *apo*-MDM2 simulations in various force fields, with trajectory lengths up to 1 μ s. The projection of these data onto the tICA landscapes shows that, regardless of the force field chosen, these simulations do not sample the full extent of lid motion seen in the MSM (Figure 5b).

An inspection of the MDM2 receptor structures found on the far right of the tICA landscape, in the region of lowest DOCK scores, reveals many receptor conformations with their lid region associating with the MDM2 binding cleft. Indeed, the lowest-scoring receptor structure in this region for the p53 ligand (Figure 5a, green star) is revealed to have a helical conformation, closely mimicking the bound pose of the p53 transactivation domain (Figure 5c). In the unbound state, residues 11-17 (DGAVTTS) of the lid region have a low propensity to form a p53-like helix, forming helical structure when bound in the cleft (Figure S7).

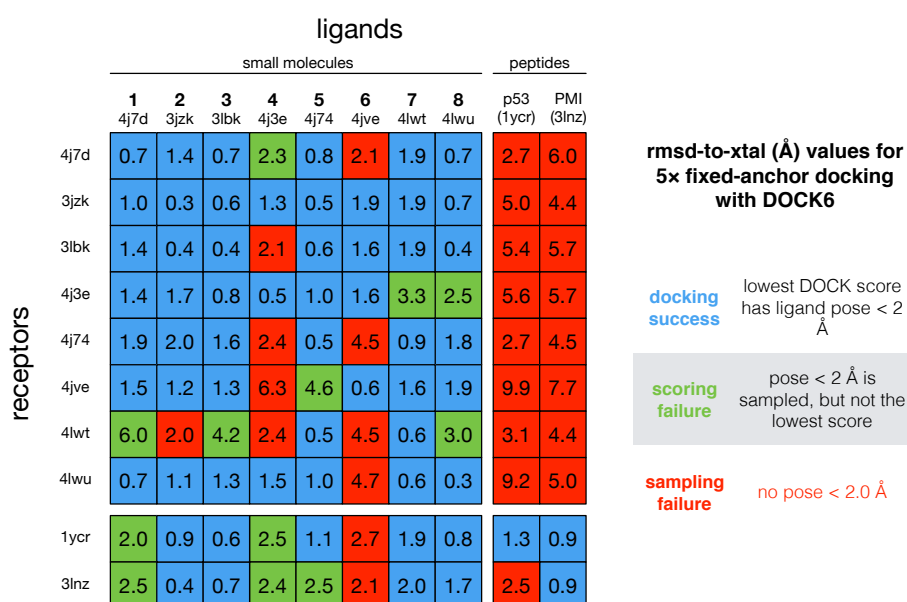


Figure 3. Self-docking and cross-docking results for a test set of 10 MDM2 ligands with available co-crystal structures (listed by PDB ID). Values shown are the rmsd (in Å) of the best-scoring docked ligand pose to the crystal ligand pose. Docking successes are shown in blue, scoring failures are shown in green, and sampling failures are shown in red.

Discussion

Recent computational studies have examined how bound ligands and/or post-translational modifications modulate the conformational dynamics of the lid region.^{33,34} The results presented here complement this work, and strongly suggest that lid region association helps to induce binding-competent, open-cleft receptor structures amenable to computational docking. This idea is consistent with the induced-fit “fly-casting” mechanism that has been proposed as the dominant mechanism of many intrinsically disordered peptides that fold upon binding,³⁵ including the p53 TAD of MDM2.¹⁷ Previous 200-ns and 1-μs simulations of *apo*-MDM2 starting from an initial closed-cleft NMR structure sample a range of open- and closed-cleft structures, but do not visit receptor structures highly competent for p53 binding, presumably because in these trajectories the lid region doesn’t sufficiently associate with the cleft to induce such structures. These findings underscore the utility of large-scale conformational sampling and analysis made possible by Markov State Model approaches. Indeed, the large-scale MD+MSM approach we use here is able to identify MDM2 receptor structures as good or better than available crystal structures for computational docking. In the future, MSMs are likely to be a valuable component of emerging molecular simulation-based methods for ensemble-based virtual screening,^{36–38} especially for homology models.³⁹

Given the known limitations in the accuracy of scoring functions for computational docking, we expect that the use of MD+MSM simulated receptor ensembles will perform even better in conjunction with more accurate energy functions, especially as a starting point for more sophisticated methods such as free energy perturbation,³² for which elucidation of relative binding modes is especially important.⁴⁰

Finally, we note that many drug targets are cell signaling proteins regulated in some way by intrinsically disordered binding partners. Many of these also have intrinsically disordered auto-inhibitory sequences that can mimic these natural substrates. For example, p53 binding partner MDMX was recently found to have an auto-inhibitory domain that inhibits binding through structural mimicry of the p53-MDMX interaction,⁴¹ a discovery which helps explain the failure of prior small-molecule drug screening efforts that did not utilize the full-length target. Similarly, our results suggest that explicit consideration of such disordered regions in simulation models may be much more important than currently appreciated, and could lead to greater functional insights and more successful computational drug discovery efforts.

Conclusion

Large-scale molecular simulation combined with Markov State Model analysis of simulated *apo*-MDM2 dynamics predicts diffuse, yet two-state binding of its disordered lid region to the p53 cleft, consistent with experiment. Computational docking of known MDM2 ligands to this simulated receptor ensemble achieves success comparable to crystal structure cross-docking, suggesting that virtual screening studies can benefit from Markov State Model approaches. These results underscore the

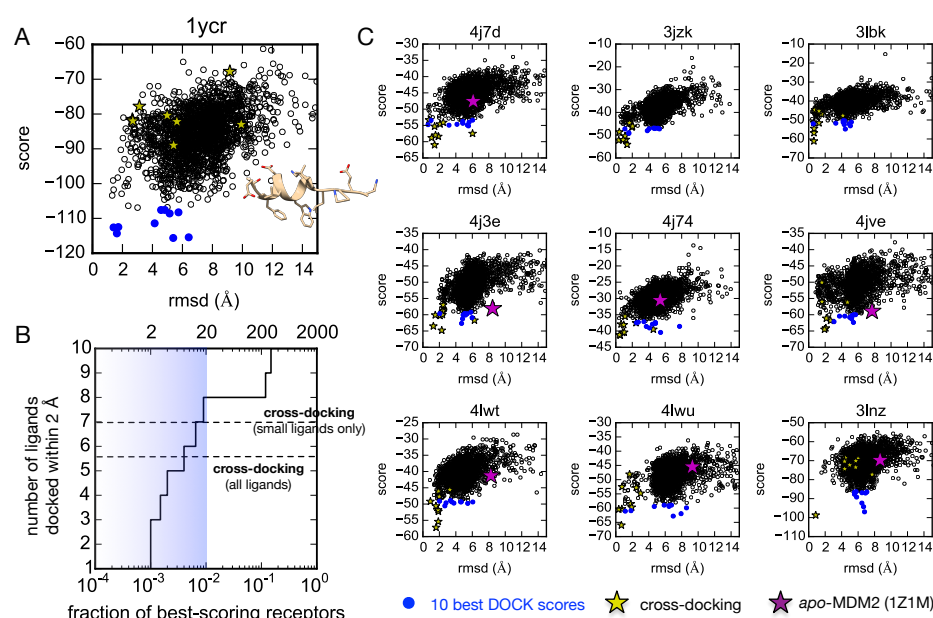


Figure 4. Scatter plots of DOCK scores versus the rmsd of the docked pose for all 2000 MSM receptor microstates show correlated funnel-like landscapes. (A) For the p53 ligand, the MSM receptor ensemble is more suitable for docking than any of the co-crystal receptor structures with other ligands. (B) The number of correct docking predictions found in some number of best-scoring poses (the true positive rate) for our test set is comparable to the cross-docking results. (C) Scatter plots for all ligands in our test set, shown with the 10 best-scoring poses docked to the MSM microstates (blue dots), cross-docking results (yellow stars), and docking results to the *apo*-MDM2 NMR structure (purple star, absences denote DOCK failures).

importance of the disordered lid region in both understanding MDM2 functional motions and in computational drug discovery.

Methods

Molecular Simulation

GROMACS 4.5 was used for all simulation preparation and production.⁴² Twenty-four initial conformations of the p53-binding region of *apo*-MDM2 (residues 1-119) were taken from the NMR-derived structure (PDB: 1Z1M).⁴³ The AMBER ff99sb-ildn-nmr force field⁴⁴ was chosen based on previous work demonstrating its accuracy and ability to predict initial structuring of the lid region in 1 μ s simulations.³² All systems were constructed as periodic cubic boxes solvated with 17268 explicit TIP3P waters and 0.1 M NaCl. Stochastic (Langevin) dynamics was simulated using a leap-frog integrator with a time step of 2 fs and an inverse friction constant of 1 ps. Non-bonded cutoffs of 0.9 nm were used for both real-space Particle-Mesh Ewald (PME) electrostatic and vdW interactions. Protein and non-protein atoms were temperature- and pressure-coupled as separate groups in the Berendsen thermostat, at 300K and 1 atm, using a 1 ps time constant, compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. Prior to production runs, all systems were equilibrated in the isothermal-isobaric (NPT) ensemble until the system volume converged to 538.71 nm³. Pro-duction runs in the canonical (NVT) ensemble were performed on the Foldinghome distributed computing network,⁴⁵ obtaining 175.7 μ s of aggregate trajectory data. The distribution of trajectory lengths is roughly exponential, with a maximum trajectory length of 945 ns, and average trajectory length of 67.0 ns (Figure S1).

Markov State Model (MSM) construction

MSMBuilder⁴⁶ was used to construct MSMs from the trajectory data. Time-lagged independent component analysis (tICA) was performed using a tICA lag time of one snapshot (100 ps), to find a low-dimensional subspace best capturing the slowest motions of the N-terminus and its binding cleft. The subspace consists of linear combinations of the set of 2304 pairwise distances between all C α atoms either in residues 1-24 of MDM2, or within 5 Å of any atom of the p53 helix in the crystal structure of *holo*-MDM2 (PDB: 1YCR). Conformational clustering in this low-dimensional subspace was used to define a set of 2000 metastable microstates. A generalized matrix Rayleigh quotient (GMRQ) method⁴⁷ was used to find optimal MSM model parameters. This analysis, which involves a cross-validation procedure wherein the trajectory data is partitioned in testing and training sets, determined that (1) *k*-centers clustering produced marginally better models than *k*-means, (2) only two tICA components were needed to accurately capture the slowest conformational motions, (3) an MSM lag time of 100 ps produced

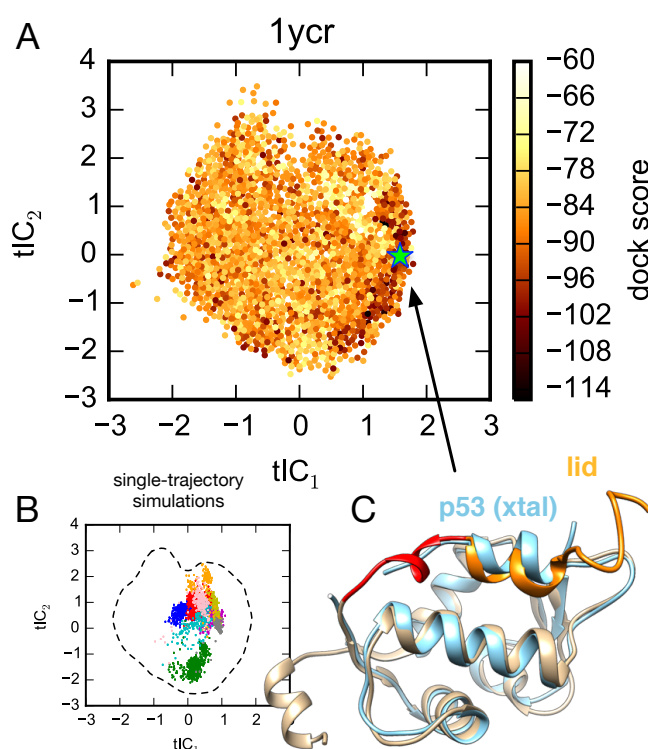


Figure 5. (A) Projection of the DOCK scores for p53+MDM2 to the tICA landscape reveals a significant clustering of low-scoring poses corresponding to lid-associated structures. (B) Previous 200-ns and 1-μs single-trajectory simulations of apo-MDM2 by Pantelopulos et al.³² projected to the tICA landscape show that simulations do not sample the full extent of lid motion seen in the MSM. Simulations were performed in force fields AMBER ff14sb (1 μs, blue), ff99sb-ildn-nmr (1 μs, red), ff99sb-ildn (200 ns, cyan), ff99sb (200 ns, yellow), ff99sb-ildn-phi (200 ns, orange), ff14sb (200 ns, magenta), CHARMM22* (1 μs, green), CHARMM36 (200 ns, pink). (C) The receptor structure with the lowest DOCK score (green star, panel A) exhibits a lid conformation closely mimicking the structure of p53 TAD bound to MDM2.

the most accurate MSMs, and (4) the GMRQ score (reflecting model quality) plateaus around 2000 microstates (Figure S2). With metastable microstates suitably defined, the matrix of transition probabilities $T_{ji}^{(\tau)}$ of transitioning from state i to state j within lag time τ was computed using a maximum-likelihood estimator from the observed transition counts.⁴⁶ Coarse-graining of MSM microstates into a 150-macrostate model was performed using the BACE algorithm.⁴⁸

Structural analysis

Analysis of trajectory data was performed using the MDTraj python library. Secondary structure populations were computed using the DSSP algorithm, with helical states corresponding to DSSP assignments G, H, I; sheet states corresponded to DSSP assignments B and E. The SHIFTX2 algorithm²⁹ was used to predict chemical shift values, using 10x subsampling of trajectory snapshots, for each MSM macrostate. To quantify the significance of interresidue contacts formed in specific conformational states, we compute a Bayes Factor (BF) contact metric for each residue pair in MDM2²⁵. More details about this are given in the Supporting Information.

Computational docking with DOCK6

Computational docking was performed using UCSF DOCK version 6.7.^{49–51} The crystal structure coordinates were downloaded from the PDB and processed using the UCSF Chimera dockprep tool.^{51,52} Small molecules were assigned AM1-BCC ligand partial charges with AmberTools antechamber,⁵³ while peptide ligands were assigned ff14SB charges. Frames taken from each of the 2000 microstate clusters were converted into DOCK-compatible MOL2 files. Owing to inconsistencies in hydrogen atom naming schemes, each such frame was reassigned optimized instantaneous protonation states using the REDUCE tool.⁵⁴ Grids at 0.3 Å-resolution were computed for each of the 2000 MD-derived frames. In order to improve sampling, each rigid segment with five or more atoms (e.g. pyrroles or larger) was used as an anchor during small molecule docking. A unique feature of the DOCK program is the anchor and grow algorithm.⁴⁹ A rigid section of the molecule, often a large aromatic

scaffold (anchor) is first oriented in the binding site. The remaining torsions are then grown one-by-one, clustering and pruning unfavorable conformations at every step until a final set of viable fully grown conformers remain. This breadth-first search approach takes exponential computational time, which severely limits docking of larger molecules. DOCK 5 was only tested on a set of molecules with seven or fewer rotatable bonds.⁵⁵ For DOCK 6.2 onwards, the addition of a fast internal energy score, coupled with aggressive pruning and rmsd symmetry, allowed reasonable performance with larger molecules (65.5% success on 8-15 torsions and 48% with >15 torsions).⁴⁹ Earlier work⁴⁹ demonstrated that despite these gains, docking success drops linearly with the number of rotatable bonds, while runtime increases exponentially. DOCK considers closed cycles in molecules to be rigid when sampling torsions. However, the simplex minimizer still relaxes local backbone conformations within these cycles. Thus, for peptide ligands, we introduced an artificial bond between the N- and C-termini to ‘rigidify’ the backbone for the purposes of docking. This ameliorates the need to fold alpha helical ligands *ab initio* with the limited molecular mechanics scoring function van der Waals and electrostatics with a distance dependent dielectric) in DOCK. In the case of the p53 TAD fragment, this reduces 66 torsions to 29 torsions after rigidifying the backbone. DOCK thus considers the backbone to be an anchor, with each sidechain torsion grown *in situ* for each receptor microstate. Cases where the receptor conformation does not (1) accommodate the backbone, or (2) allow all the sidechains to complete growth, forces the docked ligand out of the binding site, resulting in a poor interaction score.

Acknowledgements

This research was supported in part by the National Science Foundation through major research instrumentation grant number CNS-09-58854, and MCB-1412508. G.A.P. is supported by a National Science Foundation Graduate Research Fellowship under Grant DGE-1545957.

Author contributions statement

V.V. conceived the experiments, G.P. and S.M. conducted the experiments. All authors analyzed the results and reviewed the manuscript.

Additional information

Competing financial interests. No competing financial interests exist for any of the authors.

References

1. Zhao, Y., Aguilar, A., Bernard, D. & Wang, S. Small-molecule inhibitors of the mdm2–p53 protein–protein interaction (mdm2 inhibitors) in clinical trials for cancer treatment. *Journal of Medicinal Chemistry* **58**, 1038–1052 (2015). URL <http://pubs.acs.org/doi/10.1021/jm501092z>.
2. Estrada-Ortiz, N., Neochoritis, C. G. & Dömling, A. How to design a successful p53-mdm2/x interaction inhibitor: A thorough overview based on crystal structures. *ChemMedChem* (2015). URL <http://doi.wiley.com/10.1002/cmdc.201500487>.
3. Rew, Y. *et al.* Structure-based design of novel inhibitors of the mdm2-p53 interaction. *Journal of Medicinal Chemistry* **55**, 4936–4954 (2012). URL <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22524527&retmode=ref&cmd=prlinks>.
4. Baek, S. *et al.* Structure of the stapled p53 peptide bound to mdm2. *Journal of the American Chemical Society* **134**, 103–106 (2012). URL <http://pubs.acs.org/doi/abs/10.1021/ja2090367>.
5. Kritzer, J. A., Lear, J. D., Hodsdon, M. E. & Schepartz, A. Helical β -peptide inhibitors of the p53-hdm2 interaction. *Journal of the American Chemical Society* **126**, 9468–9469 (2004). URL <http://pubs.acs.org/doi/abs/10.1021/ja031625a>.
6. Brown, Z. Z. *et al.* A spirologomer α -helix mimic that binds hdm2, penetrates human cells and stabilizes hdm2 in cell culture. *PLoS ONE* **7**, e45948 (2012). URL <http://dx.plos.org/10.1371/journal.pone.0045948.s001http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3475717/pdf/pone.0045948.pdf>.
7. Liu, M. *et al.* A left-handed solution to peptide inhibition of the p53-mdm2 interaction. *Angewandte Chemie International Edition* **49**, 3649–3652 (2010). URL <http://doi.wiley.com/10.1002/anie.201000329>.
8. Pazgier, M. *et al.* Structural basis for high-affinity peptide inhibition of p53 interactions with mdm2 and mdmx. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 4665–4670

- (2009). URL <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19255450&retmode=ref&cmd=prlinks>.
9. Zhan, C. *et al.* An ultrahigh affinity d-peptide antagonist of mdm2. *Journal of Medicinal Chemistry* **55**, 6237–6241 (2012). URL <http://pubs.acs.org/doi/abs/10.1021/jm3005465><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3426300/pdf/nihms388782.pdf>.
10. Henchey, L. K., Jochim, A. L. & Arora, P. S. Contemporary strategies for the stabilization of peptides in the α -helical conformation. *Current Opinion in Chemical Biology* **12**, 692–697 (2008). URL <http://linkinghub.elsevier.com/retrieve/pii/S1367593108001257>.
11. Lao, B. B. *et al.* Rational design of topographical helix mimics as potent inhibitors of protein–protein interactions. *Journal of the American Chemical Society* **136**, 7877–7888 (2014). URL <http://pubs.acs.org/doi/abs/10.1021/ja502310r>.
12. Michel, J., Harker, E. A., Tirado-Rives, J., Jorgensen, W. L. & Schepartz, A. In silico improvement of β 3-peptide inhibitors of p53•hdm2 and p53•hdmx. *Journal of the American Chemical Society* **131**, 6356–6357 (2009). URL <http://pubs.acs.org/doi/abs/10.1021/ja901478e>.
13. Fuller, J. C., Jackson, R. M., Edwards, T. A., Wilson, A. J. & Shirts, M. R. Modeling of arylamide helix mimetics in the p53 peptide binding site of hdm2 suggests parallel and anti-parallel conformations are both stable. *PLoS ONE* **7**, e43253 (2012). URL <http://dx.plos.org/10.1371/journal.pone.0043253.g010>.
14. Fuller, J. C., Jackson, R. M. & Shirts, M. R. Configurational preferences of arylamide α -helix mimetics via alchemical free energy calculations of relative binding affinities. *The Journal of Physical Chemistry B* **116**, 10856–10869 (2012). URL <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22920218&retmode=ref&cmd=prlinks>.
15. ElSawy, K. M., Lane, D. P., Verma, C. S. & Caves, L. S. D. Recognition dynamics of p53 and mdm2: Implications for peptide design. *The Journal of Physical Chemistry B* **120**, 320–328 (2016). URL <http://pubs.acs.org/doi/abs/10.1021/acs.jpcc.5b11162>.
16. 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. *Chemical Biology & Drug Design* **83**, 631–642 (2014). URL <http://doi.wiley.com/10.1111/cbdd.12284>.
17. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* **16**, 18–29 (2015). URL <http://dx.doi.org/10.1038/nrm3920>.
18. Borchers, W. *et al.* Disorder and residual helicity alter p53-mdm2 binding affinity and signaling in cells. *Nature Chemical Biology* **1**–5 (2014). URL <http://dx.doi.org/10.1038/nchembio.1668>.
19. Showalter, S. A., Bruschweiler-Li, L., Johnson, E., Zhang, F. & Bruschweiler, R. Quantitative lid dynamics of mdm2 reveals differential ligand binding modes of the p53-binding cleft. *Journal of the American Chemical Society* **130**, 6472–6478 (2008). URL <http://pubs.acs.org/doi/abs/10.1021/ja800201j>.
20. Bista, M. *et al.* Transient protein states in designing inhibitors of the mdm2-p53 interaction. *Structure* **21**, 2143–2151 (2013). URL <http://dx.doi.org/10.1016/j.str.2013.09.006>.
21. Chodera, J. D. & Noé, F. Markov state models of biomolecular conformational dynamics. *Current opinion in structural biology* **25**, 135–144 (2014). URL <http://dx.doi.org/10.1016/j.sbi.2014.04.002>.
22. Schwantes, C. R. & Pande, V. S. Improvements in markov state model construction reveal many non-native interactions in the folding of ntl9. *Journal of Chemical Theory and Computation* **9**, 2000–2009 (2013). URL <http://pubs.acs.org/doi/abs/10.1021/ct300878a>.
23. Perez-Hernandez, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noé, F. Identification of slow molecular order parameters for markov model construction. *The Journal of Chemical Physics* **139**, 015102 (2013). URL <http://scitation.aip.org/content/aip/journal/jcp/139/1/10.1063/1.4811489>.
24. Razavi, A. M. & Voelz, V. A. Kinetic network models of tryptophan mutations in β -hairpins reveal the importance of non-native interactions. *Journal of Chemical Theory and Computation* **11**, 2801–2812 (2015). URL <http://pubs.acs.org/doi/abs/10.1021/acs.jctc.5b00088>.
25. Zhou, G. & Voelz, V. A. Using kinetic network models to probe non-native salt-bridge effects on α -helix folding. *Journal of Physical Chemistry B* **120**, 926–935 (2016).

26. Boninsegna, L., Gobbo, G., Noé, F. & Clementi, C. Investigating molecular kinetics by variationally optimized diffusion maps. *Journal of Chemical Theory and Computation* **11**, 5947–5960 (2015). URL <http://pubs.acs.org/doi/10.1021/acs.jctc.5b00749>.
27. Lin, Y.-S., Bowman, G. R., Beauchamp, K. A. & Pande, V. S. Investigating how peptide length and a pathogenic mutation modify the structural ensemble of amyloid beta monomer. *Biophysical journal* **102**, 315–324 (2012). URL <http://linkinghub.elsevier.com/retrieve/pii/S0006349511053653>.
28. Qiao, Q., Bowman, G. R. & Huang, X. Dynamics of an intrinsically disordered protein reveal metastable conformations that potentially seed aggregation. *Journal of the American Chemical Society* 130910133351004 (2013). URL <http://pubs.acs.org/doi/abs/10.1021/ja403147m>.
29. Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. Shiftx2: significantly improved protein chemical shift prediction. *Journal of Biomolecular NMR* **50**, 43–57 (2011). URL <http://www.springerlink.com/index/10.1007/s10858-011-9478-4>.
30. Kussie, P. H. *et al.* Structure of the mdm2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **274**, 948–953 (1996). URL <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=8875929&retmode=ref&cmd=prlinks>.
31. Li, C. *et al.* Systematic mutational analysis of peptide inhibition of the p53–mdm2/mdmx interactions. *Journal of molecular biology* **398**, 200–213 (2010). URL <http://dx.doi.org/10.1016/j.jmb.2010.03.005http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2856455/pdf/nihms187647.pdf>.
32. Pantelopulos, G. A., Mukherjee, S. & Voelz, V. A. Microsecond simulations of mdm2 and its complex with p53 yield insight into force field accuracy and conformational dynamics. *Proteins: Structure, Function, and Bioinformatics* **83**, 1665–1676 (2015). URL <http://onlinelibrary.wiley.com/doi/10.1002/prot.24852/full>.
33. Bueren-Calabuig, J. A. & Michel, J. Elucidation of ligand-dependent modulation of disorder-order transitions in the oncoprotein mdm2. *PLoS computational biology* **11**, e1004282 (2015). URL <http://dx.plos.org/10.1371/journal.pcbi.1004282.s010>.
34. Bueren-Calabuig, J. A. & Michel, J. Impact of ser17 phosphorylation on the conformational dynamics of the oncoprotein mdm2. *Biochemistry* **acs.biochem.6b00127–36** (2016). URL <http://pubs.acs.org/doi/abs/10.1021/acs.biochem.6b00127>.
35. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025 (2007).
36. Cheng, L. S. *et al.* Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. *Journal of Medicinal Chemistry* **51**, 3878–3894 (2008). URL <http://pubs.acs.org/doi/abs/10.1021/jm8001197>.
37. Malmstrom, R. D., Lee, C. T., Van Wart, A. T. & Amaro, R. E. Application of molecular-dynamics based markov state models to functional proteins. *Journal of Chemical Theory and Computation* 140611130813008 (2014). URL <http://pubs.acs.org/doi/abs/10.1021/ct5002363>.
38. Shukla, D., Hernández, C. X., Weber, J. K. & Pande, V. S. Markov state models provide insights into dynamic modulation of protein function. *Accounts of Chemical Research* **48**, 150103080101005 (2015). URL <http://pubs.acs.org/doi/abs/10.1021/ar5002999>.
39. Choi, J., Choi, K.-E., Park, S. J., Kim, S. Y. & Jee, J.-G. Ensemble-based virtual screening led to the discovery of new classes of potent tyrosinase inhibitors. *Journal of Chemical Information and Modeling* **56**, 354–367 (2016). URL <http://pubs.acs.org/doi/abs/10.1021/acs.jcim.5b00484>.
40. Mobley, D. L. & Klimovich, P. V. Perspective: Alchemical free energy calculations for drug discovery. *The Journal of Chemical Physics* **137**, 230901 (2012). URL <http://link.aip.org/link/JCPSA6/v137/i23/p230901/s1&Agg=doi>.
41. Chen, L. *et al.* Autoinhibition of mdmx by intramolecular p53 mimicry. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 4624–4629 (2015). URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1420833112>.
42. Pronk, S. *et al.* Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013). URL [GotoISI://WOS:000316695700004](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3695700/).

43. Uhrinova, S. *et al.* Structure of free mdm2 n-terminal domain reveals conformational adjustments that accompany p53-binding. *Journal of Molecular Biology* **350**, 587–598 (2005). URL GotoISI://WOS:000230296700015.
44. Li, D.-W. & Brüschweiler, R. Nmr-based protein potentials. *Angewandte Chemie International Edition* **49**, 6778–6780 (2010). URL <http://doi.wiley.com/10.1002/anie.201001898>.
45. Shirts, M. & Pande, V. S. Screen savers of the world, unite! *Science* **290**, 1903–1904 (2000).
46. Beauchamp, K. A. *et al.* Msmbuilder2: Modeling conformational dynamics on the picosecond to millisecond scale. *Journal of Chemical Theory and Computation* **7**, 3412–3419 (2011). URL <http://pubs.acs.org/doi/abs/10.1021/ct200463m>.
47. McGibbon, R. T. & Pande, V. S. Variational cross-validation of slow dynamical modes in molecular kinetics. *The Journal of Chemical Physics* **142**, 124105 (2015). URL <http://scitation.aip.org/content/aip/journal/jcp/142/12/10.1063/1.4916292>.
48. Bowman, G. R. Improved coarse-graining of markov state models via explicit consideration of statistical uncertainty. *The Journal of Chemical Physics* **137**, 134111 (2012). URL <http://link.aip.org/link/JCPSA6/v137/i13/p134111/s1&Apg=doi>.
49. Allen, W. J. *et al.* Dock 6: Impact of new features and current docking performance. *Journal of Computational Chemistry* **36**, 1132–1156 (2015). URL <http://doi.wiley.com/10.1002/jcc.23905>.
50. Brozell, S. R. *et al.* Evaluation of dock 6 as a pose generation and database enrichment tool. *Journal of Computer-Aided Molecular Design* **26**, 749–773 (2012). URL <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22569593&retmode=ref&cmd=prlinks>.
51. Lang, P. T. *et al.* Dock 6: combining techniques to model rna-small molecule complexes. *RNA* **15**, 1219–1230 (2009). URL <http://rnajournal.cshlp.org/content/15/6/1219.full>.
52. Pettersen, E. F. *et al.* Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605–1612 (2004). URL <http://doi.wiley.com/10.1002/jcc.20084>.
53. Case, C. *et al.* Amber 14. University of California, San Francisco (2015).
54. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *Journal of molecular biology* **285**, 1735–1747 (1999). URL <http://linkinghub.elsevier.com/retrieve/pii/S0022283698924019>.
55. Moustakas, D. T. *et al.* Development and validation of a modular, extensible docking program: Dock 5. *Journal of Computer-Aided Molecular Design* **20**, 601–619 (2006). URL <http://link.springer.com/article/10.1007/s10822-006-9060-4>.
56. Nüske, F., Keller, B. G., Perez-Hernandez, G., Mey, A. S. J. S. & Noé, F. Variational Approach to Molecular Kinetics. *Journal of Chemical Theory and Computation* **10**, 1739–1752 (2014).
57. Noé, F. & Clementi, C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *Journal of Chemical Theory and Computation* **11**, 5002–5011 (2015).
58. Boninsegna, L., Gobbo, G., Noé, F. & Clementi, C. Investigating Molecular Kinetics by Variationally Optimized Diffusion Maps. *Journal of Chemical Theory and Computation* **11**, 5947–5960 (2015).

Supplementary Information

Markov models of the *apo*-MDM2 lid region reveal diffuse yet two-state binding dynamics and receptor poses for computational docking

Sudipto Mukherjee, George A. Pantelopulos and Vincent A. Voelz
Department of Chemistry, Temple University, Philadelphia, PA

Supporting Text

Generalized Matrix Raleigh Quotient (GMRQ) Analysis

To choose optimal parameters for MSM construction, we use the GMRQ method recently developed by McGibbon et al.¹ This method exploits the variational principle of conformational dynamics,⁵⁶⁻⁵⁸ that any approximation to the true eigenvectors of a dynamical operator will necessarily underestimate its eigenvalues, which correspond to the relaxation timescales. For an MSM of n discrete conformational states indexed by $k = 1 \dots n$, approximations to the eigenfunctions are linear combinations of indicator basis set functions $\phi_k(x)$ (equal to 1 if x is in state k and 0 otherwise.) The generalized matrix Raleigh quotient R quantifies how well a given linear combination of basis functions captures the m slowest eigenvectors/timescales. It is computed as

$$R(A|C, S) \equiv \text{Tr}\left(\frac{A^T C A}{A^T S A}\right) \quad (2)$$

where A is an $n \times m$ matrix whose j^{th} column contains the linear coefficients to approximate the j^{th} c eigenvector as $\sum_k A_{kj} \phi_k(x)$, C is an $n \times n$ time-lagged correlation matrix of the indicator basis functions calculated from the trajectory data, and S is an $n \times n$ covariance matrix of the indicator basis functions. For an MSM, both C and S can be estimated from the number of transitions between states observed in the simulation trajectories. Finding the coefficient matrix A that maximizes R is the same eigenproblem solved in the tICA method.^{2,3} Once maximized, R can be used to evaluate the quality of the MSM state decomposition by how well it captures the true conformational dynamics.

To avoid overfitting to the data, a cross-validation approach is used in which the simulation data is partitioned into 5-fold leave-one-out training/testing sets. In five separate trials, R is maximized using 4/5 of the trajectory data, which we report as the GMRQ training score. Using these optimal values of the coefficients in A , the remaining 1/5 of the data is then used to compute R , which we report as the GMRQ testing score.

To select the MSM that most accurately captures the conformational dynamics of the MDM2 lid region, we explored various model construction parameters and chose the model with the largest GMRQ testing score (Figure S2). The time step between collected trajectory snapshots is $\tau_0 = 100$ ps. We explored MSMs built using different tICA lag times (i.e. the lag time to calculate the distance correlation matrix needed to find the tICA components) and different MSM lag times. For MSMs constructed using a range of 200 to 2000 microstates, a lag time of 1 (in units of τ_0) is optimal (Figure S2a,b). We also explored MSMs constructed using different numbers of tICA components (2 to 15 tICs) as the subspace to project trajectory data and perform conformational clustering. We found that, for MSMs constructed using a range of 200 to 2000 microstates, projections utilizing 2 tICs give the best GMRQ testing scores (Figure S2c). In all of our tests, GMRQ testing scores versus the number of MSM microstates plateau near 2000 microstates, so we chose this number of microstates for MSM construction.

Given that typical MSM lag times for protein folding and binding studies range from 1-200 ns, it is somewhat surprising that such a short lagtime (100 ps) is preferred for MSM construction. To validate this result, we additionally built MSMs using lag times of 2, 5, 10, 100 and 1000 (in units of τ_0). We find that the equilibrium populations and slowest relaxation mode eigenvector ϕ_1 are remarkably robust at all lag times, with the exception of 1000, for which finite sampling artifacts become pronounced (Figure S3a). A likely explanation for the short lag time being optimal is the highly diffusive nature of the lid region dynamics, coupled with a trajectory data set enriched in many short simulations (see Figure S1). To test this idea, we built MSMs using lag times of 1, 10, 100 and 1000 using the same construction parameters, but without using a standard ergodic trimming step, which is usually employed to avoid statistical bias from non-equilibrium trajectories.⁴ This bias is particularly pronounced for MSMs built from distributed computing simulations, due to the use of many short trajectories that make forward transitions to new states, but no backward transitions. Enforcing detailed balance on a MSM built from this data can therefore introduce “trap” artifacts in which states can have incorrectly high population estimates. Indeed, without ergodic trimming, the tICA model strongly exhibits the presence of traps, indicating the non-ergodicity of the underlying trajectory data arising from diffusivity and trajectory length (Figure S3b).

Bayes Factor analysis of inter-residue contacts

To quantify the significance of inter-residue contacts formed in specific conformational states, we compute a Bayes Factor (BF) contact metric for each residue pair in MDM2.⁵ The $BF_k(i, j)$ for contacts between residues i and j , given the protein is in some

conformational state k , is computed as

$$BF_k(i, j) = \frac{P(k|c_{ij} = 1)}{P(k|c_{ij} = 0)} = \frac{(P(c_{ij} = 1|k) P(c_{ij} = 0))}{P(c_{ij} = 1) P(c_{ij} = 0|k)}. \quad (3)$$

Here, c_{ij} is an indicator variable that takes the value of 1 if a contact between residues i and j is present, and 0 otherwise. The $BF_k(i, j)$ value can be thought as the statistical over-representation of contact c_{ij} in conformational state k , and hence a measure of its importance in uniquely defining the structural features of that state. For example, if $BF_k = 2$, that means that the equilibrium constant for contact formation between residues i and j is twice as large for state k as it is for the whole ensemble. We compute Bayes factors for contacts separated by three or more residues in sequence, and define a contact formed between two residues if any two non-hydrogen atoms are closer than 4 Å.

Supporting Figures

Figure S1. Trajectory length distributions Figure S2. GMRQ Analysis Figure S3. Effects of lag time and/or ergodic trimming on constructed MSMs Figure S4. Changes in inter-residue contacts and secondary structure along eigenmode relaxations ϕ_1 and ϕ_2 . Figure S5. Bayes Factors of inter-residue contacts for tICA landscape quadrants. Figure S6. DOCK scores projected onto the 2D tICA landscape for all ligands Figure S7. Backbone RMSDs of MDM2 lid residues 11-17 to bound-state p53 helix. Supporting References

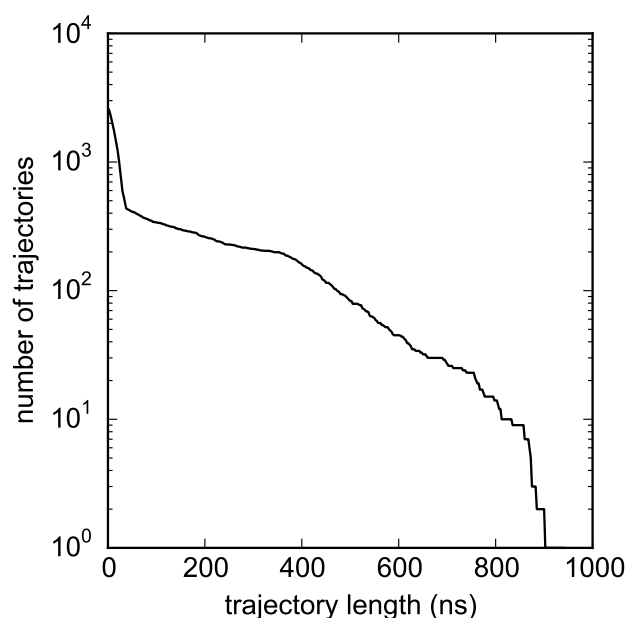


Figure S1. Distribution of trajectory lengths for the 175.7 μ s of aggregate trajectory data simulated on the Foldinghome distributed computing network. The maximum trajectory length is 945 ns, and average trajectory length is 67.0 ns.

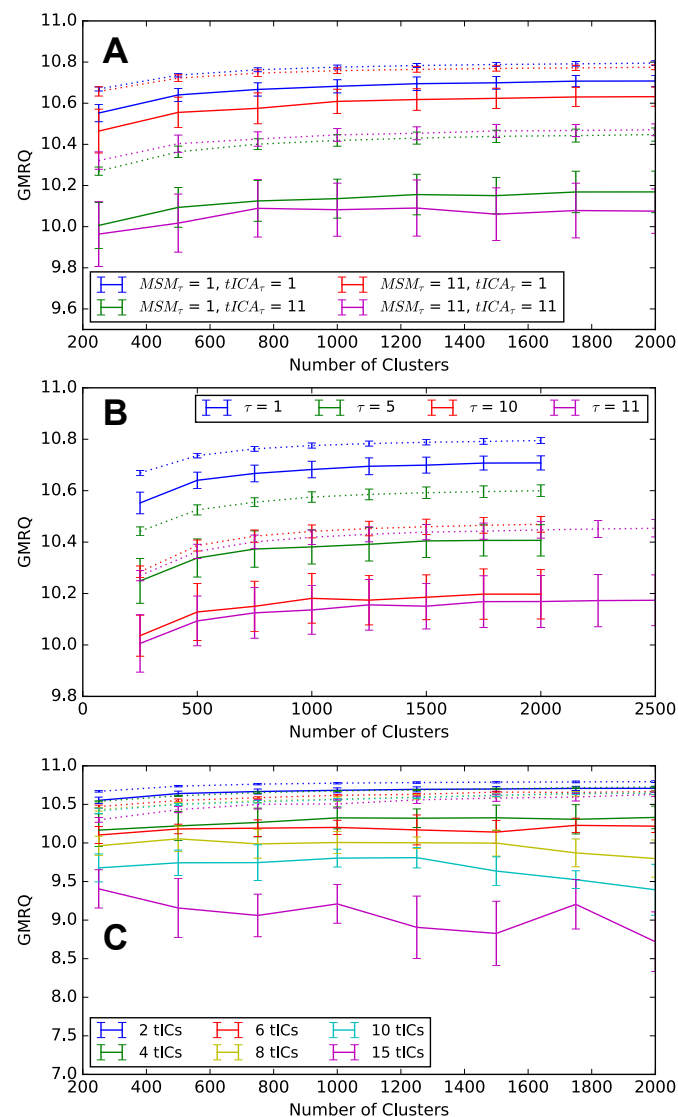


Figure S2. Mean GMRQ training scores (dotted line) and testing scores (solid line) computed using 5-fold cross validation, shown as a function of the number of k -centers clusters used to build the MSM. (A) GMRQ scores calculated using slow ($11\tau_0$) and fast ($1\tau_0$) MSM and tICA lag times. (B) GMRQ scores calculated using a range ($2\tau_0$ to $11\tau_0$) of MSM lag times; each MSM was constructed using clustering in the subspace defined by the 2 largest tICA components. (C) GMRQ scores calculated for MSMs built from various numbers of tICA components (2 to 15). Error bars in all figures show standard deviations of the 5-fold cross-validation trials.

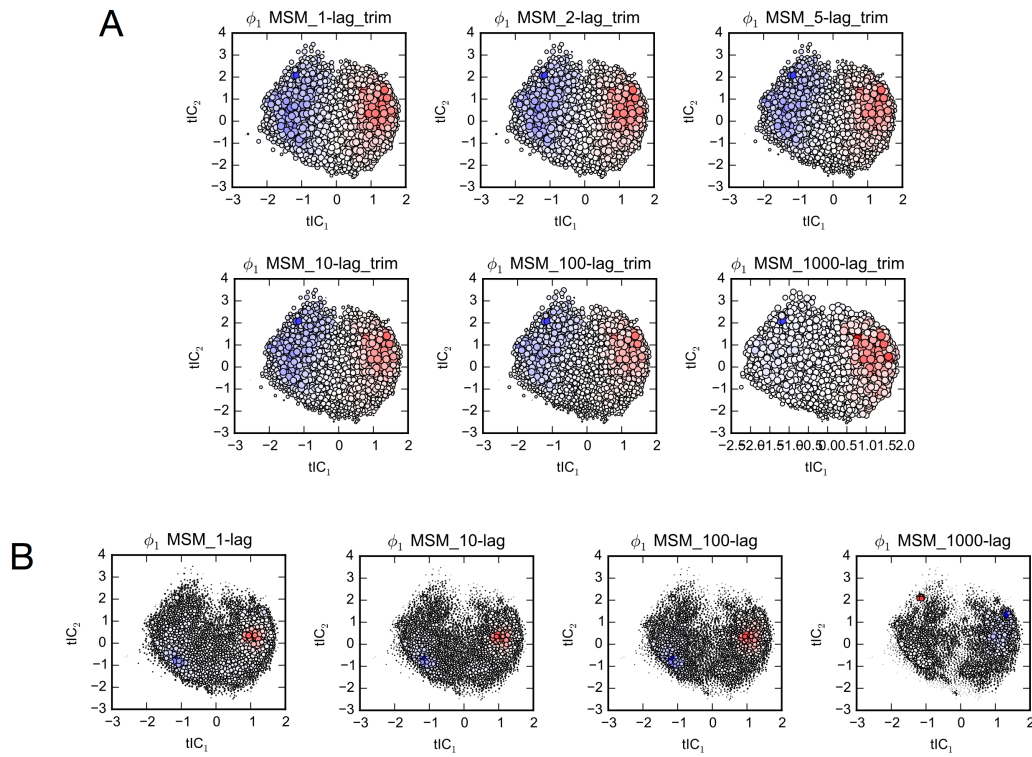


Figure S3. (A) Projections of the 2000 MSM microstates (filled circles) to tIC_1 and tIC_2 coordinates from MSMs built using lag times 1, 2, 5, 10, 100 and 1000 (in units τ_0). The size of each circle is proportional to the logarithm of its equilibrium population, and is colored according to the slowest relaxation eigenmode, ϕ_1 . (B) MSMs built using lag times 1, 10, 100, 1000, without the standard ergodic trimming step.

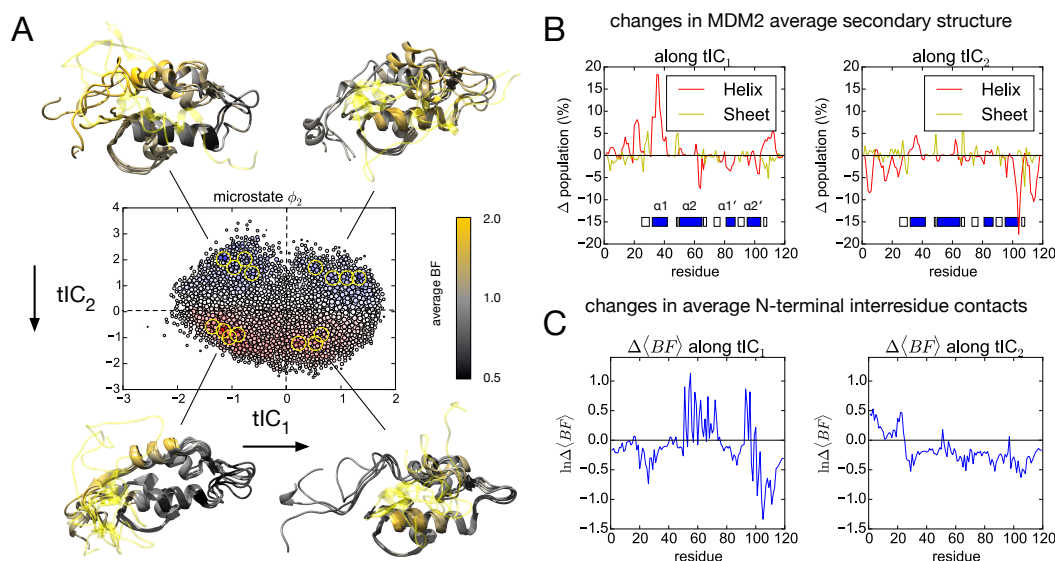


Figure S4. (A) Projection of the 2000 MSM microstates (filled circles) to tIC_1 and tIC_2 coordinates. The size of each circle is proportional to the logarithm of the equilibrium population, and is colored according to the second slowest relaxation eigenmode, ϕ_2 . Population flux along this mode is from blue to red, predominantly representing the structuring of the MDM2 termini, which we visualize using five representative structures from each basin (circled in yellow on the $tICA$ landscape). Average Bayes Factors for contacts between lid region (yellow ribbon) and non-lid residues are shown as a color gradient (black to orange) on the ribbon structure of MDM2. (B) Differences in per-residue helix and sheet content of MDM2 are shown along tIC_1 (i.e. differences between conformations with positive vs. negative values of tIC_1) and along tIC_2 . (C) Changes in the ratio of mean Bayes Factors for contacts between the lid region (residues 1–25 of MDM2) and all other residues, along tIC_1 and tIC_2 .

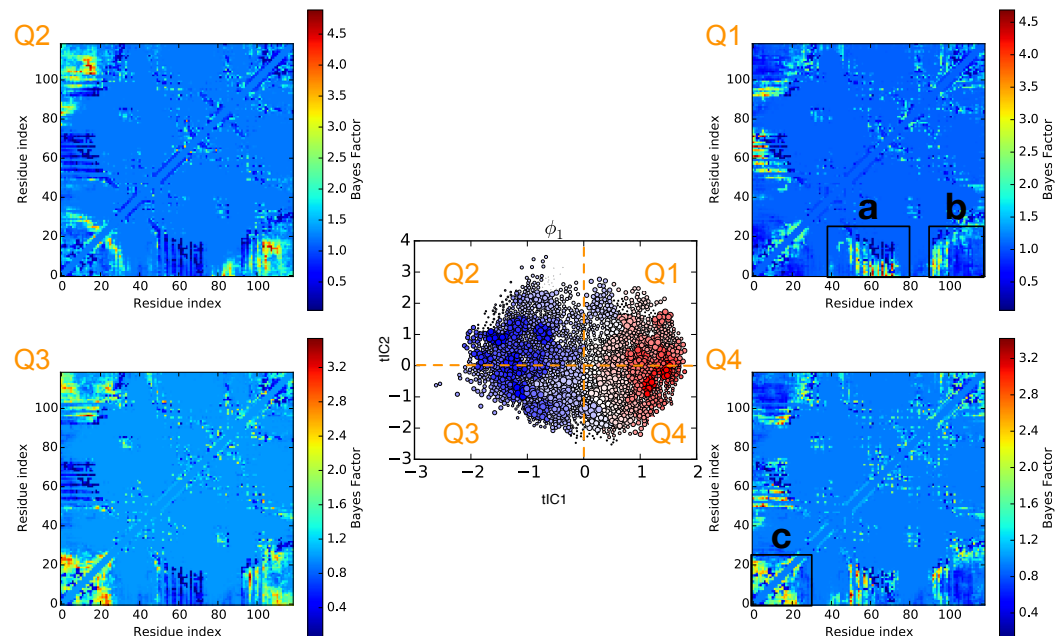


Figure S5. Bayes Factors of inter-residue contacts calculated for each quadrant in (tIC_1, tIC_2) -space. At center is shown the 2000 MSM microstates projected to the $tICA$ landscape, divided into quadrants Q1, Q2, Q3 and Q4. (a) Differences in the Bayes factors for Q2+Q3 versus Q1+Q4 reveal that contacts are formed between the N-terminus and the active site helix α_2 along tIC_1 , the slowest relaxation mode, and (b) contacts are lost between the N- and C-terminus along tIC_1 . Differences in the Bayes factors for Q1+Q2 versus Q3+Q4 reveal that (c) conformational changes along tIC_2 correspond to structuring in the N-terminus.

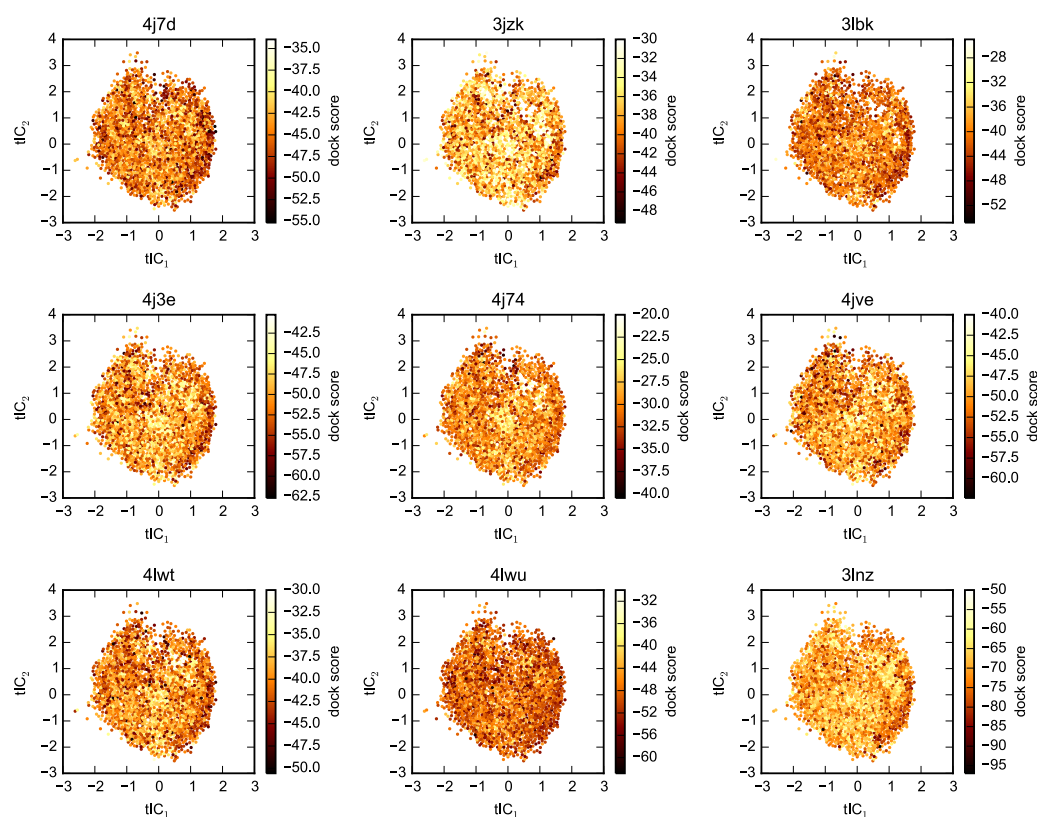


Figure S6. DOCK scores projected onto the 2D tICA landscape for all ligands. Results for p53 (1ycr) are shown in the main text.

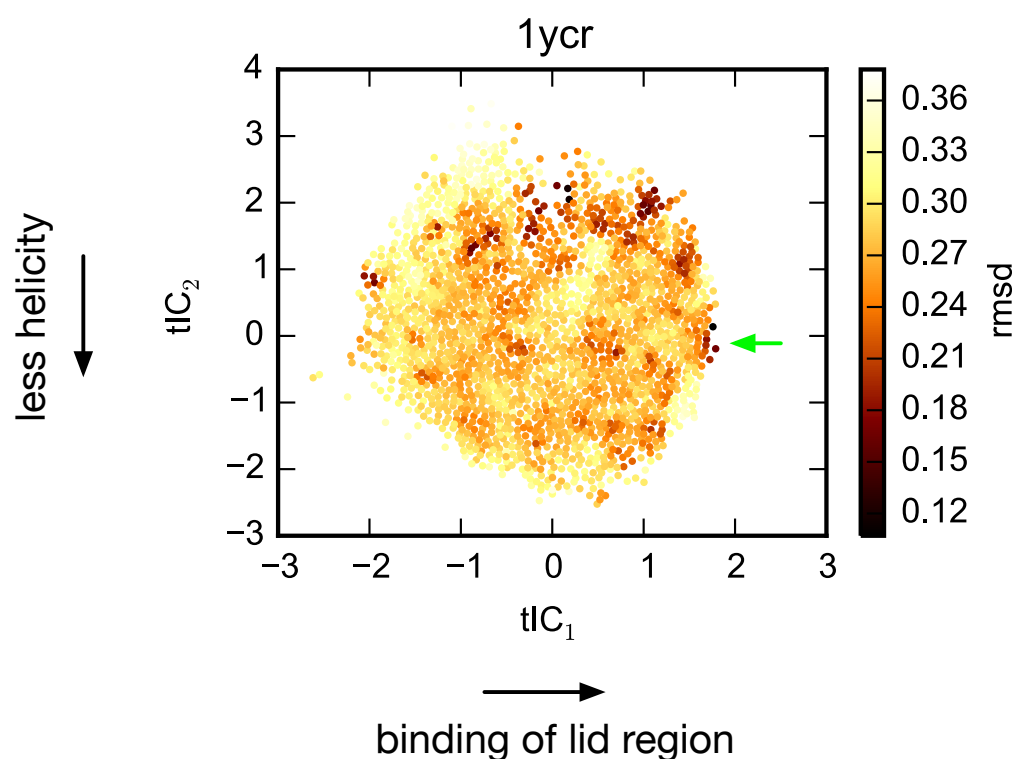


Figure S7. Backbone RMSDs of MDM2 lid residues 11-17 (DGAVTTS) to the bound-state p53 helix, shown for all 2000 MSM microstates on the tICA landscape. In the unbound state, this sequence has low propensity to form a p53-like helix, with two exceptions: (1) There is some residual helicity in the unbound state, which is lost along the ϕ_2 eigenmode relaxation (positive to negative tIC_2 values, see Figure S4) as the N-terminus self-associates, and (2) association of the lid region with the p53 binding cleft induces structuring of residues 11-17 to a conformation very similar to bound p53 (green arrow).