bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Bridging microscopic and macroscopic mechanisms of p53-MDM2 binding using molecular simulations and kinetic network models

Guangfeng Zhou,[1] George A. Pantelopulos,[1] Sudipto Mukherjee,[1] and Vincent A. Voelz[1, a)]
Department of Chemistry, Temple University, Philadelphia, PA 19122, USA

Under normal cellular conditions, the tumor suppressor protein p53 is kept at a low levels in part due to ubiquitination by MDM2, a process initiated by binding of MDM2 to the intrinsically disordered transactivation domain (TAD) of p53. Although many experimental and simulation studies suggest that disordered domains such as p53 TAD bind their targets nonspecifically before folding to a tightly-associated conformation, the molecular details are unclear. Toward a detailed prediction of binding mechanism, pathways and rates, we have performed large-scale unbiased all-atom simulations of p53-MDM2 binding. Markov State Models (MSMs) constructed from the trajectory data predict p53 TAD binding pathways and on-rates in good agreement with experiment. The MSM reveals that two key bound intermediates, each with a non-native arrangement of hydrophobic residues in the MDM2 binding cleft, control the overall on-rate. Using microscopic rate information from the MSM, we parameterize a simple four-state kinetic model to (1) determine that induced-fit pathways dominate the binding flux over a large range of concentrations, and (2) predict how modulation of residual p53 helicity affects binding, in good agreement with experiment. These results suggest new ways in which microscopic models of bound-state ensembles can be used to understand biological function on a macroscopic scale.

AUTHOR SUMMARY

Many cell signaling pathways involve protein-protein interactions in which an intrinsically disordered peptide folds upon binding its target. Determining the molecular mechanisms that control these binding rates is important for understanding how such systems are regulated. In this paper, we show how extensive all-atom simulations combined with kinetic network models provide a detailed mechanistic understanding of how tumor suppressor protein p53 binds to MDM2, an important target of new cancer therapeutics. A simple four-state model parameterized from the simulations shows a binding-then-folding mechanism, and recapitulates experiments in which residual helicity boosts binding. This work goes beyond previous simulations of small-molecule binding, to achieve pathways and binding rates for a large peptide, in good agreement with experiment.

INTRODUCTION

The transcription activator p53 plays a central role in tumor suppression[1]. Cellular levels of p53 are normally kept low by targeted degradation by the E3 ubiquitin ligase MDM2 (mouse double minute 2), whose N-terminal domain binds residues 17-29 of the p53 transactivation domain (TAD) in a deep hydrophobic cleft[2]. The p53 TAD is intrinsically disordered[3], but forms a helix when bound to MDM2. Various types of cellular stresses such as DNA damage leads to disruption of p53-MDM2 binding and an increase in p53 expression, which in turn promotes cellular repair or apoptosis. Thus, the discovery of potent competitive inhibitors that can disrupt the p53-MDM2 binding interaction has been an important strategy for developing new cancer therapeutics[4,5]. The availability of structural information has also made p53-MDM2 a valuable model system for the study of protein-protein interactions and the development of new classes of peptidomimetics[6–9] often alongside computational design efforts[10,11].

A consensus of experimental and simulation studies suggest that intrinsically disordered protein (IDP) domains such as p53 TAD bind their receptors through an induced-fit "fly-casting" mechanism, whereby binding occurs first, followed by structuring into higher-affinity poses[12–17]. It has been proposed that this mechanism facilitates binding to multiple partners in complex regulatory networks, and may enable fast association rates important for signaling. The structural and kinetic properties of IDPs are thought to fine-tune many signaling interactions[18]. Recently, Borcherds et al. have shown that the extent of residual helicity of p53 TAD can modulate p53-MDM2 binding affinity as well as signaling dynamics in cells[19]. An important challenge for molecular simulation is thus to predict binding pathways and association rates of IDPs to their targets, and the detailed molecular mechanisms responsible for shaping them.

In this work, we use extensive all-atom molecular simulations in explicit solvent, combined with state-of-the-art Markov State Model (MSM) approaches, to investigate the p53-MDM2 binding mechanism. Recent MSM studies have examined the mechanisms by which protein receptors recognize small molecules[20–23] and here we extend similar methods to model the coupled folding and binding of a larger peptide (p53 TAD) to MDM2.

a) Electronic mail: voelz@temple.edu

bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

METHODS

Molecular simulation

Simulations starting from unbound states were performed on the the Folding@home distributed computing platform[24] with Gromacs 4.5.4[25] using the ff99sb-ildn-nmr force field[26] and TIP3P explicit solvent. A number of initial starting configurations were selected from conformational clustering of implicit-solvent REMD simulations, each placed at different distances within 12 Å from the binding site. A total 2776 trajectories were generated amounting to $\sim 831\mu$s of aggregate simulation data (Figure S1).

MSM construction

Recent methodological advances have exploited the variational approach to conformational dynamics[27] to enable the construction of optimal MSMs given the available trajectory data[28,29]. To implement this approach, we used time-structure-based Independent Component Analysis (tICA)[30,31] to project the trajectory data to a low-dimensional subspace that best preserves the slowest conformational transitions. Using all pair distances between $C_\alpha + C_\beta$ atoms of p53 and the binding pocket of MDM2 (see Supporting Information for details), we constructed a time-lagged correlation matrix $\mathbf{C}^{(\Delta t)}$ and corresponding covariance matrix from the pair distances using a $\Delta t = 5$ ns lagtime. The tICA components α are found by maximizing the objective function $\langle \alpha_i | \mathbf{C}^{(\Delta t)} | \alpha_i \rangle$ subjected to certain constraints[30]. Once projected to the tICA subspace, distance-based clustering using the k-means algorithm was performed to obtain MSM metastable state definitions. To select hyper-parameters such as the number of tICA components, clustering method, MSM lag time, and the number of MSM microstates, we performed variational cross-validation using the GMRQ method of McGibbon et al.[28] on over 120 MSMs. As in previous work[32], we find that tICA distance metrics are better than rmsd or dihedral angle metrics, and k-means clustering performs better than k-centers. The optimal MSM, used in all subsequent analysis, is constructed using 10 tICA components, 600 microstates and a 5 ns MSM lag time τ_{lag} (Figure S2). The MSM transition matrix $\mathbf{T}(\tau_{\text{lag}})$ was estimated using a maximum likelihood method[33,34]. The model is validated by implied timescales which plateau near the chosen lag time of 5 ns (Figure S3), and Chapman-Kolmogorov tests (Figure S4). All models were built using MSMbuilder 3.3[33] and MDTraj 1.5[35] software packages.

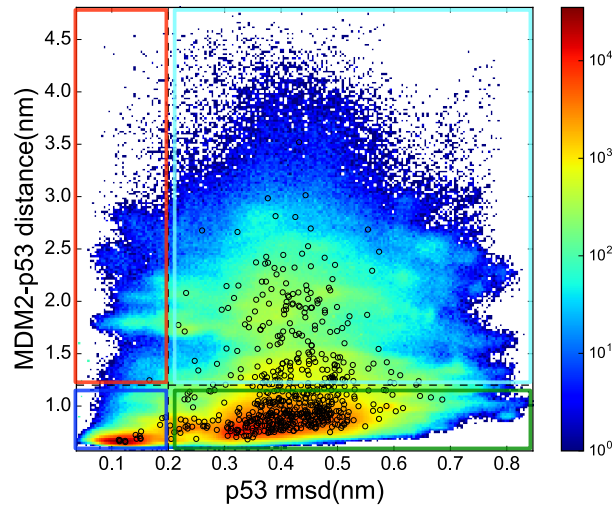

FIG. 1. Simulation trajectory data projected to p53-MDM2 distance and the p53 rmsd-to-native show a preference for binding before folding. Colored boxes highlight the partitioning into four states based on these reaction coordinates: folded-bound (blue), unfolded-bound (green), folded-unbound (red), and unfolded-unbound (cyan). Black circles show the locations of cluster centers of the 600 MSM microstates.

RESULTS

Binding precedes folding. A projection of the simulation data to two reaction coordinates–the rmsd of p53 to its native structure, and the distance of p53 to the MDM2 binding pocket–suggests that binding of p53 precedes folding of p53, consistent with the "fly-casting" mechanism (Figure 1). The distance vs. rmsd landscape can be manually partitioned into four states: folded-bound (blue), unfolded-bound (green), folded-unbound (red), and unfolded-unbound (cyan). These states were defined using a bound-state distance cutoff of 1.2 nm, and rmsd cutoff of 0.2 nm. Projecting the 600 MSM microstates to this landscape, we find most of the population in the bound states, with only one microstate corresponding to the folded-unbound state (red).

Projections of the simulation data to the two largest tICA components, corresponding to the slowest conformational dynamics, show a very different landscape (Figure 2). The folded-bound state (blue) is composed of a single well-populated basin, closely matching within 1.3 Å backbone rmsd to the native co-crystal structure, with the side chains of F19, W23 and L26 correctly inserted into the binding pocket of MDM2. In contrast, the unfolded-bound state (green) is distributed throughout the tICA landscape. The two predominant basins of the unfolded-bound state correspond to p53 bound in two different misfolded states, each with residue F19 in its native binding groove, but W23 outside of the binding cleft. As can be seen by the eigenvector structure of the MSM (Figure S5) transitions from these basins control

bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

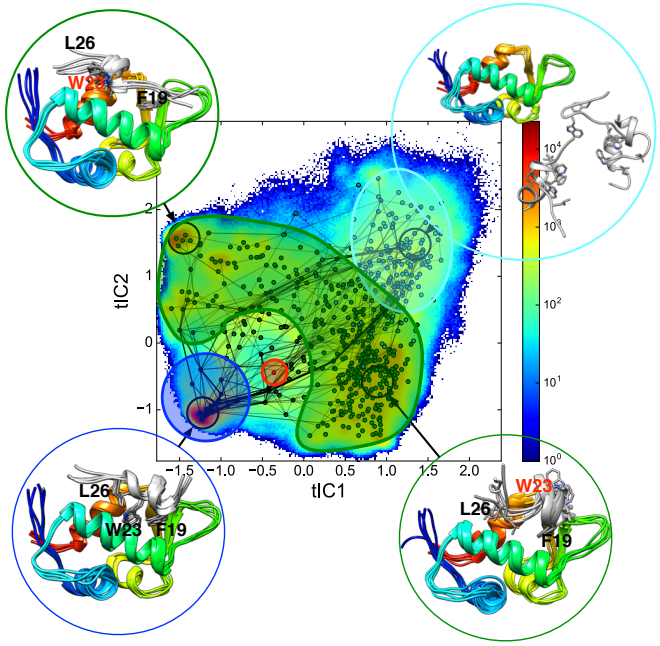

FIG. 2. Trajectory data projected to the 2D tICA landscape, with microstates color-coded and overlaid with color-shaded cartoons to highlight the locations of the four states. The 100 highest-flux binding pathways calculated from TPT (black lines) fall into two groups, each dominated by a different unfolded-bound basin with W23 misregistered outside of the binding cleft. Representative structures for these two basins, the folded-bound native state, and the unfolded-unbound state are shown in circles.

the slowest timescales of binding.

Transition pathways and rates. To estimate pathways, fluxes and rates of p53 association, we used Transition Path Theory (TPT), which we briefly describe here and refer readers to other references for more details[36–38]. In TPT, source states (A) and sink states (B) first need to be defined for the transition process of interest. The remaining states are considered to be intermediate states (I). Next, committor probabilities q_i^+, defined as the probability that a trajectory started from state i will reach B before state A, are computed from the MSM transition matrix. The total folding flux giving the expected number of observed $A \to B$ transitions per time unit τ is: $F = \sum_{i \in A} \sum_{j \notin A} \pi_i T_{ij} q_j^+$. The rate of reaction $A \to B$, k_{AB} can then be computed as:

$$k_{AB} = \frac{F}{\tau \sum_{i=1}^{m} \pi_i q_i^-} \quad (1)$$

We first used TPT to estimate overall p53 binding on- and off-rates (k_{on} and k_{off}) from the 600-microstate MSM, using the unfolded-unbound and folded-bound states as the source and sink states for k_{on}, respectively (and vice versa for estimating k_{off}). For comparison, we constructed a four-macrostate MSM by manually lumping the 600 microstates according to our four-state definitions. The results (Table I) show that the

TABLE I. Comparison of experimental and simulated rates.

Methods	k_{on} ($\text{M}^{-1}\ \text{s}^{-1}$)	k_{off} (s^{-1})
Experiment[39]	9.2×10^6	2.0
TPT 600-state MSM	2.5×10^7	1.9×10^5
TPT 4-state MSM	5.2×10^7	6.0×10^5

k_{on} estimated from the 600-microstate MSM is very close to the experimental k_{on} (within a factor of 2.7). Macrostate lumping into four states further accelerates the dynamic timescales, with k_{on} predictions still within a factor of 5.6. As expected given the available trajectory lengths, k_{off} estimates from both models are severely over-estimated.

A four-state kinetic model predicts an induced-fit mechanism for p53 binding. To analyze whether the binding mechanism follows a conformational selection (CS) or induced fit (IF) mechanism (or aspects of both), we computed the reactive flux for each mechanism according to the method introduced by Hammes et al.[40], illustrated in Figure 3. In this model, association of a ligand can occur either through a weak-binding (w) form of p53, or a tight-binding (t) form, with interconversions between these two possible when unbound or bound. Here, we slightly modify our interpretation of the model for use with disordered peptide binding; in our case it is the *receptor* MDM2 which can select or induce folded (helical) states of p53 TAD. Following Hammes et al., we compute the reactive flux for conformational selection pathways as:

$$F_{\text{CS}} = \left(\frac{1}{k_{wt}[\text{p53}_{\text{weak}}]} + \frac{1}{k_{\text{on}}^{t}[\text{p53}_{\text{tight}}][\text{MDM2}]_f} \right)^{-1} \quad (2)$$

and the flux for induced fit pathways as

$$F_{\text{IF}} = \left(\frac{1}{k_{\text{on}}^{w}[\text{p53}_{\text{weak}}][\text{MDM2}]_f} + \frac{1}{k_{wt}^{\text{MDM2}}[\text{MDM2} \cdot \text{p53}_{\text{weak}}]} \right)^{-1} \quad (3)$$

where $[\text{MDM2}]_f$ is the free MDM2 concentration. The derivation is shown in the Supporting Information.

To obtain the relative amounts of reactive flux that occur by conformational selection vs. induced fit pathways, we use our MSM model to make initial estimates of all eight rates in the four-state kinetic model shown in Figure 3. We tried two different approaches to make these initial estimates: (1) directly from the transition probabilities of a four-macrostate MSM derived for our state classifications, and (2) using TPT with pairs of relevant states selected as the source and sink states. Both sets of estimates are listed in Table S1. We find that the two methods yield very similar results, except that k_{on}^{t} estimates from the transition matrix are more than an order of magnitude larger than the estimate from TPT, due to enforcing detailed balance with a low equilibrium population predicted for the folded-unbound state. Therefore, in the following analysis we use only the parameters estimated from TPT. From this initial estimate, we then

bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

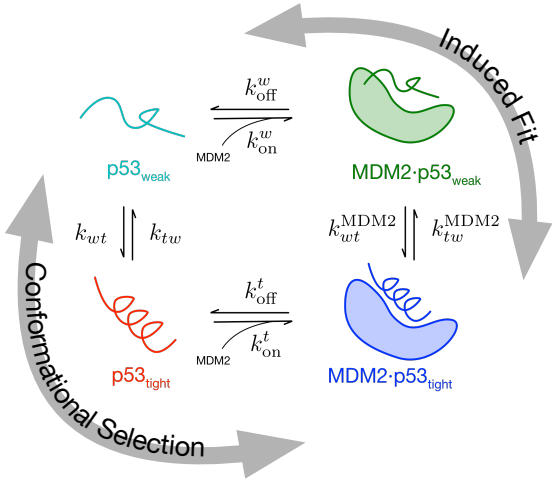

FIG. 3. A four-state model of possible p53-MDM2 binding mechanisms includes both induced-fit (IF) and conformational selection (CS) pathways.

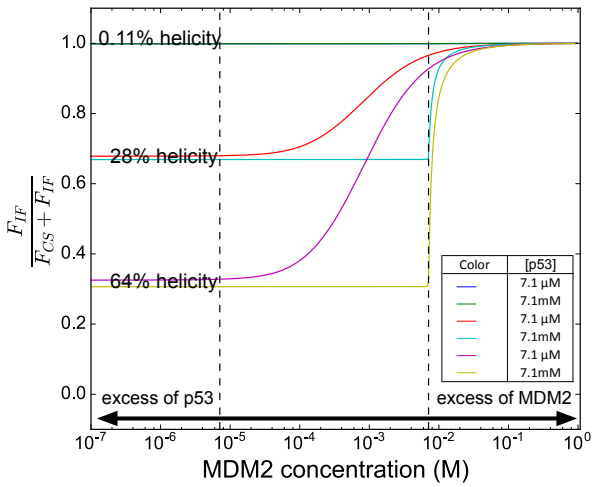

FIG. 4. Fraction of binding flux going through induced-fit (IF) pathways at different p53 concentrations and helicities versus total MDM2 concentration. Dashed lines denote the p53 concentrations shown. 7.1 mM is the concentration of p53 and MDM2 in the molecular simulations.

scale the off-rates $k^w_{\rm off}$ and $k^t_{\rm off}$ to reproduce the experimental binding affinity of p53 to MDM2 (see Supporting Information).

Unlike our MSM, which was constructed from simulations performed at a fixed concentration ([p53] = [MDM2] = 7.1 mM), the resulting four-state kinetic model can be used to extrapolate the binding fluxes at any desired concentrations. In all cases, we find that binding is dominated by an induced-fit mechanism, consistent with "fly-casting". The fraction of flux that occurs by an induced-fit mechanism, $F_{\rm IF}/(F_{\rm CS}+F_{\rm IF})$ is nearly 100% regardless of the concentrations of p53 and MDM2. This is mainly due to the fact that the simulated helicity of p53 is very low (0.11%, Figure 4).

Increased residual helicity leads to enhanced p53 binding and a shift toward conformational selection. To estimate the effect of residual helicity of p53 on binding mechanism and affinity, we use a maximum caliber approach to infer how the rates k_{wt} and k_{tw} between unfolded-unbound states (cyan) and folded-unbound states (red) change in response to new helix-coil equilibrium populations (see Supporting Information), using the relation $k_{ab} \propto \sqrt{\pi_b/\pi_a}$[41]. The helicity of p53 predicted by our 4-state model is 0.11%. To model the experimental system of Borcherds et al., we increase the helix population of unbound p53 to 28% and 64%, values measured for wild-type and P27A variants of the p53 TAD[19]. The inferred rates are shown in Table S3.

The four-state kinetic model predicts that increasing residual helicity increases the flux of conformational selection at low MDM2 concentration; however, in the limit of excess of MDM2, induced-fit binding flux increases to almost 100% (Figure 4). The reason for this is the relatively high k^w_{on} value, a key feature of intrinsically disordered proteins that we have calculated directly from the MSM model. In the limit of excess of MDM2, k^w_{on} would have to be reduced by several orders of magnitude to convert the binding mechanism to conformational selection. In the limit of excess p53, a shift towards a conformational selection binding mechanism is observed, although a strong preference for induced-fit binding pathways (more than 30% of the binding flux) remains even at high levels of residual helicity (64%) and in the excess of p53.

In agreement with experiment, the four-state model predicts a greater apparent binding affinity of P27A p53 TAD compared to wild-type, with absolute and relative binding free energies similar to experimental values (see Figure 5). The predicted apparent ΔG of binding for p53 wild-type and P27A are -7.5 and -9.0 kcal·mol^{-1}, respectively, while the experimental values are -9.1 ± 0.2 and -10.4 ±0.1 kcal·mol^{-1}, respectively. We predict that the $\Delta\Delta G$ incurred by increasing the helicity of p53 from 28% to 64% is -1.5 kcal·mol^{-1}, which also agrees very well with experiment (-1.3±0.3 kcal·mol^{-1}).

CONCLUSION

Recently, Zwier et al. has reported efficient implicit-solvent simulations of p53 TAD binding to MDM2 using GPU-enabled weighted-ensemble path sampling methods which predict similarly accurate on-rates and a binding-then-folding mechanism[42,43]. By comparison, the work we present here achieves nearly an order of magnitude more sampling in full explicit solvent, and uses an MSM-based framework for analysis which enables (1) the estimation of rates directly from Transition Path Theory, and (2) predictive connections to a more macroscopic

bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

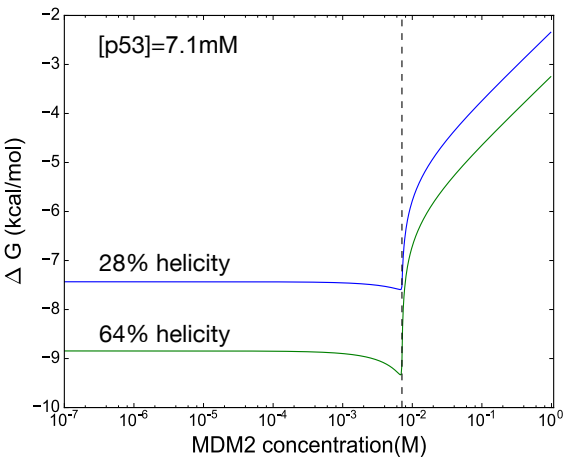

FIG. 5. Predicted p53-MDM2 binding affinities versus MDM2 concentration from the four-state kinetic model. Increasing the residual helicity of p53 TAD from 28% to 64% increases affinity by $\Delta\Delta G = -1.5$ kcal/mol, in good agreement with experiment (-1.3 ± 0.3 kcal/mol).

model of conformational selection vs. induced fit. Both studies suggest a bright future using adaptive sampling simulations to model protein-peptide binding. With a high-quality MSM of p53 binding now constructed, we aim to exploit new MSM-based adaptive sampling approaches to model binding rates and mechanisms for multiple sequences[44]. A remaining challenge of course is to efficiently sample off-rates as well as on-rates. With the advent of multiple-ensemble MSM techniques[45], this too may be within reach in the near future. We also expect that the mechanistic detail provided by MSM approaches may suggest new ways to design inhibitors that compete with natural substrates, by rationalizing how specific binding modes may modulate key rate processes.

In conclusion, we have used *ab initio* binding simulations and Markov State Models to construct a detailed kinetic network model of p53 TAD binding to MDM2. The MSM predicts binding on-rates in agreement with experiment, as well an ensemble of encounter complex structures that control the overall binding pathways and rates. Predicted MSM rates, along with experimental affinities, were used to parameterize a four-state kinetic model, which predicts an induced-fit "fly-casting" mechanism over a wide range of concentrations, and shows increased binding affinity for p53 variants with higher amounts of residual structure, in agreement with recent experiments. This work demonstrates how combining detailed all-atom MSMs and simple few-state kinetic models can be very useful in understanding how disordered protein domains bind their target receptors. The results also suggest new ways to design inhibitors that compete with natural substrates, by rationalizing how specific binding modes may modulate key rate processes.

ACKNOWLEDGMENTS

The authors thank the participants of Folding@home, without whom this work would not be possible. This research was supported in part by the National Science Foundation through major research instrumentation grant number CNS-09-58854 and MCB-1412508.

[1] A. C. Joerger and A. R. Fersht, Annual Review of Biochemistry **85**, 375 (2016).
[2] P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, and N. P. Pavletich, Science **274**, 948 (1996).
[3] M. Wells, H. Tidow, T. J. Rutherford, P. Markwick, M. R. Jensen, E. Mylonas, D. I. Svergun, M. Blackledge, and A. R. Fersht, Proceedings of the National Academy of Sciences of the United States of America **105**, 5762 (2008).
[4] K. K. Hoe, C. S. Verma, and D. P. Lane, Nature Reviews Drug Discovery **13**, 217 (2014).
[5] P. Chène, Nature Reviews Cancer **3**, 102 (2003).
[6] J. A. Kritzer, J. D. Lear, M. E. Hodsdon, and A. Schepartz, Journal of the American Chemical Society **126**, 9468 (2004).
[7] Z. Z. Brown, K. Akula, A. Arzumanyan, J. Alleva, M. Jackson, E. Bichenkov, J. B. Sheffield, M. A. Feitelson, and C. E. Schafmeister, PLoS ONE **7**, e45948 (2012).
[8] S. Baek, P. S. Kutchukian, G. L. Verdine, R. Huber, T. A. Holak, K. W. Lee, and G. M. Popowicz, Journal of the American Chemical Society **134**, 103 (2012).
[9] N. Estrada-Ortiz, C. G. Neochoritis, and A. Dömling, ChemMedChem **11**, 757 (2015).
[10] B. B. Lao, K. Drew, D. A. Guarracino, T. F. Brewer, D. W. Heindel, R. Bonneau, and P. S. Arora, Journal of the American Chemical Society **136**, 7877 (2014).
[11] J. Michel, E. A. Harker, J. Tirado-Rives, W. L. Jorgensen, and A. Schepartz, Journal of the American Chemical Society **131**, 6356 (2009).
[12] K. Sugase, H. J. Dyson, and P. E. Wright, Nature **447**, 1021 (2007).
[13] L. N. Wafer, W. W. Streicher, S. A. McCallum, and G. I. Makhatadze, Biochemistry **51**, 7189 (2012).
[14] P. E. Wright and H. J. Dyson, Nature Reviews Molecular Cell Biology **16**, 18 (2015).
[15] Y. Huang and Z. Liu, Journal of molecular biology **393**, 1143 (2009).
[16] J. Chen, Journal of the American Chemical Society **131**, 2088 (2009).
[17] D. De Sancho and R. B. Best, Molecular BioSystems **8**, 256 (2012).
[18] V. Iesmantavicius, J. Dogan, P. Jemth, K. Teilum, and M. Kjaergaard, Angewandte Chemie International Edition **53**, 1548 (2014).
[19] W. Borcherds, F.-X. Theillet, A. Katzer, A. Finzel, K. M. Mishall, A. T. Powell, H. Wu, W. Manieri, C. Dieterich, P. Selenko, A. Loewer, and G. W. Daughdrill, Nature Chemical Biology **10**, 1000 (2014).
[20] S. Gu, D.-A. Silva, L. Meng, A. Yue, and X. Huang, PLoS computational biology **10**, e1003767 (2014).
[21] X. Huang and G. De Fabritiis, Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation **797**, 107 (2014).
[22] K. J. Kohlhoff, G. R. Bowman, D. Shukla, M. Lawrenz, D. E. Konerding, D. Belov, R. B. Altman, and V. S. Pande, Nature Chemistry **6**, 15 (2014).
[23] N. Plattner and F. Noé, Nature Communications **6**, 7653 (2015).
[24] M. Shirts and V. S. Pande, Science **290**, 1903 (2000).
[25] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, Bioinformatics **29**, 845 (2013).
[26] D.-W. Li and R. Brüschweiler, Angewandte Chemie (International ed. in English) **49**, 6778 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[27]F. Noé and F. Nüske, Multiscale Modeling & Simulation **11**, 635 (2013).
[28]R. T. McGibbon and V. S. Pande, The Journal of Chemical Physics **142**, 124105 (2015).
[29]L. Boninsegna, G. Gobbo, F. Noé, and C. Clementi, Journal of Chemical Theory and Computation **11**, 5947 (2015).
[30]C. R. Schwantes and V. S. Pande, Journal of Chemical Theory and Computation **9**, 2000 (2013).
[31]G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, The Journal of chemical physics **139**, 015102 (2013).
[32]G. Zhou and V. a. Voelz, The Journal of Physical Chemistry B **120**, 926 (2016).
[33]K. A. Beauchamp, G. R. Bowman, T. J. Lane, L. Maibaum, I. S. Haque, and V. S. Pande, Journal of Chemical Theory and Computation **7**, 3412 (2011).
[34]B. Trendelkamp-Schroer and F. Noé, The Journal of Chemical Physics **138**, 164113 (2013).
[35]R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, and V. S. Pande, Biophysical Journal **109**, 1528 (2015).
[36]E. Vanden-Eijnden, Journal of Statistical Physics **123**, 503 (2006).
[37]C. Schütte, P. Metzner, and E. Vanden-Eijnden, Multiscale Modeling & Simulation **7**, 1192 (2009).
[38]F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, Proceedings of the National Academy of Sciences **106**, 19011 (2009).
[39]O. Schon, A. Friedler, M. Bycroft, S. M. V. Freund, and A. R. Fersht, Journal of Molecular Biology **323**, 491 (2002).
[40]G. G. Hammes, Y.-C. Chang, and T. G. Oas, Proceedings of the National Academy of Sciences of the United States of America **106**, 13737 (2009).
[41]P. D. Dixit, A. Jain, G. Stock, and K. A. Dill, Journal of Chemical Theory and Computation **11**, 5464 (2015).
[42]M. C. Zwier, J. L. Adelman, J. W. Kaus, A. J. Pratt, K. F. Wong, N. B. Rego, E. Suárez, S. Lettieri, D. W. Wang, M. Grabe, D. M. Zuckerman, and L. T. Chong, Journal of Chemical Theory and Computation **11**, 800 (2015).
[43]M. C. Zwier, A. J. Pratt, J. L. Adelman, J. W. Kaus, D. M. Zuckerman, and L. T. Chong, The Journal of Physical Chemistry Letters , 3440 (2016).
[44]V. a. Voelz, B. Elman, A. M. Razavi, and G. Zhou, Journal of Chemical Theory and Computation **10**, 5716 (2014).
[45]H. Wu, F. Paul, C. Wehmeyer, and F. Noé, Proceedings of the National Academy of Sciences of the United States of America , 201525092 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SUPPORTING INFORMATION

Supporting Text

Preparation of initial p53-MDM2 configurations for simulation. Four non-native structures of the p53 fragment were chosen from 500 ns of replica exchange molecular dynamics (REMD) simulations of the p53 fragment (residues 17-26) taken from PDB structure 1YCR. The Amber ff96 force field with the OBC GBSA implicit solvent model were used to simulate 16 exponentially-spaced temperature replicas between 300 and 450 K. Conformational clustering was performed on the lowest-temperarture replica using a backbone-RMSD distance metric, from which representative structures of p53 were chosen. Of these structures, one was partial helical, two were hairpin-like and one was random coil. These four p53 TAD structures, in addition to the native structure, were used to prepare 30 initial structures of the p53-MDM2 complex at a variety of distances between the p53 fragment and the MDM2 binding pocket by placing six replicas of each p53 fragment at 2 Å intervals (2, 4, 6, 8, 10 and 12 Å) along a vector from the MDM2 center of mass to the p53 center of mass the 1YCR crystal structure.

Pairwise atom distances used for time-lagged independent component analysis (tICA). To perform tICA, the simulation trajectory data was first projected to a set of 1953 distance coordinates, from which a time-lagged correlation matrix was constructed. We used all pairwise distances between C_α and C_β atoms in selected residues of p53 (Glu17, Thr18, Phe19, Ser20, Asp21, Leu22, Trp23, Lys24, Leu25, Leu26, Pro27, Glu28, Asn29) and MDM2 (Glu25, Thr26, Met50, Lys51, Leu54, Leu57, Gly58, Ile61, Met62, Tyr67, Gln71, Gln72, His73, Val75, Phe91, Val93, Lys94, His96, Ile99).

Distance between p53 and the MDM2 binding pocket. Figure 1 in the main text uses as a structural observable the distance of p53 to the MDM2 binding pocket. We define this distance as the average of six p53-MDM2 pairwise C_α distances: (p53-Phe19, MDM2-Met62), (p53-Phe19, MDM2-Gln72), (p53-Trp23, MDM2-Gly58), (p53-Trp23, MDM2-Val93), (p53-Leu26, MDM2-Leu54), and (p53-Leu26, MDM2-His96). Selected residues are highlighted in red in Figure S6.

Correction of off-rates to accurately reproduce experiment. While TPT analysis of our Markov State Model predicts binding on-rates accurately (see Table 1, main text), off-rates are over-estimated by about five orders of magnitude. This is not surprising given the timescales sampling in our simulations. To obtain realistic off-rates for the four-state binding mechanism model, we correct the values of k^w_{off} and k^t_{off} (see Figure 2, main text) according to the experimental off-rate through a scaling factor, γ.

$$\gamma = \frac{k^{exp}_{\text{off}}}{k^{sim}_{\text{off}}} = \frac{2.0\ \text{s}^{-1}}{6.0 \times 10^5\ \text{s}^{-1}} \approx 3.3 \times 10^{-6} \tag{4}$$

$$k^w_{\text{off}}(\text{corrected}) = \gamma k^w_{\text{off}} \tag{5}$$

$$k^t_{\text{off}}(\text{corrected}) = \gamma k^t_{\text{off}} \tag{6}$$

Here, k^{sim}_{off} is the off-rate estimated using TPT from the 4-state MSM (see Table 1, main text), and k^{exp}_{off} is the experimentally measured off-rate. The original estimates used for correction are from TPT estimation (600-state MSM) as shown in Table S1 and the corrected values are shown in Table S2.

Correction of p53 folding rates to model residual helicity. The *helicity* of p53 is defined to be the fraction of population in the helical state:

$$h_{p53} = \frac{\pi_{helix}}{\pi_{helix} + \pi_{coil}} = \frac{\pi_{p53_{tight}}}{\pi_{p53_{tight}} + \pi_{p53_{weak}}} \tag{7}$$

where $\pi_{p53_{tight}}$ and $\pi_{p53_{weak}}$ are estimated equilibrium populations. The helicity estimated from the MSM model is $h_{p53} = 0.11\%$.

To model other values of residual helicity in the context of the four-state kinetic model, we must adjust the folding and unfolding rates of unbound p53 TAD, i.e. k_{wt} and k_{tw}, respectively. This is an under-constrained problem, as many ratios of rates can lead to the same helix-coil equilibrium constant. Therefore, we turn to a maximum-caliber (MaxCal) approach to infer these rates from the helix/coil population changes, using the relation $k_{ab} \propto \sqrt{\frac{\pi_b}{\pi_a}}$ derived in Dixit et al. Given a desired helicity h'_{p53}, the inferred rates are

$$k'_{wt} = k_{wt}\left(\frac{\pi'_t \pi_w}{\pi'_w \pi_t}\right)^{0.5} = k_{wt}\left[\frac{h'_{p53}(1-h_{p53})}{h_{p53}(1-h'_{p53})}\right]^{0.5} \tag{8}$$

$$k'_{tw} = k_{tw}\left(\frac{\pi'_w \pi_t}{\pi'_t \pi_w}\right)^{0.5} = k_{tw}\left[\frac{h_{p53}(1-h'_{p53})}{h'_{p53}(1-h_{p53})}\right]^{0.5} \tag{9}$$

The inferred values are shown in Table S3.

Derivation of key equations for flux analysis. The four-state model of Hammes et al. gives expressions for the reactive flux of conformational selection (CS) and induced-fit (IF) pathways in terms of rates and concentrations $[\text{p53}_{\text{weak}}]$, $[\text{p53}_{\text{tight}}]$, $[\text{MDM2} \cdot \text{p53}_{\text{weak}}]$ and $[\text{MDM2}]_f$, the *free* (unbound) concentration of MDM2. These concentrations can be computed entirely from the rate parameters and the total concentrations $[\text{MDM2}]_{\text{tot}}$ and $[\text{p53}]_{\text{tot}}$, as described in Hammes et al. In their original publication, we noticed a typographical error in the expression for $[\text{MDM2}]_f$, and present the correct expression here:

bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$
\begin{aligned}
{}_f &= [\mathrm{MDM2}]_{\mathrm{tot}} - [\mathrm{p53}_{\mathrm{weak}} \cdot \mathrm{MDM2}] - [\mathrm{p53}_{\mathrm{tight}} \cdot \mathrm{MDM2}] \\
&= [\mathrm{MDM2}]_{\mathrm{tot}} - [\mathrm{p53}_{\mathrm{weak}}][\mathrm{MDM2}]_f \left(\frac{1}{K_d^{\mathrm{W}}} + \frac{K_{eq}^{\mathrm{WT}}}{K_d^{\mathrm{T}}} \right)
\end{aligned} \tag{10}
$$

Here, K_d^{W} and K_d^{T} are the dissociation constants of $[\mathrm{p53}_{\mathrm{weak}} \cdot \mathrm{MDM2}]$ and $[\mathrm{p53}_{\mathrm{tight}} \cdot \mathrm{MDM2}]$, respectively; K_{eq}^{WT} is the equilibrium constant for the reaction $\mathrm{p53}_{\mathrm{weak}} \rightleftharpoons \mathrm{p53}_{\mathrm{tight}}$.

The concentration of $\mathrm{p53}_{\mathrm{weak}}$ is calculated using the following equation:

$$
[\mathrm{p53}_{\mathrm{weak}}] = \frac{[\mathrm{p53}]_{\mathrm{tot}}}{1 + \frac{[\mathrm{MDM2}]_f}{K_d^{\mathrm{W}}} + K_{eq}^{\mathrm{WT}} \left(1 + \frac{[\mathrm{MDM2}]_f}{K_d^{\mathrm{T}}} \right)} \tag{11}
$$

Substituting Eq. 11 into Eq. 10 yields a quadratic equation for $[\mathrm{MDM2}]_f$:

$$
[\mathrm{MDM2}]_f^2 + \big([\mathrm{p53}]_{\mathrm{tot}} - [\mathrm{MDM2}]_{\mathrm{tot}} + a\big)[\mathrm{MDM2}]_f - a[\mathrm{MDM2}]_{\mathrm{tot}} = 0 \tag{12}
$$

where

$$
a = \frac{1 + K_{eq}^{\mathrm{WT}}}{\frac{1}{K_d^{\mathrm{W}}} + \frac{K_{eq}^{\mathrm{WT}}}{K_d^{\mathrm{T}}}} \tag{13}
$$

The positive solution of Eq. 12 is the concentration of $[\mathrm{MDM2}]_f$. Once this value is obtained, the concentration of other species can be calculated using the following equations:

$$
\begin{aligned}
[\mathrm{p53}_{\mathrm{tight}}] &= K_{eq}^{\mathrm{WT}}[\mathrm{p53}_{\mathrm{weak}}] \\
[\mathrm{p53}_{\mathrm{weak}} \cdot \mathrm{MDM2}] &= \frac{[\mathrm{p53}_{\mathrm{weak}}]}{K_d^{\mathrm{W}}}[\mathrm{MDM2}]_f \\
[\mathrm{p53}_{\mathrm{tight}} \cdot \mathrm{MDM2}] &= \frac{K_{eq}^{\mathrm{WT}}[\mathrm{p53}_{\mathrm{weak}}]}{K_d^{\mathrm{T}}}[\mathrm{MDM2}]_f
\end{aligned} \tag{14}
$$

We note that the Hammes et al. model was derived in the context of protein-ligand association, and applied to systems in which the receptor is thought to undergo conformational change that can be either selected or induced by the association of a ligand. In our model, it is p53 that is analogous to the protein receptor, as it undergoes conformational change between a tight-binding helical state and weak-binding coil state, and MDM2 is analogous to the ligand, as it can select or induce helical states of p53 by association.

Estimation of p53-MDM2 binding affinities at different values of residual helicities. We calculated the binding affinity ΔG at different values of p53 helicities according to $\Delta G = RT \ln K_d$. We estimate the disassociation constant K_d based on the four-state model, considering $\mathrm{p53}_{\mathrm{tight}} \cdot \mathrm{MDM2}$ as the bound state, and the other three states to be unbound states. Thus, the equation used to estimate K_d is

$$
\begin{aligned}
K_d &= \frac{[\mathrm{p53}][\mathrm{MDM2}]}{[\mathrm{p53} \cdot \mathrm{MDM2}]} \\
K_d &= \frac{([\mathrm{p53}_{\mathrm{weak}}] + [\mathrm{p53}_{\mathrm{tight}}] + [\mathrm{p53}_{\mathrm{weak}} \cdot \mathrm{MDM2}])([\mathrm{MDM2}]_f + [\mathrm{p53}_{\mathrm{weak}} \cdot \mathrm{MDM2}])}{[\mathrm{p53}_{\mathrm{tight}} \cdot \mathrm{MDM2}]}
\end{aligned} \tag{15}
$$

SI Figures

bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

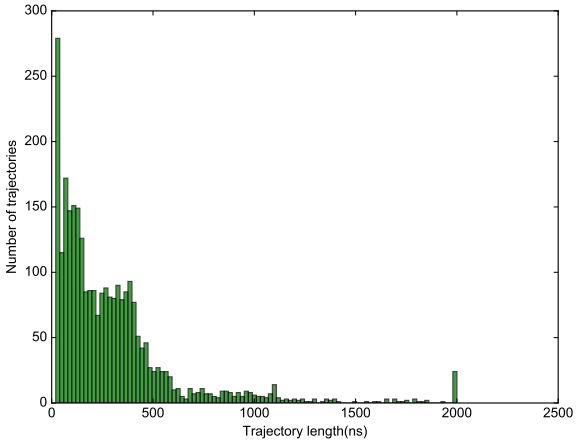

FIG. S1. Distribution of simulation trajectory lengths. A total of 2776 trajectories were generated on the Folding@home distributed computing platform, amounting to $\sim$ 831μs of aggregate simulation data.

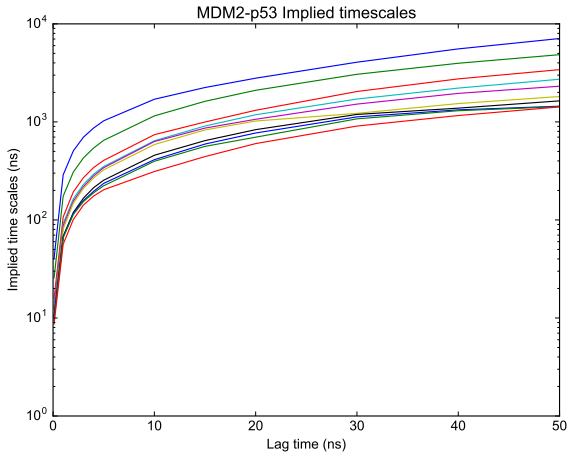

FIG. S3. Implied timescales vs MSM lag time. Shown are the ten slowest implied timescales.

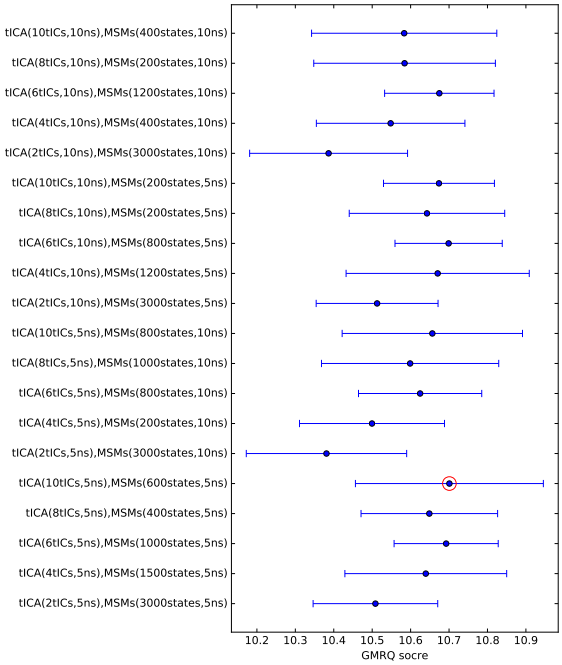

FIG. S2. GMRQ scores for MSMs built from various sets of hyper-parameters. Error bars are from 5-fold cross-validation. The hyper-parameters with the best score, marked by a red circle, was chosen for the construction of the final MSM.

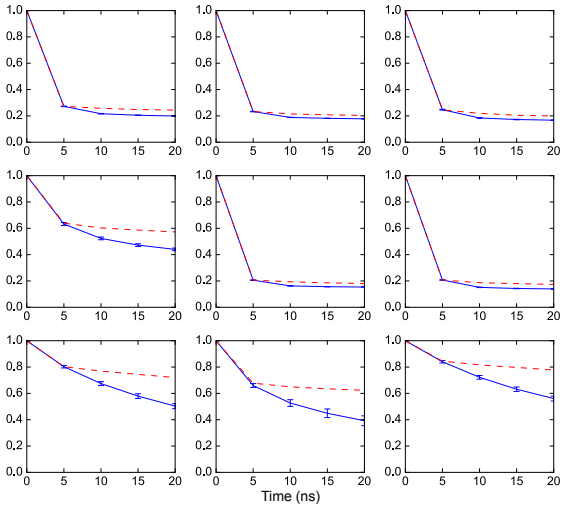

FIG. S4. Chapman-Kolmogorov test applied to the nine most populated states in the 600-microstate MSM. Red dashed curves are $\langle P(i, t+\tau|i, t)\rangle$ for state i, calculated from the MD simulations, and blue solid curves are populations $[(\mathbf{T}^{(\Delta t)})^n \delta_i]_i$ propagated from the MSM transition matrix, where δ_i is an initial population vector with state i containing the entire population, and $n\Delta t = \tau$ where $\tau = 5$ ns. Error bars are calculated from five bootstraps.

bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

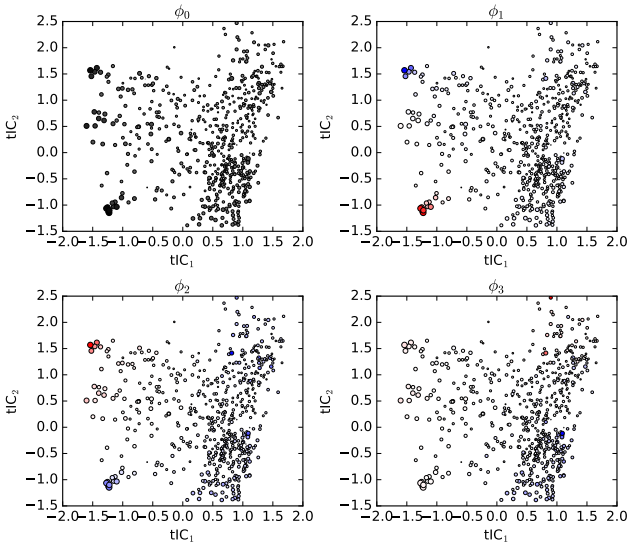

FIG. S5. Relaxation eigenmodes of the 600-state MSM. Relaxation dynamics from an initial population vector $\mathbf{p}(0)$ is given by $\mathbf{p}(t) = \sum_n \langle \psi_n | \mathbf{p}(0) \rangle \phi_n e^{-t/\tau_n}$, where ψ_n and ϕ_n are the left and right eigenvectors of the MSM transition matrix, respectively. Shown are the four slowest relaxation eigenmodes ϕ_n, including the stationary state (i.e. equilibrium populations), ϕ_0, projected to coordinates tIC$_1$ and tIC$_2$. The size of each circle is proportional to the equilibrium population, while the color corresponds to the eigenvector structure, with population flux along each eigenmode flowing from blue to red. The slowest relaxation, ϕ_1 mainly corresponds to binding flux along tIC$_1$ while ϕ_2 corresponds to binding flux along tIC$_2$, both on ~ 1 μs timescales (see Figure S3).

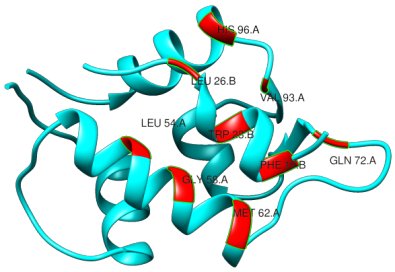

FIG. S6. A visualization of residues selected for the computation of p53 distances to the MDM2 binding pocket (see Supporting Methods).

bioRxiv preprint doi: https://doi.org/10.1101/086272; this version posted November 7, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supporting Tables

TABLE S1. Estimated rates for flux analysis.

Parameter	Value (Tmatrix)	Value (TPT)
k^w_{on}	1.8×10^9 M^{-1} s^{-1}	1.2×10^9 M^{-1} s^{-1}
k^w_{off}	2.9×10^6 s^{-1}	1.8×10^6 s^{-1}
k_{wt}	1.1×10^4 s^{-1}	1.7×10^4 s^{-1}
k_{tw}	1.0×10^7 s^{-1}	9.0×10^5 s^{-1}
k^{MDM2}_{wt}	9.6×10^5 s^{-1}	1.1×10^6 s^{-1}
k^{MDM2}_{tw}	3.9×10^6 s^{-1}	4.4×10^6 s^{-1}
k^t_{on}	4.0×10^9 M^{-1} s^{-1}	8.7×10^7 M^{-1} s^{-1}
k^t_{off}	2.9×10^4 s^{-1}	1.8×10^4 s^{-1}

TABLE S2. Corrected off-rates for the four-state binding mechanism model

Parameter	Corrected value
k^w_{off}	6.0 s^{-1}
k^t_{off}	0.06 s^{-1}

TABLE S3. Inferred rates for high p53 helicity

Parameter	Inferred rates
k_{wt} ($h\% = 28\%$)	3.2×10^5 s^{-1}
k_{tw} ($h\% = 28\%$)	4.7×10^4 s^{-1}
k_{wt} ($h\% = 64\%$)	6.9×10^5 s^{-1}
k_{tw} ($h\% = 64\%$)	2.2×10^4 s^{-1}