Supporting Information

AlaScan: A graphical user interface for alanine scanning free-energy calculations

Vijayaraj Ramadoss,[†] François Dehez,[†] and Christophe Chipot[*,†,‡]

Laboratoire International Associé Centre National de la Recherche Scientifique et University of Illinois at Urbana-Champaign, Unité Mixte de Recherche n°7565, Université de Lorraine, B.P. 70239, 54506 Vandœuvre-lès-Nancy cedex, France, and Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801

E-mail: chipot@ks.uiuc.edu

[*]To whom correspondence should be addressed

[†]Laboratoire International Associé CNRS/UIUC

[‡]Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801

Methods

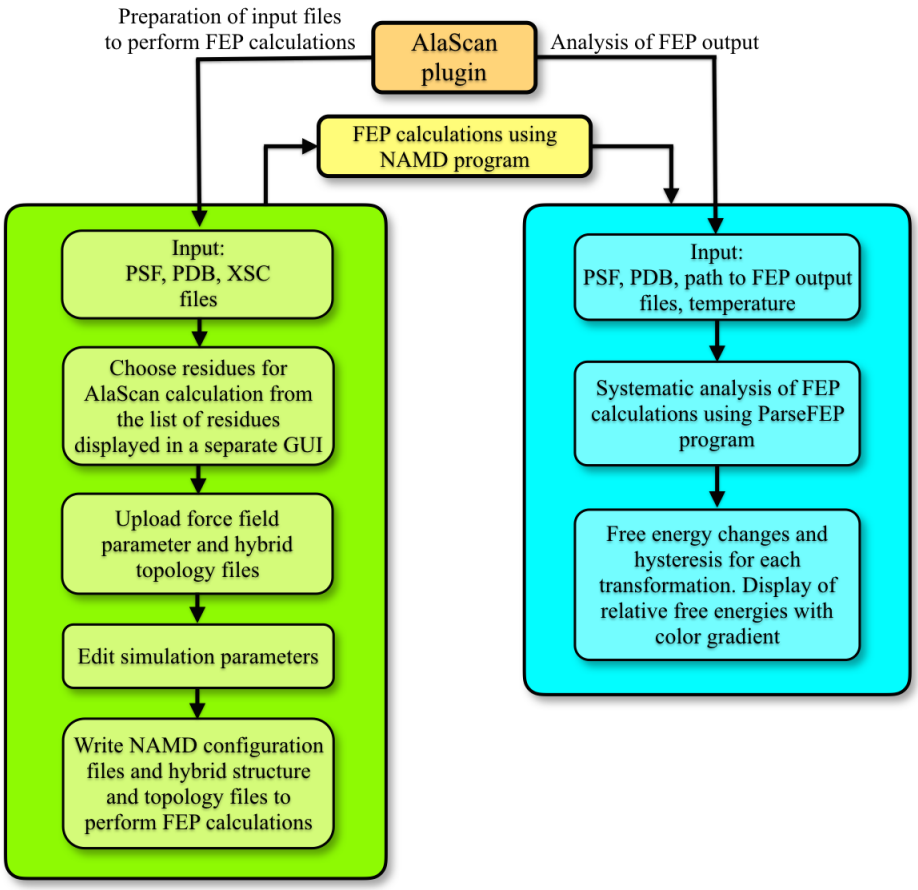
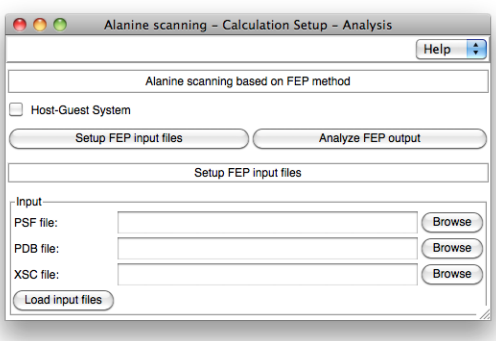
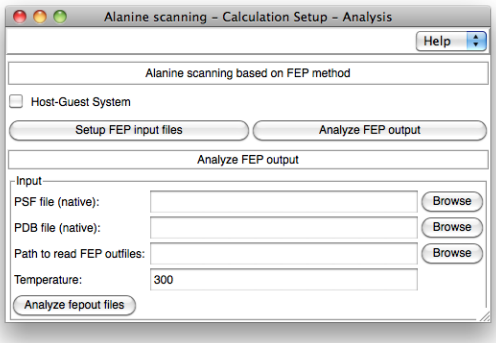


Figure S1: Schematic workflow of the AlaScan plug–in. A variety of options are available to select residues for in–silico alanine scanning — e.g., VMD Extension Functions[1] can be employed to highlight residues forming the host–guest interface.



(a)



(b)

Figure S2: Main graphical user interface (GUI) of the AlaScan plug–in designed to prepare the input files required for in–silico alanine scanning (a), and to perform systematic analysis of ASM free-energy calculations (b).

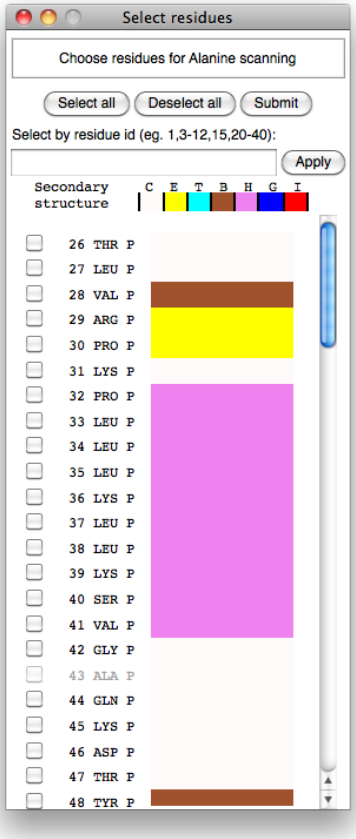


Figure S3: Secondary graphical user interface (GUI) of the AlaScan plug–in designed to select amino acids towards ASM free–energy calculations.

Illustration and computational details

For in–silico ASM experiments on an isolated protein, a lookup table containing the free energies associated with the alanine replacement of standard amino acids has been constructed. The reported values correspond to the lower, horizontal transformation, i.e., the unfolded state ($\Delta G^0_{\text{mutation}}$) of the thermodynamic cycle presented in Figure 1b of the main text. To estimate the free–energy change due to alanine substitution, using FEP, capped single amino acids, i.e., ACE—X—CT3 were generated, where X represents any amino acid, except alanine.

The choice of capped amino acid has proven to constitute a reasonable option to model a fully unfolded protein chain — while side–chain analogs are admittedly too rudimentary.[2] The present

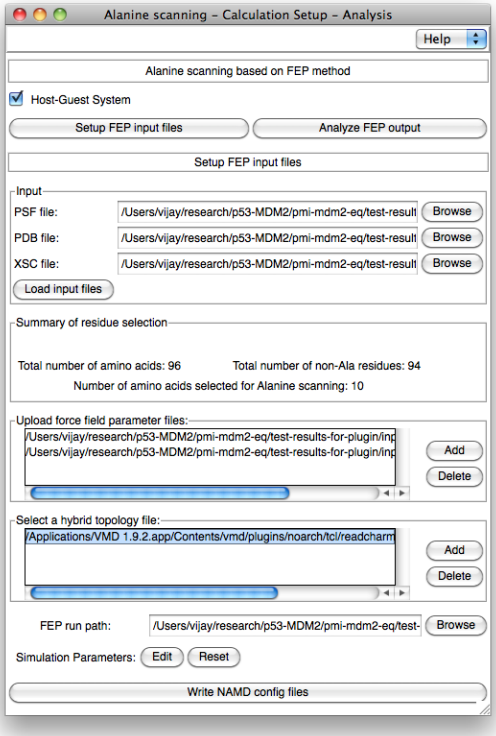


Figure S4: Screenshot of the GUI of AlaScan plug–in to setup FEP input files for alanine scanning calculations.

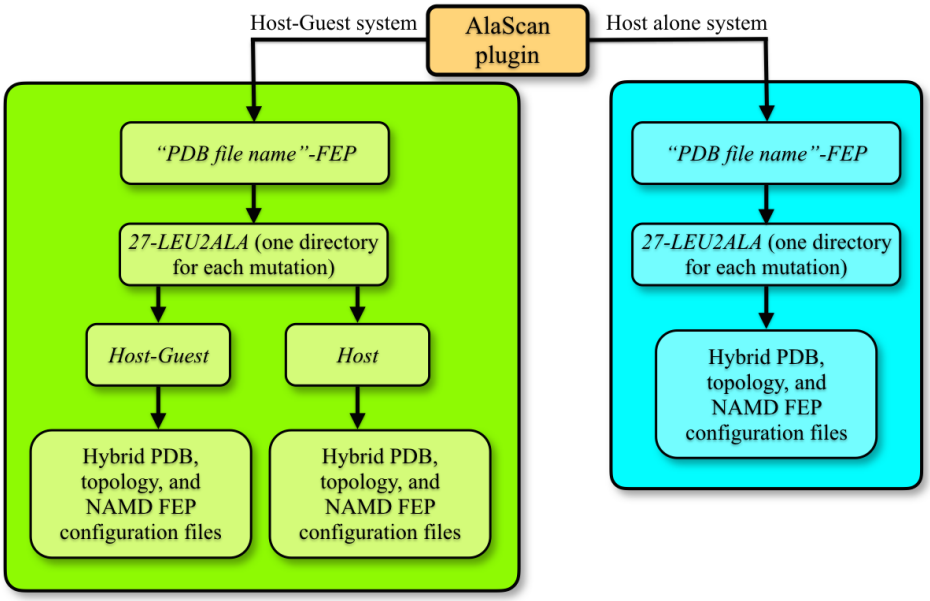


Figure S5: Overview of the different directories and files generated by "Setup FEP input files" section of AlaScan plug–in. The directory names are shown in italics.

choice rests upon the arguable assumption of an extended chain, wherein the side chain interacts at most with the backbone, but does not interact with other amino acids. The modular design of the plug–in allows the end–user to substitute the current lookup table based on capped single amino acids by an alternate one obtained using more elaborate models.[3]

Each capped amino acid was hydrated in a periodic cubic cell with 12–Å headspace in the three directions of Cartesian space, using the TIP3P water model.[4] The free–energy calculations were prefaced by a 2–ns thermalization.

Conversely, for in–silico ASM experiments on a host–guest complex, use was made of the short α–helical peptide inhibitor PMI (TSFAEYWNLLSP) bound to the protein MDM2 (PDB id 3EQS[5]). This protein–ligand complex has been the object of a number of previous investigations and, thus, constitutes an ideal candidate for alanine–scanning calculations and a cogent test of the AlaScan plug–in.[6] The stability of the PMI–MDM2 complex and the contributions of various PMI residues to the formation of the complex has been studied experimentally.[7] The PMI–MDM2 complex structure was placed in a periodic cubic cell together with of 6,836 TIP3P water molecules, corresponding to initial dimensions of $60 \times 60 \times 60$ Å^3. With the counterions to ensure electric neutrality, the molecular assembly consisted of 22,142 atoms. It was energy minimized for 10,000 steps prior to a 5–ns thermalization.

All MD simulations and FEP calculations were performed in the isobaric–isothermal ensemble, using NAMD 2.10[8] with the all–atom CHARMM36 force field.[9] The AlaScan is provided with an updated library of dual topologies for all possible mutations of naturally occurring amino acids, consistent with the CHARMM36 force field. The temperature and pressure were maintained at 300 K and 1 atm, respectively, using softly damped Langevin dynamics and the Langevin piston.[10] The van der Waals interactions were smoothly switched to zero between 10 and 12 Å. The equations of motion were integrated with the Verlet–I/r–RESPA multiple time–step algorithm,[11] with a time step of 2 fs for all bonded and short–range non–bonded interactions, and of 4 fs for long–range electrostatic interactions. Long–range electrostatic interactions were computed using the PME al-

gorithm.[12] Chemical bonds involving hydrogen atoms were constrained to their equilibrium length by means of the Rattle algorithm.[13] The Lennard-Jones potential was scaled–shifted to avoid the singularities arising from particle creation or deletion at the end points of the reaction pathway in FEP calculations.[14]

Results

In–silico alanine–scanning experiments on PMI bound to MDM2. The oncoproteins, MDM2 and MDMX, interact with the p53 transactivation domain (p53TAD) and negatively regulates the function of p53.[15,16] Inhibition of the interaction of p53TAD and its negative regulators MDM2 and MDMX is of topical interest in cancer therapy.[17,18] The dodecamer PMI inhibitor analogues to p53TAD competes with p53 for MDM2 binding, and is a stronger binder about two orders of magnitude compared to the native p53TAD peptide (ETFSDLWKLLPE).[6] To illustrate the versatility of the AlaScan plug–in for host–guest complexes, in–silico ASM calculations were performed on the PMI peptide, employing FEP. The equilibrated complex has been utilized as an initial structure and input of the plug–in to prepare the different input files required for each free–energy calculation carried out on the PMI peptide. Since alanine scanning is performed on the latter, the dodecamer will be treated by the plug–in as the host. The relative binding free energies were determined considering the peptide in the presence and in the absence of the MDM2 protein. Prior to carrying out the FEP calculations, each hybrid structure, either in its unbound state in water (host), or in its bound state, as part of the PMI–MDM2 complex (host–guest), was subjected to 10,000 cycles of energy minimization followed by 100 ps of thermalization.

Table 1 indicates the number of λ intermediate states and the simulation lengths of the forward and backward FEP calculations, alongside the free–energy contributions determined for the host–guest PMI–MDM2 complex. As can be observed from the experimental binding affinities, PMI residues Phe^3, Tyr^6, Trp^7 and Leu^{10} contribute significantly to the association. Conversely, PMI

Table S1: FEP calculation parameters and free–energy differences (in kcal/mol) obtained from the in–silico ASM experiments on the PMI–MDM2 complex. The $\Delta\Delta G$ value (in kcal/mol) represents the relative binding free energy of PMI to MDM2 upon mutation to alanine.

residue and mutation[a]		PMI–MDM2 (host–guest)		PMI (host)		$\Delta\Delta G$	Experiment[c]
		λ states[b]	ΔG (hysteresis)	λ states[b]	ΔG (hysteresis)		
1	T2A	75 (0.5)	12.4 ($\pm$0.0)	54 (0.5)	12.5 ($\pm$0.1)	-0.1	0.39
2	S2A	68 (0.5)	-2.5 ($\pm$0.1)	61 (0.5)	-2.9 ($\pm$0.1)	0.4	1.24
3	F2A	94 (1.0)	-3.2 ($\pm$0.3)	75 (0.5)	-7.7 ($\pm$0.2)	4.5	5.46
6	Y2A	70 (1.0)	10.7 ($\pm$0.3)	69 (1.0)	10.6 ($\pm$0.1)	0.1	3.06
7	W2A	77 (1.0)	-14.8 ($\pm$0.5)	78 (1.0)	-20.5 ($\pm$0.2)	5.7	6.31
8	N2A	75 (0.2)	77.2 ($\pm$0.2)	52 (0.5)	78.1 ($\pm$0.4)	-0.9	-1.10
9	L2A	64 (0.2)	12.0 ($\pm$0.2)	50 (0.5)	11.8 ($\pm$0.3)	0.2	-0.17
10	L2A	65 (0.5)	15.1 ($\pm$0.0)	51 (0.5)	13.2 ($\pm$0.3)	1.9	3.28
11	S2A	79 (0.5)	-2.4 ($\pm$0.0)	112 (0.5)	-3.0 ($\pm$0.1)	0.6	0.12
12	P2A	107 (0.5)	-16.4 ($\pm$0.0)	50 (0.5)	-17.3 ($\pm$0.3)	0.9	-0.25

[a]Native and mutated amino acids described by a single–letter code. [b]Steps (Simulation length in ns per λ intermediate state). [c]The experimental data were taken from reference 7.

residues Thr^1, Asn^8 and Leu^9, which exhibit considerable solvent accessibility, are much lesser contributors to the binding with MDM2. A user–friendly graphical representation, generated by the AlaScan plug–in, of various alanine–scanning values obtained from the FEP calculations is shown in Figure 2a of the main text. The free energy (ΔG) column in Figure 2a contains the free–energy change due to mutating a residue in the PMI peptide into alanine in the presence and in the absence of the guest protein. The hysteresis measured from each bidirectional free–energy calculation is given in parentheses. The green circles in the GUI Figure 2a are suggestive that all FEP calculations have properly converged (see probability distribution functions in Figure S6). It is noteworthy that the relative binding free energies, $\Delta\Delta G$, for residues Thr^1, Ser^2, Phe^3, Trp^7, Asn^8, Leu^9 and Ser^{11} depart by less than 1 kcal/mol from the experimental counterpart (see Table 1). Conversely, residues Tyr^6, Leu^{10} and Pro^{12} yield more discrepant results, in particular Tyr^6, for which the difference between theory and experiment amounts to about 3 kcal/mol. In the latter case, although the small hysteresis suggests appropriate convergence, replacement of the bulky tyrosine side chain by a mere methyl group represents a significant entropic change and entails local reorganization in the complex, which may not be easily captured in the relatively short simulations performed here for illustrative purposes. As can be seen in Figure 2a, a color–gradient

box is included in the $\Delta\Delta G$ column. It is constructed using the minima and maxima of the $\Delta\Delta G$ values (highlighted on a scale next to the secondary–structure color codes), and reveals at a glance whether or not the proposed alanine mutation is energetically favorable. For illustrative purposes, Figure S7 shows the results of poorly converged simulations. The orange and red "traffic lights" allow the end–user to pinpoint rapidly which free–energy calculation ought to be improved, or possibly rerun.

Table S 2: FEP simulation parameters and calculated free–energy change (ΔG in kcal/mol) for the alanine mutation on a series of capped single amino acids modeling the unfolded state of the protein in an aqueous environment.

mutation	λ–states	simulation length (in ns) per λ–state	ΔG (hysteresis)
ACE–I2A–CT3	50	0.5	-7.1 ($\pm$0.2)
ACE–L2A–CT3	50	1.0	11.5 ($\pm$0.1)
ACE–M2A–CT3	61	0.5	1.1 ($\pm$0.2)
ACE–F2A–CT3	76	0.5	-8.4 ($\pm$0.1)
ACE–W2A–CT3	70	1.0	-22.3 ($\pm$0.1)
ACE–Y2A–CT3	59	1.0	8.8 ($\pm$0.1)
ACE–V2A–CT3	50	0.2	-1.3 ($\pm$0.1)
ACE–S2A–CT3	50	0.2	-3.9 ($\pm$0.1)
ACE–T2A–CT3	50	0.2	17.1 ($\pm$0.0)
ACE–N2A–CT3	50	0.2	77.5 ($\pm$0.2)
ACE–Q2A–CT3	53	0.5	56.1 ($\pm$0.1)
ACE–C2A–CT3	50	0.2	-1.0 ($\pm$0.2)
ACE–G2A–CT3	50	0.5	6.9 ($\pm$0.0)
ACP–P2A–CT3	50	0.5	-20.6 ($\pm$0.0)
ACE–R2A–CT3	87	0.5	263.5 ($\pm$0.1)
ACE–H2A–CT3	66	0.2	36.7 ($\pm$0.1)
ACE–K2A–CT3	78	1.0	43.9 ($\pm$0.2)
ACE–D2A–CT3	100	1.0	131.2 ($\pm$0.2)
ACE–E2A–CT3	125	0.5	109.9 ($\pm$0.0)

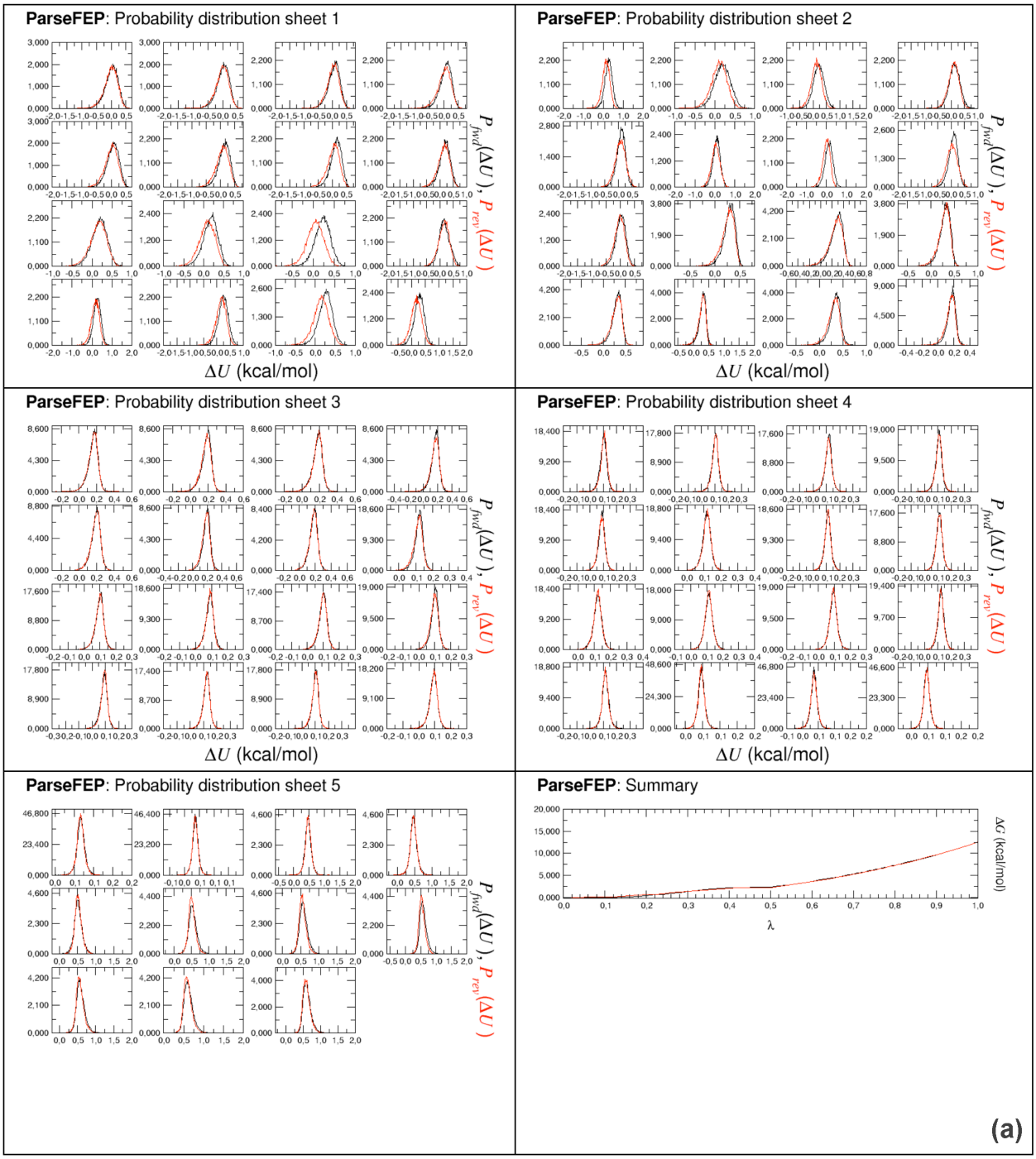
In–silico alanine–scanning experiments on the unfolded state. The final structure obtained from the equilibration of each capped single amino acid, ACE–X–CT3, in an aqueous environment was used as a starting structure for the FEP calculations. Prior to the FEP calculations, each dual–topology structure generated with the AlaScan plug–in was subjected to 1,000 steps of energy minimization and 100 ps of thermalization. The number of intermediate λ–states, the forward

and backward FEP simulation lengths, and the free–energy change associated with each alanine mutation are shown in Table 2. The very moderate hysteresis associated with each value of ΔG suggests that convergence has been attained for all calculations performed in the unfolded state (Table 2). The theoretical ΔG values were used to construct a lookup table in the AlaScan plug–in. This database serves as a reference for in–silico alanine–scanning experiments on isolated hosts aimed at probing their thermostability.

Put together, the alanine–scanning calculations performed on the PMI peptide in water (see Table 1) can be viewed as a proof of concept to demonstrate the applicability of the AlaScan plug–in to the host alone systems. Figure 2b of the main text gathers the results obtained from the alanine scanning calculations performed on the PMI peptide in its folded state (see Table 1) and in its unfolded state (see Table 2). It should be stressed, however, that no experimental data is available to appraise the reliability of the theoretical predictions. The results for this simple test system are presented here for illustrative purposes and to highlight the versatility of the AlaScan toolkit.

Availability of the plug-in

Version 1.0 of AlaScan is available as part of the popular visualization code VMD 1.9.2.[19] The plug–in is supplied with a library of hybrid topologies compliant with the macromolecular CHARMM36 force field.[9] Detail can be found in http://www.ks.uiuc.edu/Research/vmd/plugins/ alascanfep/.


ParseFEP: Probability distribution sheet 1
ParseFEP: Probability distribution sheet 2
ParseFEP: Probability distribution sheet 3
ParseFEP: Probability distribution sheet 4
ParseFEP: Probability distribution sheet 5
ParseFEP: Summary
ΔU (kcal/mol)
$P_{fwd}(\Delta U)$, $P_{rev}(\Delta U)$
ΔG (kcal/mol)
λ
(a)

…/…

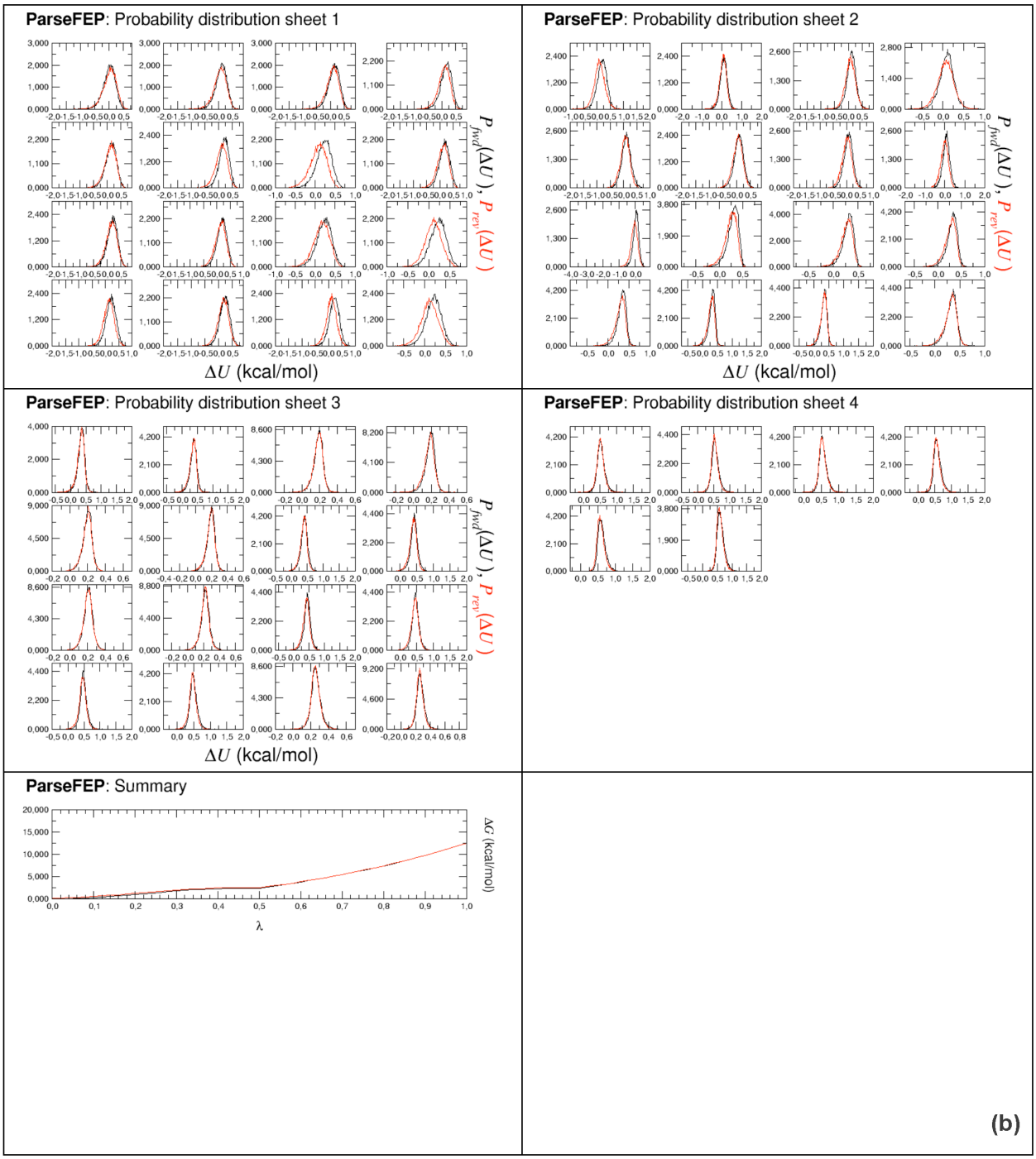


Figure S6: Probability distribution functions for the forward and backward T2A transformation in the PMI–MDM2 host–guest complex (a) and in the peptide only (b). The last graph depicts the net free–energy change in the bidirectional simulation. The different distributions are generated by the ParseFEP plug–in, invoked by AlaScan.

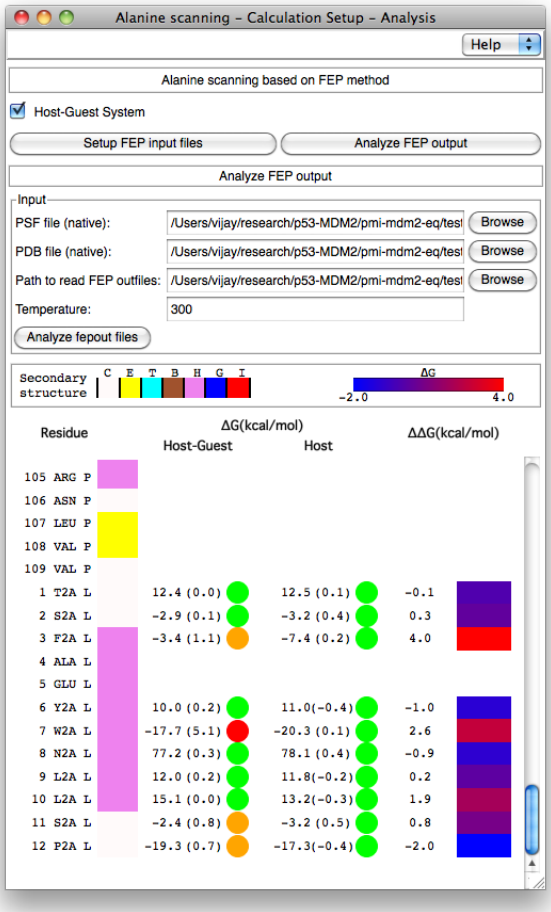


Figure S7: AlaScan GUI showing the results of a systematic analysis of a prototypically unconverged series of FEP calculations. The orange and red circles denote, respectively, a hysteresis greater than $k_{\mathrm{B}}T$, but less than $2k_{\mathrm{B}}T$, and a hysteresis greater than $2k_{\mathrm{B}}T$, suggestive that the corresponding free–energy calculations ought to be either improved, e.g., with a finer stratification strategy, or simply rerun with additional sampling.

References

(1) Giorgino, T. VMD Extension Functions. http://tonigi.github.io/vmd_extensions/index.html, 2014; GNU General Public License.

(2) König, G.; Boresch, S. Hydration Free Energies of Amino Acids: Why Side Chain Analog Data Are Not Enough. *J. Phys. Chem. B* **2009**, *113*, 8967–8974.

(3) Veenstra, D. L.; Kollman, P. A. Modeling Protein Stability: A Theoretical Analysis of the Stability of T4 Lysozyme Mutants. *Protein Eng.* **1997**, *10*, 789–807.

(4) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(5) Pazgier, M.; Liu, M.; Zou, G.; Yuan, W.; Li, C.; Li, C.; Li, J.; Monbo, J.; Zella, D.; Tarasov, S. G.; Lu, W. Structural Basis for High-affinity Peptide Inhibition of p53 Interactions with MDM2 and MDMX. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4665–4670.

(6) Li, C.; Pazgier, M.; Liu, W. Y., MinM.and Lu; Lu, W. Apamin as a Template for Structure–based Rational Design of Potent Peptide Activators of p53. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 8712–8715.

(7) Li, C.; Pazgier, M.; Li, C.; Yuan, W.; Liu, M.; Wei, G.; Lu, W. Y.; Lu, W. Systematic Mutational Analysis of Peptide Inhibition of the p53–MDM2/MDMX Interactions. *J Mol Biol* **2010**, *398*, 200–213.

(8) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, L., R. D. Kalé; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(9) MacKerell Jr., A. D. et al. All–atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(10) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulations — The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(11) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Phys. Chem. B* **1992**, *97*, 1990–2001.

(12) Darden, T. A.; York, D. M.; Pedersen, L. G. Particle Mesh Ewald: An $NLogN$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(13) Andersen, H. C. Rattle: a "velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.

(14) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994**, *100*, 9025–9031.

(15) Momand, J.; Zambetti, G. P.; Olson, D. C.; George, D.; Levine, A. J. The *Mdm*–2 Oncogene Product Forms a Complex with the p53 Protein and Inhibits p53–mediated Transactivation. *Cell* **1992**, *69*, 1237–1245.

(16) Oliner, J. D.; Pietenpol, J. A.; Thiagalingam, S.; Gyuris, J.; Kinzler, K. W.; Vogelstein, B. Oncoprotein MDM2 Conceals the Activation Domain of Tumour Suppressor p53. *Nature* **1993**, *362*, 857–860.

(17) Alarcon-Vargas, D.; Ronai, Z. p53–Mdm2 — The Affair That Never Ends. *Carciogenesis* **2002**, *23*, 541–547.

(18) Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; Fotouhi, N.; Liu, E. A. In Vivo Activation of the p53 Pathway by Small-molecule Antagonists of MDM2. *Science* **2004**, *303*, 844–848.

(19) Humphrey, W.; Dalke, A.; Schulten, K. VMD — Visual Molecular Dynamics. *J. Molec. Graphics* **1996**, *14*, 33–38.