JOURNAL
OF THE
AMERICAN
CHEMICAL
SOCIETY

Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies

Irina Massova and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446

Received March 22, 1999

Abstract: Noncovalent interactions are important in many physiological processes of complexation which involve all components of the living cells. Here we report an approach to computationally study the interaction free energies in protein−protein complexes which allows from a single simulation an estimate of the individual contribution of each residue to the binding. We developed this new technique—computational alanine scanning— and applied it to study the interactions of the oncoprotein Mdm2 to the *N*-terminal stretch of tumor suppressor protein p53. Excellent agreement has been found between the calculated and experimental data. This residue mutation methodology could prove to be a useful general design tool for molecules—nucleotides, peptides, lipids, or any other organic compound—optimized for interactions or stability, since one can qualitatively estimate the free energy consequences of many mutations from a single molecular dynamics trajectory.

Introduction

Noncovalent interactions are driving forces in numerous physiological processes. They are major determinants in many events which involve interactions between biomolecules. Noncovalent interactions play an important role in signal transduction, regulation, complexation, immune recognition, and adhesion. They are critical components of many processes such as the folding of proteins and nucleic acids, biocatalysis, inhibition, activation, partition, and distribution and are essential part of ligand−receptor and host−guest interactions.

The challenge is to computationally determine the free energy for noncovalent interactions in complex molecules. Recently Srinivasan et al.[1] developed the approach which combines molecular mechanical energies of the nucleic acid with a continuum solvent model to describe nucleic acid solvation free energies to evaluate the nucleic acid conformational dependent free energies. Here we report an extension of this approach to study protein−protein interactions. We have also expanded the methodology to estimate changes in the binding free energies as a result of mutating the residues of the interacting molecules.[2] We have applied this method to evaluate the binding free energies for complex formation between the tumor suppressor p53 and oncoprotein Mdm2, a protein−protein interaction. However, it can be easily extended to any noncovalent ligand− receptor, drug−target, or host−guest interaction.[3]

The protein complex of cellular regulatory phosphoprotein p53 with oncoprotein Mdm2[4] has been chosen as a model for our study because of the implication of these proteins in cancer. Transcription activator p53 plays an important role in the cell cycle. Its main purpose is to preserve the DNA material from

---
* To whom correspondence should be addressed. Phone: (415) 476-4637; Fax: (415) 476-0688; E-mail: pak@cgl.ucsf.edu.

(1) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.

(2) The program will be distributed with the AMBER6 release.

(3) Cieplak, P.; Kollman, P. A. DNA−Ligand interactions; work in progress; and Chong, L. T. et al., Antibody−Ligand Interactions. Submitted for publication.

(4) Abbreviations: hMdm2, human Mdm2; xMdm2, *Xenopus laevis* Mdm2.

the alterations[5,6] caused by diverse stress stimuli such as exposure to IR/UV,[5] γ radiation,[7] hypoxia,[8] depletion of nucleotide pool,[9] redox effect,[10] oncogeneses,[11] or chemotherapeutic drugs[12] (for review see Giaccia and Kastan[13]). As a response to these various stimuli, p53 causes cell growth arrest[5,6] through the transactivation-dependent and -independent mechanisms[14] to provide time for DNA-damage repair, or the damaged cell undergoes apoptosis depending on the origin of the tissue.[15]

The oncoprotein Mdm2 forms tight complexes with p53, thus preventing it from binding to DNA and regulating transcription activation. The complications in functioning of p53 and Mdm2 have been associated with many tumors. Thus, p53 has been found altered or missing in about half of human tumors.[16] Mdm2 is often amplified in a variety of cancers.[17] Although, p53 was first identified two decades ago, there is wide research ongoing on p53 and Mdm2 today, and many functions of these enzymes have just become known.[13,18]

The crystal structures of the p53-binding domain of human Mdm2 and its analogue from *Xenopus laevis* bound to the *N*-terminal stretch of human p53 shed light on the interactions of the oncoprotein and the tumor suppressor.[19] We have used both structures in our studies of Mdm2–p53 complexes. The crystal structure revealed that the helical *N*-terminal part of p53 with the three important residues—Phe-19, Trp-23, and Leu-26—fits perfectly into the extended hydrophobic binding site of Mdm2.[19] The critical role of these three residues was also emphasized in experiments with phage display libraries of Böttger et al.,[20] who identified the small peptides that are liable to inhibit the complexation of Mdm2 and p53. We have computationally mutated all 11 residues of the peptides derived from the Mdm2-binding stretch of p53 to alanine and have evaluated the effect of such mutations on the binding free energy. This work represents the first theoretical computational approach to alanine-scanning mutagenesis.

Methods

The structure of human and *X. laevis* Mdm2 (hMdm2 and xMdm2[4]) with human p53 reported in the PDB have been used in this research (reference codes 1ycr and 1ycq, respectively).[19] Though the 15–29 stretch of p53 has been used in crystallization experiments, residues 15 and 16 were not resolved due to the absence of electron density. The replacement experiments on small peptides reported in ref 20 were used as experimental criteria for our computational alanine-scanning mutagenesis calculations. These sequences corresponded to the 12-residue stretch of p53 from Gln-16 to Pro-27. Therefore, we inserted the 12-residue version of p53 (Ac-QETFSDLWKLLP-NH$_2$) into the active site of both Mdm2's. We will call the 12-residue p53 part (residues 16–27) a *peptide* throughout the text vs the term *protein* used for Mdm2 (residues 25–109 for human enzyme and 21–108 for *X. laevis*). Several sets of the dynamics simulation experiments were carried on complexes, peptides, and proteins separately. All molecules were solvated by a cubic box of TIP3P waters extended at least 10 Å in each direction from the solute. The periodic boundary conditions and the 8 Å cutoff for nonbonded van der Waals interactions were applied. The particle mesh Ewald method[21] was used to treat the long-range electrostatics. All systems were gradually heated to 300 K in three intervals, allowing a 5 ps interval per each 100 K, and then equilibrated for 25 ps at 300 K, followed by the 400 ps data collection runs. The 2 fs time step has been used with constant temperature and pressure, and periodic boundary conditions were applied. The temperature and pressure coupling parameters were both 0.2 ps. A total of 400 snapshots were saved during data collection period, one snapshot per each 1 ps of dynamic simulation.

The binding free energies ($\Delta G_{\text{binding}}$) were estimated from the absolute energies in the gas phase (E_{gas}), the solvation free energies ($G_{\text{PB}} + G_{\text{nonpolar}}$), and the vibration, rotation, and translation entropies for the complex, protein, and peptide. These terms were determined as reported elsewhere.[1] Here, we briefly describe the procedure:

$$\Delta G_{\text{binding}} = \Delta G_{\text{water}}(\text{complex}) - [\Delta G_{\text{water}}(\text{protein}) + \Delta G_{\text{water}}(\text{peptide})]$$

where the free energies for each species were evaluated by the following scheme:

$$\Delta G_{\text{water}} = E_{\text{gas}} + \Delta G_{\text{solvation}} - TS$$

$$G_{\text{solvation}} = G_{\text{PB}} + G_{\text{nonpolar}}$$

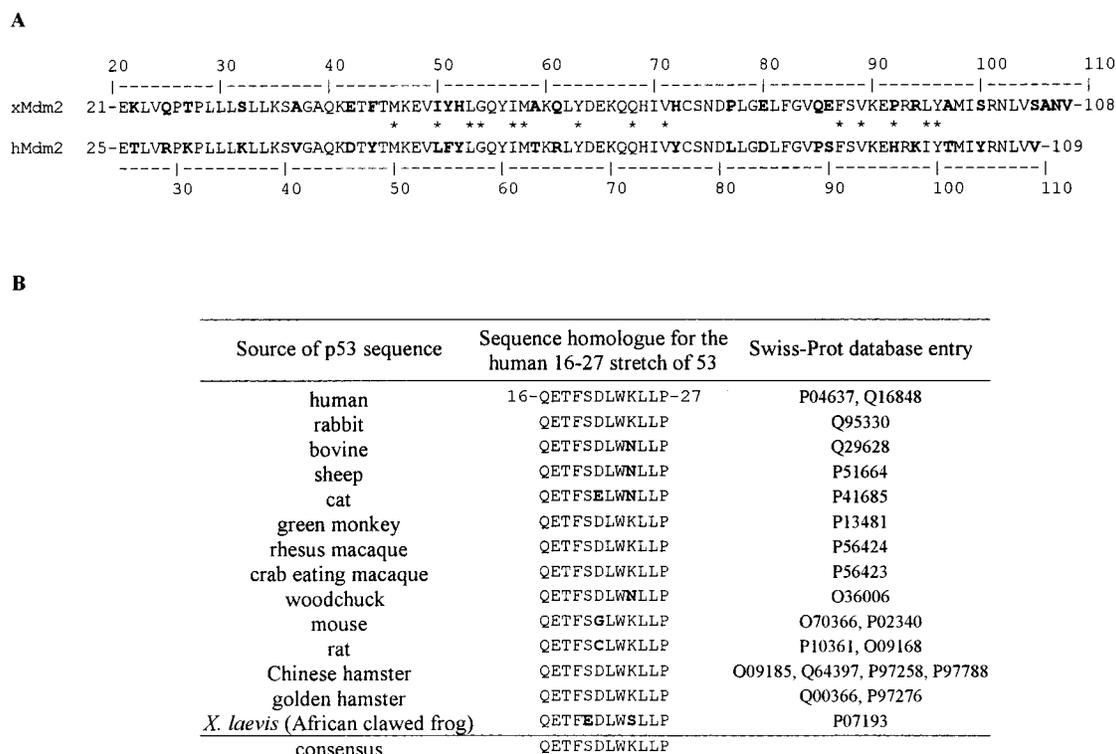$$E_{\text{gas}} = E_{\text{internal}} + E_{\text{electrostatic}} + E_{\text{vdW}}$$

$$E_{\text{internal}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}}$$

where T is temperature, S is the entropy, and the E_{bond}, E_{angle}, and E_{torsion} are contributions to the internal energy (E_{internal}) caused by the strain from the deviation of the bonds, angles, and torsion angles from their equilibrium values. The energy terms $E_{\text{electrostatic}}$ and E_{vdW} are the electrostatic and van der Waals interaction energies, respectively. The absolute energies in the gas phase (E_{gas}) were calculated with the *anal* module from the AMBER molecular modeling software package. The infinite cutoffs for all interactions and parm94 force field parameters were applied.[22] The entropy was estimated with the *nmode* module from AMBER based on the crystallographic structures of the complex and protein without water molecules. The normal-mode analysis was carried out for the energy-minimized structures of the complex, protein, and 12-residue peptide using a dielectric constant of $4r_{ij}$ (where r_{ij} is the distance between *i*th and *j*th atoms). As has been elaborated in more detail in ref 1, the entropy change is the most difficult term to estimate. The harmonic approximation has been used to assess the order of magnitude for the entropy loss during the complex formation. The polar contribution to the solvation free energy (G_{PB}) was evaluated using the *DelPhi*[23] program. A 0.5 Å spacing grid which extended 20% in all directions beyond the dimensions of the solute was used. A total of 1000 iterations has been performed to achieve a better convergence.[1] A dielectric constant of 1 was used for the solute and a dielectric constant of 80 for the exterior medium. All atoms had PARSE[23] van

(5) Kastan, M. B.; Onyekwere, O.; Sidransky, D.; Vogelstein, B.; Craig, R. W. *Cancer Res.* **1991**, *51*, 6304–6311.

(6) Lane, D. P. *Nature* **1992**, *358*, 15–16.

(7) Kuerbitz, S. J.; Plunkett, B. S.; Walsh, W. V.; Kastan, M. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7491–7495.

(8) Graeber, T. G.; Peterson, J. F.; Tsai, M.; Monica, K.; Fornace, A. J., Jr.; Giaccia, A. J. *Mol. Cell. Biol.* **1994**, *14*, 6264–6277.

(9) Linke, S. P.; Clarkin, K. C.; Di Leonardo, A.; Tsou, A.; Wahl, G. M. *Genes Dev.* **1996**, *10*, 934–947.

(10) Schwartz, J. L.; Antoniades, D. Z.; Zhao, S. *Ann. N.Y. Acad. Sci.* **1993**, *686*, 262–278.

(11) Debbas, M.; White, E. *Genes Dev.* **1993**, *7*, 546–554.

(12) Sinkovics, J. G. *Acta Microbiol. Hung.* **1991**, *38*, 321–334.

(13) Giaccia, A. J.; Kastan, M. B. *Genes Dev.* **1998**, *12*, 2973–2983.

(14) Bennett, M.; Macdonald, K.; Chan, S. W.; Luzio, J. P.; Simari, R.; Weissberg, P. *Science* **1998**, *282*, 290–293.

(15) Symonds, H.; Krall, L.; Remington, L.; Saenz-Robles, M.; Lowe, S.; Jacks, T.; Van Dyke, T. *Cell* **1994**, *78*, 703–711.

(16) Kaelin, W. G. *Science* **1998**, *281*, 57–58.

(17) Oliner, J. D.; Kinzler, K. W.; Meltzer, P. S.; George, D. L.; Vogelstein, B. *Nature* **1992**, *358*, 80–83.

(18) Sun, P.; Dong, P.; Dai, K.; Hannon, G. J.; Beach, D. *Science* **1998**, *282*, 2270–2272.

(19) Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. *Science* **1996**, *274*, 948–953.

(20) Böttger, A.; Böttger, V.; Garcia-Echeverria, C.; Chène, P.; Hochkeppel, H. K.; Sampson, W.; Ang, K.; Howard, S. F.; Picksley, S. M.; Lane, D. P. *J. Mol. Biol.* **1997**, *269*, 744–756.

(21) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(22) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(23) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144–1149.

A

```
            20        30        40        50        60        70        80        90       100       110
            |---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
xMdm2   21-EKLVQPTPLLLSLLKSAGAQKETFTMKEVIYHLGQYIMAKQLYDEKQQHIVHCSNDPLGELFGVQEFSVKEPRRLYAMISRNLVSANV-108
                              *  *   **  **     *     *  *          * *  *  **
hMdm2   25-ETLVRPKPLLLKLLKSVGAQKDTYTMKEVLFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYTMIYRNLVV-109
            -----|---------|---------|---------|---------|---------|---------|---------|---------|---------|
                 30        40        50        60        70        80        90       100       110
```

B

Source of p53 sequence	Sequence homologue for the human 16-27 stretch of 53	Swiss-Prot database entry
human	16-QETFSDLWKLLP-27	P04637, Q16848
rabbit	QETFSDLWKLLP	Q95330
bovine	QETFSDLWNLLP	Q29628
sheep	QETFSDLWNLLP	P51664
cat	QETFSELWNLLP	P41685
green monkey	QETFSDLWKLLP	P13481
rhesus macaque	QETFSDLWKLLP	P56424
crab eating macaque	QETFSDLWKLLP	P56423
woodchuck	QETFSDLWNLLP	O36006
mouse	QETFSGLWKLLP	O70366, P02340
rat	QETFSCLWKLLP	P10361, O09168
Chinese hamster	QETFSDLWKLLP	O09185, Q64397, P97258, P97788
golden hamster	QETFSDLWKLLP	Q00366, P97276
X. laevis (African clawed frog)	QETFEDLWSLLP	P07193
consensus	QETFSDLWKLLP	

Figure 1. (A) Amino acid sequence alignment for p53-binding domains of xMdm2 and hMdm2. Different residues in both sequences are shown in bold type. Residues participating in interaction to the p53 are marked with asterisks (*). (B) The amino acid sequence alignment for Mdm2-binding *N*-terminal stretch of p53 from various species. Residues that are different from the human sequence are shown in bold type.

der Waals radii and Cornell et al. charges.[22] The nonpolar contribution to the solvation free energy was calculated from the equation: $G_{nonpolar} = \gamma SA + b$,[24] where $\gamma = 0.00542$ kcal/Å^2, $b = 0.92$ kcal/mol, and SA is the solvent accessible area estimated with the program MSMS.[25] The alanine mutant structures were generated based on the structures of the collected snapshots by truncating the mutated residue at C_γ and by replacing C_γ with a hydrogen atom at a 1.09 Å distance from C_β along the $C_\gamma - C_\beta$ bond. All parameters in the topology files for the mutated residue were accordingly replaced with the alanine residue parameters. This was done with a program developed in our laboratory.[2] Data for the processing of the original (nonmutated) trajectories were collected by two protocols. In the first protocol, the structures of the complex, protein and peptide were taken from the same snapshot of the dynamics trajectory of the complex in a water box. Then, all energy components and the binding free energy were calculated for all three molecules generated from every fourth snapshot (total of 100 snapshots out of 400 were processed) and the statistical analysis was carried out. In the second protocol, the structures of each of three species (complex, protein, and peptide) were sampled alone in a box of water. Then statistical analysis was carried out for each trajectory separately for every fourth snapshot (total of 100 snapshots out of 400 collected were processed for each trajectory). The analyses of the trajectories from the separate runs (protocol 2) resulted in higher standard deviations for the energy components and the binding energies than the analyses based on structures collected from the same trajectory (protocol 1) because the energies based on structures from the separated trajectories represented the independent measurements compared to the energy values for structures of all three species from the same trajectory.
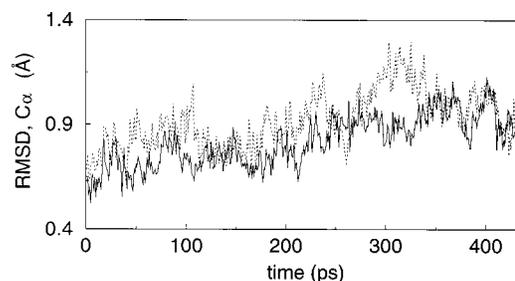
Results and Discussion

A. Free Energy Evaluation for p53-Derived 12-Residue Binding to xMdm2.
The p53-binding domain of xMdm2 shares

(24) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(25) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305–320.

Figure 2. Root-mean-square deviations for the C_α during the dynamics simulations of complexes of p53-binding domains of xMdm2 (solid lines) and hMdm2 (dotted lines) with 12-residue peptide derived from human p53.

72.9% of amino acid sequence identity with the human homologue (Figure 1A). The variable positions of both enzymes are shown in bold type, and the majority represents conservative mutations. As could be seen, two enzymes share a high degree of amino acid sequence identity, particularly among the residues lining the active sites. There are only two conservative mutations, Ile-50 and Pro-92, of xMdm2, which are replaced by Leu-54 and His-96, respectively, in human enzyme among all 14 residues which have contacts to p53. The structure of the hMdm2 complex was determined at 2.6 Å resolution compared to the 2.3 Å xMdm2 complex with human p53. In addition, the xMdm2 structure was used for the refinement of the hMdm2 complex by molecular replacement.[19] The human complex was missing water molecules because they were not reported with the lower resolution structure. Also, some instability during dynamics simulations was observed for hMdm2 in contrast to xMdm2 (Figure 2), which could be related to the unaccounted cavity waters in the hMdm2 structure. The trajectories for the human complex undergo long periodic fluctuations which might be the result of missing cavity waters and total positive charge

Table 1. Energy Terms and the Binding Free Energies for xMdm2 Complex with Ac-QETFSDLWKLLP-NH$_2$ Peptide

contribution	complex mean	complex std	protein mean	protein std	ligand mean	ligand std
(A) Protocol 1: Complex, Protein, and Peptide Structures Are from the Same Trajectory						
$E_{electrostatic}$	−2951.55	45.91	−2584.75	33.92	−265.47	24.41
E_{vdW}	−399.02	15.20	−331.01	13.66	−12.79	4.27
$E_{internal}$	1721.32	21.17	1502.01	21.26	219.31	10.55
E_{gas}	−1629.25	48.47	−1413.75	39.50	−58.95	27.50
G_{PB}	−1653.60	39.98	−1509.27	33.17	−275.33	24.41
$G_{nonpolar}$	35.52	0.52	32.66	0.44	10.22	0.19
$G_{subtotal}$	−3247.34	23.16	−2890.36	23.17	−324.06	10.42
$-TS$	−1198.45	NA	−1041.93	NA	−184.96	NA
(B) Protocol 2: Complex, Protein, and Peptide Structures Are from Separate Trajectories						
$E_{electrostatic}$	−2951.55	45.91	−2498.32	36.95	−222.19	20.30
E_{vdW}	−399.02	15.20	−336.69	14.56	−14.44	5.55
$E_{internal}$	1721.32	21.17	1497.67	24.35	213.01	8.31
E_{gas}	−1629.25	48.47	−1337.34	40.92	−23.62	21.86
G_{PB}	−1653.60	39.98	−1579.06	36.50	−318.97	19.53
$G_{nonpolar}$	35.52	0.52	32.15	0.38	10.30	0.32
$G_{subtotal}$	−3247.34	23.16	−2884.25	20.19	−332.29	7.33

(C) Energy Terms and Total Binding Energies

contribution	protocol 1 mean	protocol 1 std	protocol 2 mean	protocol 2 std
$\Delta E_{electrostatic}$	−101.33	31.27	−231.05	62.33
ΔE_{vdW}	−55.22	3.23	−47.89	21.77
ΔE_{gas}	−156.55	32.09	−268.29[a]	67.09
ΔG_{PB}	130.99	33.11	244.43	57.55
$\Delta G_{nonpolar}$	−7.37	0.38	−6.93	0.72
$\Delta G_{subtotal}$	−32.93	4.11	−30.80	31.59
$-T\Delta S$	28.40	NA	28.40	NA
$\Delta G_{binding}$	−4.52	NA	−2.39	NA

[a] There is also a contribution of the $\Delta E_{internal} = 10.64 \pm 53.82$ kcal/mol to the ΔE_{gas}, which is due to the difference in individual $\Delta E_{internal}$ terms of the complex with protein and peptide not being canceled in protocol 2. This term is always equal to zero for protocol 1.

on the hMdm2 complex (+4e). The total neutrality of the xMdm2 complex, as well as all of the above factors made it a more attractive and suitable model to probe the interactions in the human complex. Therefore, we first conducted all of our experiments with the xMdm2 structure.

Table 1 shows all energy terms and the binding free energy for the xMdm2 complex with 12-residue peptides derived from human p53 (Ac-QETFSDLWKLLP-NH$_2$). Our estimates showed that the entropy term ($-T\Delta S$) adds an approximately equal contribution of ∼30 kcal/mol to the total binding energy for the wild type and the mutants (28−36 kcal/mol depending on how well the structures were minimized before the normal-mode analysis was done). We would expect these terms to cancel when the differences in the binding free energies are calculated between the wild type and the mutants. Therefore, we report only $\Delta G_{subtotal}$ (without the $-T\Delta S$ term) as the criteria for the $\Delta G_{binding}$ estimates throughout the paper, and the total energy estimates are shown only in Table 1.

The IC$_{50}$ for the binding of the Ac-QETFSDLWKLLP-NH$_2$ sequence is 2−14 μM depending on the different assaying methods.[20] This corresponds to −6.6 to −7.8 kcal/mol in $\Delta G_{binding}$. The data in Table 1C is approximately in that range: −4.5 and −2.4 kcal/mol for the two different protocols, which are in a very good agreement with experiment given the approximation of our methodology and the cancellation of large numbers required. The relative similarity of the $\Delta G_{binding}$ numbers for both protocols was surprising, given one's expecta-

Table 2. Computational Alanine-Scanning Mutagenesis Results for xMdm2 Complex with a 12-Residue Stretch of p53 ($\Delta\Delta = \Delta_{wildtype} - \Delta_{mutant}$)

contribution	Gln16Ala mean	Gln16Ala std	Glu17Ala mean	Glu17Ala std	Thr18Ala mean	Thr18Ala std	Phe19Ala mean	Phe19Ala std
$\Delta\Delta E_{electrostatic}$	−2.69	4.08	−66.29	25.43	−1.68	1.33	−0.76	0.66
$\Delta\Delta E_{vdW}$	−0.20	0.11	−0.72	1.15	−0.68	0.34	−8.82	0.88
$\Delta\Delta E_{gas}$	−2.89	4.13	−67.01	25.22	−2.36	1.37	−9.58	1.03
$\Delta\Delta G_{PB}$	2.68	4.25	68.19	26.34	2.23	1.31	6.20	1.17
$\Delta\Delta G_{nonpolar}$	0.00	0.18	−0.09	0.15	−0.04	0.08	−0.33	0.11
$\Delta\Delta G_{subtotal}$	−0.20	0.89	1.09	2.18	−0.17	0.46	**−3.71**	1.27

contribution	Ser20Ala mean	Ser20Ala std	Asp21Ala mean	Asp21Ala std	Leu22Ala mean	Leu22Ala std	Trp23Ala mean	Trp23Ala std
$\Delta\Delta E_{electrostatic}$	−0.20	1.71	−24.45	6.82	0.12	0.28	−4.76	1.34
$\Delta\Delta E_{vdW}$	−0.30	0.19	−0.08	0.02	−3.78	0.86	−11.70	1.46
$\Delta\Delta E_{gas}$	−0.50	1.67	−24.53	6.82	−3.66	0.89	−16.47	1.62
$\Delta\Delta G_{PB}$	0.86	1.68	24.22	6.77	2.17	0.95	10.77	1.04
$\Delta\Delta G_{nonpolar}$	−0.05	0.12	−0.01	0.08	−0.20	0.08	−0.58	0.15
$\Delta\Delta G_{subtotal}$	0.31	0.71	−0.32	0.42	**−1.68**	1.05	**−6.27**	1.55

| contribution	Lys24Ala mean	Lys24Ala std	Leu25Ala mean	Leu25Ala std	Leu26Ala mean	Leu26Ala std
$\Delta\Delta E_{electrostatic}$	33.12	5.22	−0.14	0.10	−0.34	0.15
$\Delta\Delta E_{vdW}$	−0.11	0.01	−0.21	0.07	−3.86	0.73
$\Delta\Delta E_{gas}$	33.01	5.22	−0.35	0.10	−4.20	0.75
$\Delta\Delta G_{PB}$	−32.98	5.43	0.12	0.30	2.73	0.70
$\Delta\Delta G_{nonpolar}$	−0.01	0.08	−0.02	0.08	−0.02	0.10
$\Delta\Delta G_{subtotal}$	0.03	1.50	−0.25	0.34	**−1.48**	0.97

tion that the binding energy should be significantly less favorable when using separate trajectories.

Generally, the absolute numbers for the $\Delta G_{binding}$(protocol 2; snapshots are from the separate trajectories) are higher than the $\Delta G_{binding}$(protocol 1; snapshots are from the same trajectory) because of the opposing trends in the $\Delta E_{electrostatic}$ and ΔG_{PB}, which are larger in absolute values for protocol 2. However, these energy terms do not cancel each other out and their contribution—a positive number—decreases the absolute value of the $\Delta G_{binding}$ to a smaller extent for protocol 2.

B. Computational Alanine-Scanning Mutagenesis of p53-Derived Peptide. Table 2 shows the results of a computational alanine-scanning mutagenesis approach for 11 residues out of a total of 12 of p53-derived peptide. Data are also depicted as a graph in Figure 3B. The proline residue was not mutated since its backbone conformation differs significantly from that of alanine. A Pro/Ala mutation sometimes produces the anomalous changes to the binding as a result of significant conformational changes.[26] We made a reasonable assumption that the mutation with alanine for the other 11 residues of p53-based peptide introduced only local changes to the structure of the peptide, that these changes did not effect the total binding modes for the peptide to protein, and that the shape of the complex was preserved. As has been shown previously on various alanine-scanning mutagenesis experiments, such an assumption works for most mutations.[27] Negative numbers in $\Delta\Delta G_{subtotal}$ means highly unfavorable substitutions. In contrast, positive $\Delta\Delta G_{subtotal}$ indicates the preference for the alanine residue at the mutated position. The results in Table 2 show the significant loss in the binding free energies when the three critical residues Phe-19, Trp-23, and Leu-26 are mutated to alanine. These data are in a good agreement with experiments that demonstrated that these positions are the only ones which do not allow any substitution

(26) Clackson, T.; Ultsch, M. H.; Wells, J. A.; de Vos, A. M. *J. Mol. Biol.* **1998**, *277*, 1111−1128.

(27) Cunningham, B. C.; Wells, J. A. *Science* **1989**, *244*, 1081−1085.

A

B

C
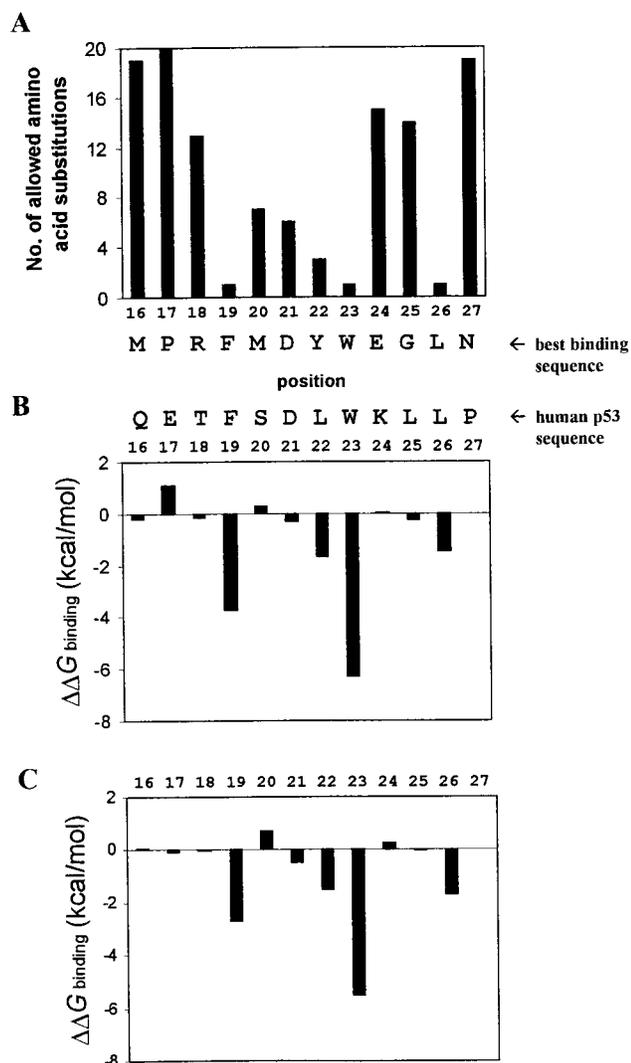
Figure 3. (A) Replacement series of peptides. An increase in IC_{50} of more than 3-fold compared to the original amino acid was considered a "not permitted" replacement. Printed from ref 20 with the permission of the Journal of Molecular Biology, Academic Press Limited. (B) The binding energy change for the computational alanine-scanning mutagenesis experiments for xMdm2 complex of 12-residue peptide derived from human p53. (C) The same as part B, only for hMdm2 complex. Negative numbers in $\Delta\Delta G_{binding}$ mean highly unfavorable substitutions. In contrast, positive $\Delta\Delta G_{binding}$ indicates the preference for alanine mutation.

with other residues for the set of small 12-residue peptides (Figure 3A).[20] The Leu22Ala mutation was also selected as unfavorable in agreement with the replacement experiments which allowed only two other substitutions of this amino acid.[20] Actually, these four positions (19, 22, 23, and 26) were the only mutations which resulted in the significant decrease of binding energies and which were the only positions selected as the least variable positions in biochemical experiments on small peptides (Figure 3A). These groups participate in hydrophobic interactions with the nonpolar groups lining the active site cleft of Mdm2 (Figure 4A). The largest input to the binding energy for these groups comes from the gain in van der Waals interactions ($\Delta\Delta E_{vdW}$ term; Table 2) which was not canceled by unfavorable changes in the solvation free energies ($\Delta\Delta G_{PB}$ term; Table 2). This gain was lost when these groups were substituted with the smaller Ala residue. In addition, the substitution of Trp-23 with alanine was also accompanied by a ∼ 4.8 kcal/mol loss in electrostatic energy, which is roughly equal to the energy of

one hydrogen bond. The proton on the Trp-23 side chain nitrogen participates in hydrogen bonding to the backbone carbonyl oxygen of Ile-50 in xMdm2 (Leu-54 in hMdm2; Figure 4A). The $\Delta\Delta G_{nonpolar}$ term is proportional to the change of the surface accessible area (see the Methods section). Though this term added ∼0.6 kcal/mol to the binding energy of Trp-23, ∼0.3 kcal/mol for Phe-19, and ∼0.2 kcal/mol for Leu-22 interactions to Mdm2, overall its contribution to the total binding energy was negligible and less than the standard deviation (std) of the $\Delta\Delta G_{subtotal}$. Presumably, the small contribution of the solvent accessible area dependent term to the free energy change is related to the larger accessibility of the backbone carbonyl and amide of alanine to the solvent although the side chain volume of alanine is smaller than that of the original residue. When the terms are combined in a different order ($\Delta\Delta E_{electrostatic}$ + $\Delta\Delta G_{PB}$) based on the common electrostatic origin of the contributions, then the nature of the favorable interactions for Phe-19, Leu-22, Trp-23, and Leu26 became more obvious. It arises from highly favorable van der Waals interactions ($\Delta\Delta E_{vdW}$). Nevertheless, all the terms except the nonpolar one are contributors to the total effect. The electrostatic interactions are the smallest for both leucines, and the total free energy change for Leu-22 and Leu-26 depends on the balance between the unfavorable loss in solvation free energy and the favorable gain in van der Waals interactions. These observations are particularly interesting in light of the fact that the substitution of the third leucine (Leu25Ala) has no effect on the binding. Position 25 allows 14 amino acid replacements, and our computational alanine mutagenesis experiment shows almost a zero effect of all contributions to the free binding energy for this mutation. Trp-23 contributes the most to the binding energy of p53 to xMdm2 among all sampled residues, and this finding is in excellent agreement with the survey of the crystallographic structures deposited to PDB which demonstrated the highest propensity of Trp residue to be found in protein−protein interfaces.[28] The Trp residue plays the main role in many protein−proteincomplexes.[26] The phenylalanine residue has the second highest propensity,[28] and p53 Phe-19 displays the second highest contribution to the binding of p53 to Mdm2 in our experiments. Obviously, nature optimized the p53-Mdm2 interactions through the steps of evolutionary mutation events to achieve the tight binding.

Excellent agreement with replacement experiments has been found.[20] Thus, the most unfavorable mutations on graph B are complimentary to the smallest peaks of graph A in Figure 3. Interestingly, the Glu17Ala and Ser20Ala mutations actually show the positive $\Delta\Delta G_{binding}$, which means that alanine residue provides a better binding. Although position 17 is almost nondiscriminatory and it could be replaced with any of 20 amino acids (Figure 3A), the best binding sequence selected in ELISA assays still had a more hydrophobic proline residue at this position rather than polar glutamic acid.[20] Obviously, the alanine replacement is a better choice than polar charged Glu and might be considered as a more favorable substitution. A similar situation occurs with Ser-20. Although seven amino acids were allowed at this position, the best binding sequence has a hydrophobic methionine rather than serine of the wild-type sequence.[20] This substitution also displayed a small increase in the binding for the alanine substitution in our analysis.

To evaluate whether the use of the wild-type trajectory for the computational alanine-scanning mutagenesis experiments, where the coordinates of the mutated residue were modified to that of alanine, would result in a correct conformational

(28) Jones, S.; Thornton, J. M. *J. Mol. Biol.* **1997**, *272*, 121−132.

A

B

Figure 4. (A) Stereoview of the crystal structure of the p53-binding domain of hMdm2 (dark) bound to 17−29 residue stretch of p53 (light). The coordinates were taken from PDB (reference code 1ycr).[19] (B) The stereoview of the energy minimized peptide derived from the 16−27 stretch of Leu22Tyr mutant of p53. The dotted lines represent hydrogen bonds in both figures. The two left structures could be viewed with stereoglasses, and the right combination represents the crossed eye projections.

Table 3. Comparison of the Results for the Computational Alanine-Scanning Mutagenesis Experiments Calculated for the Modified Trajectory of the Wild Type and for the Trajectory Collected for the p53 Trp23Ala Mutant[a]

	Trp23Ala modified trajectory of the wild type		trajectory of the Trp23Ala mutant	
contribution	mean	std	mean	std
$\Delta E_{electrostatic}$	−96.57	31.08	−141.03	18.45
ΔE_{vdW}	−43.52	3.10	−44.31	3.85
ΔE_{gas}	−140.09	31.93	−185.35	18.56
ΔG_{PB}	120.22	33.12	165.42	18.56
$\Delta G_{nonpolar}$	−6.79	0.38	−6.79	0.33
$\Delta G_{subtotal}$	−26.65	3.86	−26.72	3.70

[a] Data are collected and processed according to protocol 1.

Table 4. Comparison of the Energy Terms and the Total Binding Energy for the Wild Type and for the Trajectory Collected for the p53 Leu22Tyr Mutant[a]
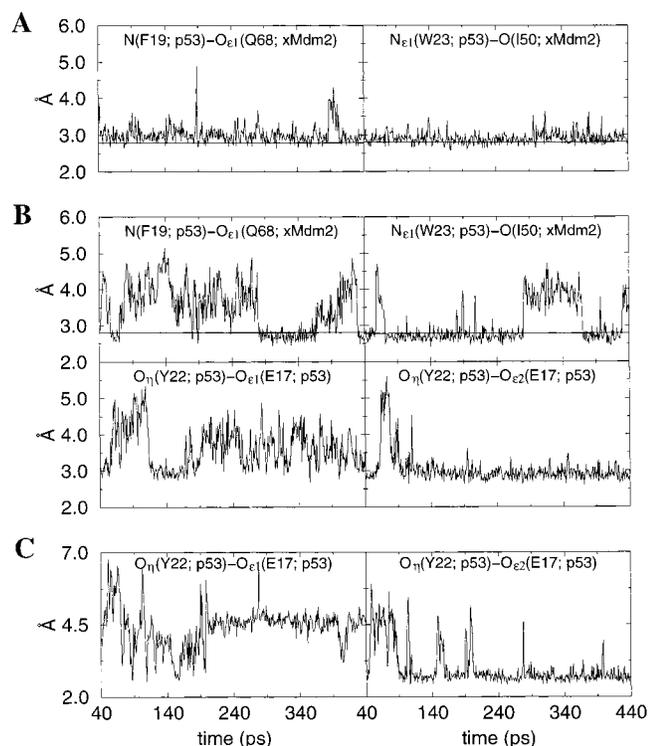
	wild type		Leu22Tyr mutant	
contribution	mean	std	mean	std
$\Delta E_{electrostatic}$	−101.33	31.27	−54.03	13.50
ΔE_{vdW}	−55.22	3.23	−56.25	3.45
ΔE_{gas}	−156.55	32.09	−110.28	13.24
ΔG_{PB}	131.00	33.11	83.90	12.95
$\Delta G_{nonpolar}$	−7.37	0.38	−7.20	0.25
$\Delta G_{subtotal}$	−32.93	4.11	−33.58	4.11

[a] Data are collected and processed according to protocol 1 for both trajectories.

sampling for the mutant, we conducted a set of separate dynamics calculations for the Trp23Ala mutant of p53-derived peptide in the complex with xMdm2 (Table 3). The values for the $\Delta G_{subtotal}$ are almost identical (−26.65 ± 3.86 kcal/mol for the mutated wild-type trajectory and −26.72 ± 3.70 kcal/mol for the trajectory of the mutant). These findings justified our use of the modified trajectories collected for the wild type for mutagenesis experiments rather than time-consuming separate sampling for each of the mutant structures.

C. Leu22Tyr Mutation in p53. Position 22 is known to be important for the binding,[29] and it was selected as such in our computational analysis. Moreover, the Leu22Tyr mutation showed the better binding for the 12-residue peptides. Thus,

(29) Lin, J.; Chen, J.; Elenbaas, B.; Levine, A. J. *Genes Dev.* **1994**, *8*, 1235−1246.

the IC$_{50}$ for the Leu22Tyr mutant of the p53-derived sequence is twice as small as the wild-type sequence.[20] This corresponds to a −0.4 kcal/mol lowering of the binding free energy. Almost all the best binding sequences found in the ELISA assays had Tyr at position 22.[20] To study this mutation further by calculating the binding energies for the Leu22Tyr mutant, we carried out two dynamic simulations sets: one for wild-type p53 sequence and another for its Leu22Tyr mutant. The results are shown in Table 4. Interestingly, the $\Delta\Delta G_{subtotal} = -0.65$ kcal/mol favors the mutant, consistent with experiment, although this qualitative agreement with experiment is fortuitous, given the precision that can be achieved with the method we used to calculate the binding free energies.

The Leu22Tyr p53 mutation was never chosen during evolution (Figure 1B). The close analysis of the Leu22Tyr p53 mutant structure reveals that the Tyr-22 phenol hydrogen can

Figure 5. The evolution of the hydrogen bond distances between heavy atoms during dynamics simulations of the Mdm2 complex with the wild-type p53-based peptide (A) and that of the p53 Leu22Tyr mutant peptide (B). (C) The same is for the intact p53-derived Leu22Tyr mutant peptide sampled alone in a water box.

participate in hydrogen bonding to at least one of the carboxylate oxygens of the Glu-17 side chain (Figure 4B). This bond helps to stabilize the helicity of the intact 12-residue p53-derived peptide. However, it introduces an additional strain in the bound conformation in the p53−Mdm2 complex, and it disrupts the two hydrogen bonds which exist between the p53 Phe-19 backbone amide hydrogen and the xMdm2 Glu-68 side chain oxygen (Glu-72 in hMdm2) and the hydrogen at p53 Trp-23 ring $N_{\epsilon 1}$ and the xMdm2 Ile-50 backbone carbonyl (Leu-54 in hMdm2; Figure 4A). An analysis of the dynamic trajectories does show these two opposing effects of the Leu22Tyr mutation (Figure 5). Figure 5 demonstrates that the hydrogen bond exists almost all the time between the p53 Tyr-22 hydroxyl and one of the p53 Glu-17 side chain oxygens during the simulations of both the complex and the intact peptide. The two important hydrogen bonds between p53 peptide and Mdm2 are broken and lost several times during the dynamic simulations for the complex of p53 Leu22Tyr mutant with Mdm2. As a result, the p53 Leu22Tyr mutant shows more stable trajectories for the small peptide in water (Figure 6B), but the wild-type complex trajectory is more stable than that of the mutant (Figure 6A). These two opposing effects could result in different outcomes for the small peptides and for the *N*-terminal stretch of the bigger protein p53 when they bind to Mdm2. In the case of a small peptide which has a very disordered and undefined structure in solution, the additional one or two hydrogen bonds could help to preserved helicity and should contribute to the favorable binding, as has been seen in ELISA assays.[20] However, when the *N*-terminal stretch of the p53 protein, which probably has a more defined structure, binds to the active site of Mdm2, the disruption of the two important hydrogen bonds between p53 and Mdm2 could result in smaller binding energies. As has been demonstrated in this work, the hydrogen bond to the Trp-23 side chain amine adds ~4.8 kcal/mol to the total binding free
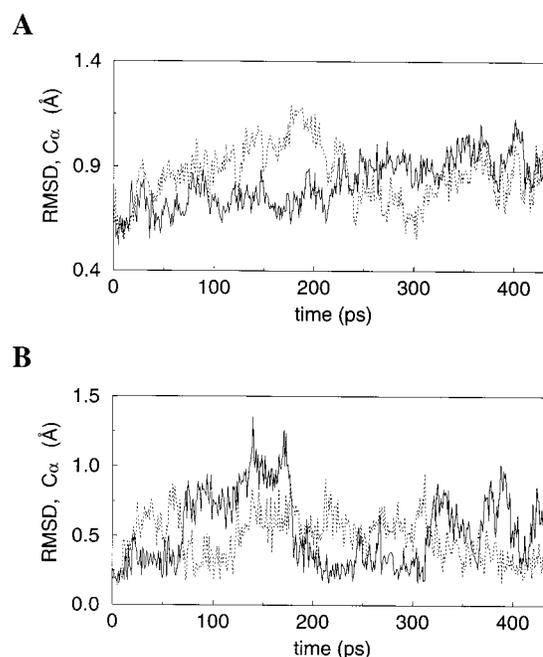
Figure 6. (A) Root-mean-square deviations along the trajectories for C_{α} of the Mdm2 complex with the wild-type p53-derived peptide (solid lines) and p53 Leu22Tyr mutant (dotted lines). (B) Root-mean-square deviations along the trajectories for C_{α} of 12-residue p53 sequence-based peptide sampled separately in a water box. The wild-type p53-derived peptide trajectory is depicted in solid lines and p53 Leu22Tyr mutant is shown in dotted lines.

energy. That corresponds to the more than 3 orders of magnitude in increase of the effective binding concentrations—a highly unfavorable outcome. Therefore, the Leu22Tyr mutation was never selected during the evolution.

D. Comparative Study of the PROFEC Method and Computational Methyl-to-Hydrogen-Scanning Mutagenesis Experiments for the Selected Positions at the Phenyl Ring of p53 Trp-23. Having two trajectory sets for the complexes of Mdm2 with the wild-type p53 and that of p53 Trp23Ala mutant, we decided to check the robustness of the PROFEC software.[30] The PROFEC program produces the three-dimensional maps which indicate the areas where the increase or decrease of the substrate steric volume or electrostatic interactions would contribute to the binding energy between a ligand and a receptor. Therefore, we generated the maps for the trajectory sets collected for the complexes of the wild type (Figure 7A) and the Trp23Ala mutant (parts B and C of Figure 7) and their peptides. Parts A−C of Figure 7 demonstrate the predictive power of the PROFEC method showing the cavity near Ala-23 of the p53 mutant, large enough to accommodate such a bulky group as Trp (parts B and C of Figure 7). In contrast, the map for the wild-type leaves little space for improvement, though contours show that there may be a small volume that can be grown near carbons $C_{\zeta 2}$ and $C_{\eta 2}$ of the phenyl ring of Trp-23. To investigate this issue further, we collected three trajectories and analyzed the binding free energies for the methylated Trp at three positions: $C_{\zeta 2}$, $C_{\eta 2}$, and $C_{\epsilon 3}$. Position $C_{\epsilon 3}$ has been chosen as a control because the PROFEC maps demonstrated that growth of a particle there would produce highly unfavorable van der Waals interactions resulting in decreased binding. All energy terms for the methylated species are shown in Tables 10−12 (Supporting Information). Table 5A gives the binding energy terms, and the differences in

(30) Radmer, R. J.; Kollman, P. A. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 215−227.
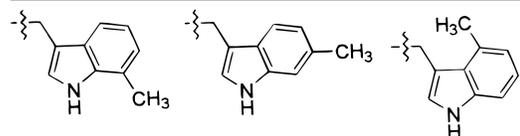
Table 5

(A) Binding Free Energies for xMdm2 Complex with Ac-QETFSDL(mW)KLLP-NH$_2$ Peptide, whereas mW Is a Trp-23 Methylated at C$_{\zeta2}$, C$_{\eta2}$, and C$_{\epsilon3}$
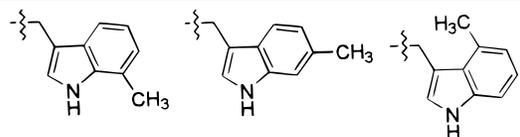
contribution	mean	std	mean	std	mean	std
$\Delta E_{\text{electrostatic}}$	−72.55	35.51	−85.16	28.59	−47.08	26.25
ΔE_{vdW}	−51.69	3.40	−56.97	3.95	−44.50	4.39
ΔE_{gas}	−124.24	35.26	−142.13	29.53	−91.58	27.91
ΔG_{PB}	96.55	35.76	116.36	29.56	68.81	26.35
$\Delta G_{\text{nonpolar}}$	−6.94	0.23	−7.66	0.41	−6.28	0.35
$\Delta G_{\text{subtotal}}$	−34.63	4.13	−33.44	3.74	−29.05	3.57

(B) Wild-Type Binding Energies Calculated from the Modified Trajectory of the Methylated Trp-23 Species (Methyl-to-Hydrogen Substitution)

trajectory set

contribution	mean	std	mean	std	mean	std
$\Delta E_{\text{electrostatic}}$	−72.44	35.51	−85.10	28.59	−47.19	26.20
ΔE_{vdW}	−48.62	3.32	−54.12	4.09	−43.63	4.43
ΔE_{gas}	−121.06	35.12	−139.23	29.65	−90.83	27.81
ΔG_{PB}	96.69	35.82	116.13	29.52	69.07	26.25
$\Delta G_{\text{nonpolar}}$	−6.84	0.24	−7.51	0.40	−6.33	0.37
$\Delta G_{\text{subtotal}}$	−31.21	4.10	−30.61	3.85	−28.09	3.68

energies are reported in parts A and B of Table 6. The results in Table 6 represent methyl-to-hydrogen-scanning mutagenesis experiments. Negative numbers in $\Delta\Delta G_{\text{subtotal}}$ mean a favorable increase in binding when the position is methylated. In contrast, positive $\Delta\Delta G_{\text{subtotal}}$ indicates the preference for the hydrogen atom (wild-type residue) at the indicated position of the Trp-23 residue. The data in Tables 6B is in excellent agreement with the PROFEC map prediction that the methylation at C$_{\zeta2}$ and C$_{\eta2}$ positions provides better van der Waals interactions resulting in tighter binding, in contrast to addition of a methyl group at the C$_{\epsilon3}$ position which results in highly unfavorable van der Waals contacts. The methylation of Trp-23 at C$_{\epsilon3}$ gives an increase of 3.9 kcal/mol in the binding free energies, opposite to the methylation at C$_{\zeta2}$ and C$_{\eta2}$, where the former is the more favorable with a gain of −1.7 kcal/mol in interaction (Table 6B). The higher standard deviations in Table 6B are due to the fact that the mutant and wild-type energy terms are collected by independent measurements (from different sets of trajectories). Data in Table 6A represents the computational methyl-to-hydrogen-scanning mutagenesis results on the same trajectory set. The trend in the data is consistent with the findings reported in Table 6B and the PROFEC maps. However, the methylation at C$_{\epsilon3}$ shows about 1 kcal/mol more favorable binding energy than that of the wild-type Trp-23 sequence when all energy terms are calculated from the same trajectory set. Still, the gain for the C$_{\epsilon3}$ is the smallest among the three considered methylation sites. This discrepancy with the PROFEC method prediction shows some limitations of the use of the structures for all species from the same sampling set collected for the complex. Interestingly, the free energies of the wild-type enzyme produced from the modified trajectory sets collected for complexes of the Trp-23 C$_{\zeta2}$ and C$_{\eta2}$ methylated species are

comparable to the values for the trajectory set of the wild type (Tables 5B and 1C; $\Delta G_{\text{subtotal}}$[wild type from C$_{\zeta2}$−CH$_3$ mutant] = −31.21 ± 4.10 kcal/mol, $\Delta G_{\text{subtotal}}$[wild type from C$_{\eta2}$−CH$_3$ mutant] = −30.61 ± 3.85 kcal/mol, and $\Delta G_{\text{subtotal}}$[wild type] = −32.93 ± 4.11 kcal/mol). These are two other examples supporting an assumption that small alterations do not introduce major changes to the sampling and the same sampling set can be used for the analysis in many cases. However, the caution should be taken when a mutation results in the sizable distortions to the binding interface. Thus, the binding free energy for the native complex produced from the Trp-23 C$_{\epsilon3}$-methylated complex trajectory ($\Delta G_{\text{subtotal}}$[wild type from C$_{\epsilon3}$−CH$_3$ mutant] = −28.09 ± 3.68 kcal/mol) differs by almost 5 kcal/mol from that of the wild-type trajectory ($\Delta G_{\text{subtotal}}$[wild type] = −32.93 ± 4.11 kcal/mol). This last observation is just another indication that C$_{\epsilon3}$ methylation introduces some changes to the sampling and its modified trajectory cannot be used to evaluate the binding energies for the wild type. Nevertheless, parts A and B of Table 6 are similar in their estimate of the changes in the binding free energies for the methylation at these three positions, ranking them in the same order in both analyses, C$_{\zeta2}$, C$_{\eta2}$, and C$_{\epsilon3}$, from the best to the worst, respectively.

E. Binding Energy Evaluation and Computational Alanine-Scanning Mutagenesis Experiments on the Human Complex. We also conducted binding energy evaluation and the computational alanine-scanning mutagenesis experiments for the human complex. Table 7 shows all energy components of the total binding energy for hMdm2 which are given in Table 8. We also looked into the trajectory stability and the convergence issues by calculating the binding free energies for the snapshots collected during the first and second halves of the dynamics simulations (Table 8B). The standard deviations for the energy terms are smaller for the second period, indicating better convergence with the binding free energy term drifting about 0.6 kcal/mol between the first and second 200 ps periods of the sampling with the both values differing by only 0.3 kcal/ mol for the binding free energy reported for the entire 400 ps collection stretch. Standard deviations are slightly decreased when the sample size was increased from 100 to 200 snapshots (processed each 2 ps). In contrast, the $\Delta G_{\text{subtotal}}$ standard deviation is increased for second period of the simulations. This is simply a consequence of the opposing changes in the individual energy terms. Nevertheless, the binding free energies in both parts A and B of Table 8 are very similar and within the standard deviation, displaying a well equilibrated and stable system with the reasonable convergence over a dynamics simulation. The alanine-scanning results are given in Table 9 and plotted as a graph in Figure 3C. The data in Tables 7−9 are very similar to the values for the xMdm2 complex (Tables 1 and 2), justifying the use of the homologue complex when a better crystallographic structure was available. The numbers in Table 9 are also in excellent agreement with experiment. Thus, the periodicity in peaks in Figure 3C reflects the amphiphilic nature of the interactions and closely traces alterations in the replacement data for the experiment. Even the positive effect of the replacement with alanine (peaks for positions 20 and 24) become more obvious when the native p53 sequence is compared to the sequence of the best binding peptide chosen by ELISA assays.[20] The hydrophobic methionine residue is found in the best binding peptide at position 20 vs the polar Ser-20 for the native sequence, and the negatively charged glutamic acid has been selected in ELISA assays vs positively charged Lys-24 of p53. The favorable contributions to the binding free energy for Ala substitutions at positions 20 and
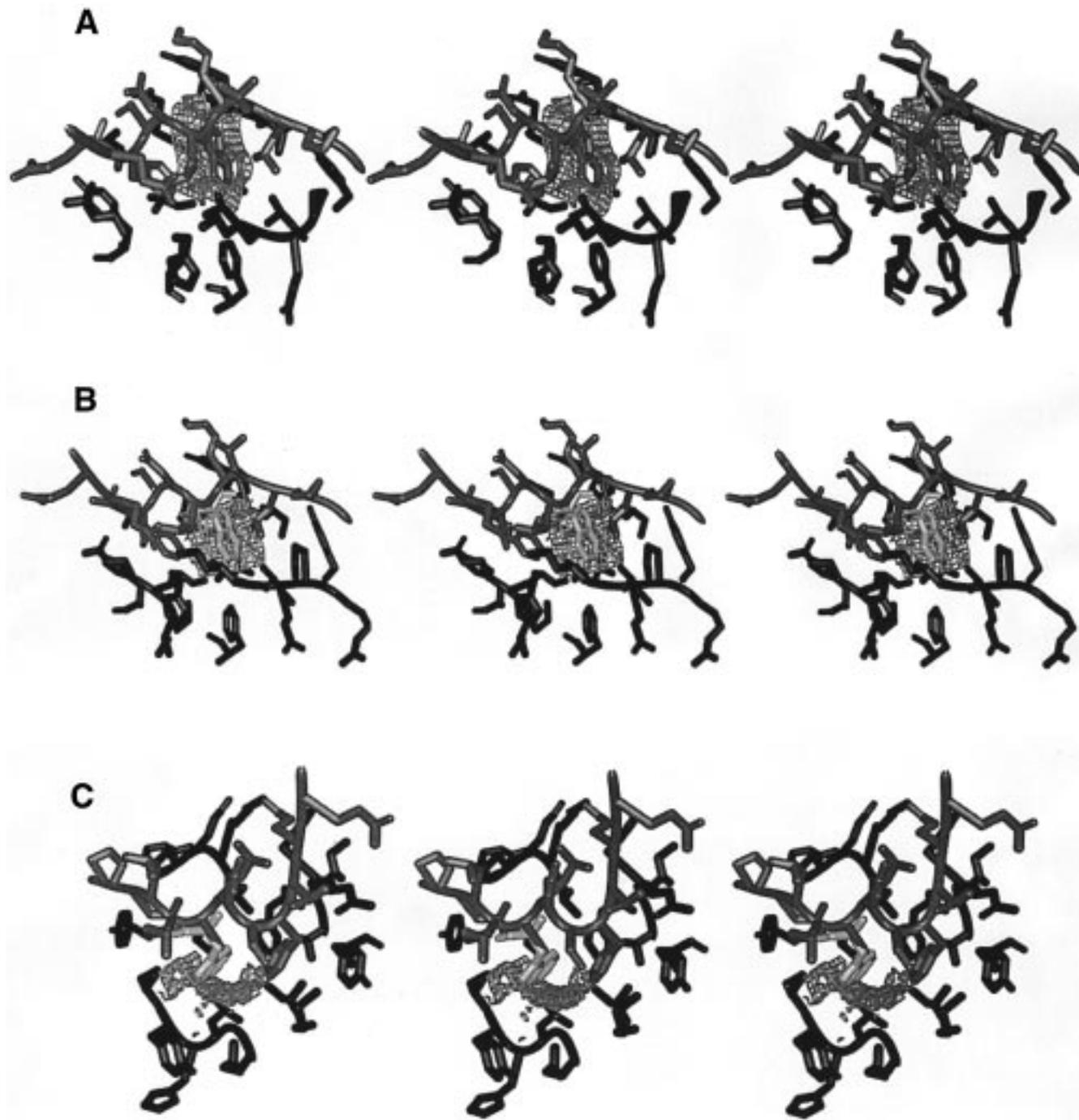
Figure 7. Stereoview of the PROFEC maps for the complex of xMdm2 (black) bound to p53-derived peptide (dark gray) for the (A) wild-type simulations and (B) for the p53 Trp23Ala peptide mutant. Part C is a different view of part B. The light gray residue in parts B and C represents a Trp in the orientation as in the wild-type structure overlapped on the Ala-23 residue of the mutant. The two left structures could be viewed with stereoglasses and the right combination represents the crossed eye projections.

24 are a result of the $\Delta\Delta G_{PB}$ not being totally balanced by $\Delta\Delta G_{electrostatic}$ and, in addition, by the small term of $\Delta\Delta G_{vdW}$ for Ser20Ala mutation (Table 9). These substitutions should contribute favorably to the binding.

F. Effect of the Charge and Ions on the Free Energy of Binding. Another aspect of the interactions is the total positive charge of $+4e$ on the hMdm2 complex and its effect on the binding free energies. The influence of the charge on the interacting species is particularly interesting because, although the total charge of the p53-binding domain of hMdm2 is $+5e$, the entire sequence of hMdm2 has 40 more negatively charged residues than positive ones and the *N*-terminal sequence of p53

is rich with negatively charged residues. To study the effect of the charges, we carried out an additional dynamics simulation for hMdm2 complex surrounded by counterions to achieve the total neutrality of the system. Five chloride anions and one sodium cation have been added to the hMdm2 complex with p53-derived peptide at the distance not closer than 15 Å to the peptide to eliminate the effect of long-range electrostatics on the binding interface. Some of the crystallographic water present in xMdm2 structure was added to the hMdm2 complex to fill the internal cavities. The entire system was solvated with a box of water molecules, and dynamics simulations were carried out as has been described in the Methods section. The computational

Massova and Kollman

Table 6. Computational Methyl-to-Hydrogen-Scanning Mutagenesis Results for xMdm2 Complex with a 12-Residue p53-Based Peptide Mutants of Trp-23 Methylated at $C_{\zeta 2}$, $C_{\eta 2}$, and $C_{\epsilon 3}$ ($\Delta\Delta = \Delta_{\text{mutant}} - \Delta_{\text{wild type}}$)

(A) Mutant and Wild-Type Energies Are from the Same Trajectory of the Methylated Trp-23 Species
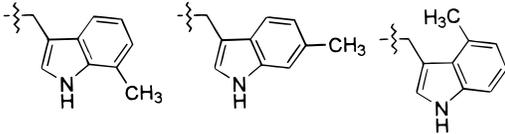
contribution	mean	std	mean	std	mean	std
$\Delta\Delta E_{\text{electrostatic}}$	−0.11	0.16	−0.06	0.10	0.12	0.13
$\Delta\Delta E_{\text{vdW}}$	−3.07	0.71	−2.85	0.65	−0.87	0.48
$\Delta\Delta E_{\text{gas}}$	−3.18	0.73	−2.91	0.65	−0.76	0.50
$\Delta\Delta G_{\text{PB}}$	−0.14	0.61	0.23	0.49	−0.26	0.66
$\Delta\Delta G_{\text{nonpolar}}$	−0.10	0.07	−0.15	0.08	0.05	0.16
$\Delta\Delta G_{\text{subtotal}}$	−3.42	0.92	−2.83	0.77	−0.97	0.77

(B) Mutant and Wild-Type Energies Are from Different Trajectories
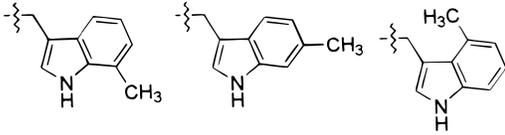
contribution	mean	std	mean	std	mean	std
$\Delta\Delta E_{\text{electrostatic}}$	28.79	47.32	16.17	42.37	54.25	40.83
$\Delta\Delta E_{\text{vdW}}$	3.53	4.69	−1.75	5.10	10.72	5.45
$\Delta\Delta E_{\text{gas}}$	32.31	47.68	14.42	43.61	64.97	42.53
$\Delta\Delta G_{\text{PB}}$	−34.45	48.73	−14.63	44.39	−62.19	42.32
$\Delta\Delta G_{\text{nonpolar}}$	0.43	0.44	−0.30	0.56	1.09	0.52
$\Delta\Delta G_{\text{subtotal}}$	−1.70	5.83	−0.51	5.56	3.87	5.44

Table 7. Energy Terms for hMdm2 Complex with Ac-QETFSDLWKLLP-NH$_2$ Peptide

	complex		protein		ligand	
contribution	mean	std	mean	std	mean	std
(A) Protocol 1: Complex, Protein, and Peptide Structures Are from the Same Trajectory						
$E_{\text{electrostatic}}$	−2875.39	46.37	−2517.71	32.49	−255.07	15.21
E_{vdW}	−385.54	16.22	−313.07	15.44	−14.63	4.59
E_{internal}	1688.14	26.34	1466.59	24.41	221.55	9.72
E_{gas}	−1572.79	53.14	−1364.19	37.15	−48.16	17.79
G_{PB}	−1731.64	49.91	−1577.72	34.68	−287.70	15.89
G_{nonpolar}	35.56	1.12	32.76	0.93	10.29	0.18
G_{subtotal}	−3268.86	21.08	−2909.15	21.15	−325.57	9.79
(B) Protocol 2: Complex, Protein, and Peptide Structures Are from Separate Trajectories						
$E_{\text{electrostatic}}$	−2875.39	46.37	−2409.64	64.91	−222.19	20.30
E_{vdW}	−385.54	16.22	−309.55	13.12	−14.44	5.55
E_{internal}	1688.14	26.34	1472.74	19.90	213.01	8.30
E_{gas}	−1572.79	53.14	−1246.45	66.46	−23.62	21.86
G_{PB}	−1731.64	49.91	−1689.74	66.17	−318.97	19.53
G_{nonpolar}	35.56	1.12	32.58	0.64	10.30	0.32
G_{subtotal}	−3268.86	21.08	−2903.60	20.02	−332.29	7.33

Table 8. Binding Free Energies for hMdm2 Complex with Ac-QETFSDLWKLLP-NH$_2$ Peptide

(A) 100 Snapshots: 1 Snapshot per Each 4 ps

	protocol 1		protocol 2	
contribution	mean	std	mean	std
$\Delta E_{\text{electrostatic}}$	−102.59	20.16	−243.56	82.31
ΔE_{vdW}	−57.85	3.94	−61.56	21.59
ΔE_{gas}	−160.44	20.37	−302.72	87.86
ΔG_{PB}	133.79	20.16	277.07	85.15
$\Delta G_{\text{nonpolar}}$	−7.49	0.36	−7.32	1.33
$\Delta G_{\text{subtotal}}$	−34.14	3.66	−32.97	29.98

(B) All Data Are Collected and Processed According to Protocol 1

	50 first snapshots (1 per each 4 ps)		50 last snapshots (1 per each 4 ps)		200 snapshots (1 per each 2 ps)	
contribution	mean	std	mean	std	mean	std
$\Delta E_{\text{electrostatic}}$	−109.47	20.53	−95.70	17.39	−102.11	19.83
ΔE_{vdW}	−57.58	4.36	−58.12	3.49	−57.80	3.85
ΔE_{gas}	−167.05	21.21	−153.82	17.30	−159.90	20.15
ΔG_{PB}	140.22	20.60	127.36	17.67	133.18	20.15
$\Delta G_{\text{non−polar}}$	−7.59	0.35	−7.39	0.35	−7.50	0.39
$\Delta G_{\text{subtotal}}$	−34.43	3.11	−33.86	4.14	−34.22	3.34

Table 9. Computational Alanine-Scanning Mutagenesis Results for hMdm2 Complex with a 12-Residue Stretch of p53 ($\Delta\Delta = \Delta_{\text{wildtype}} - \Delta_{\text{mutant}}$)

	Gln16Ala		Glu17Ala		Thr18Ala		Phe19Ala	
contribution	mean	std	mean	std	mean	std	mean	std
$\Delta\Delta E_{\text{electrostatic}}$	−0.94	4.89	−76.32	14.18	−1.51	0.56	−1.23	0.54
$\Delta\Delta E_{\text{vdW}}$	−1.18	1.13	−0.88	0.56	−0.39	0.15	−9.20	0.80
$\Delta\Delta E_{\text{gas}}$	−2.12	4.77	−77.19	14.19	−1.90	0.60	−10.44	0.85
$\Delta\Delta G_{\text{PB}}$	2.43	5.01	77.18	14.27	1.84	0.70	7.86	1.12
$\Delta\Delta G_{\text{nonpolar}}$	−0.23	0.23	−0.09	0.15	0.01	0.19	−0.13	0.10
$\Delta\Delta G_{\text{subtotal}}$	0.08	0.81	−0.10	1.03	−0.04	0.39	**−2.71**	1.31

	Ser20Ala		Asp21Ala		Leu22Ala		Trp23Ala	
contribution	mean	std	mean	std	mean	std	mean	std
$\Delta\Delta E_{\text{electrostatic}}$	−0.21	0.96	−68.51	3.66	0.30	0.29	−5.21	1.67
$\Delta\Delta E_{\text{vdW}}$	−0.37	0.32	−0.10	0.01	−4.08	0.97	−11.04	1.20
$\Delta\Delta E_{\text{gas}}$	−0.58	0.99	−68.61	3.66	−3.78	1.01	−16.25	1.92
$\Delta\Delta G_{\text{PB}}$	1.33	1.00	68.13	3.58	2.51	1.13	11.42	1.39
$\Delta\Delta G_{\text{nonpolar}}$	−0.03	0.13	−0.00	0.11	−0.24	0.16	−0.72	0.09
$\Delta\Delta G_{\text{subtotal}}$	0.72	0.61	−0.48	0.39	**−1.51**	1.30	**−5.54**	1.43

	Lys24Ala		Leu25Ala		Leu26Ala	
contribution	mean	std	mean	std	mean	std
$\Delta\Delta E_{\text{electrostatic}}$	76.81	3.89	−0.06	0.30	0.22	0.22
$\Delta\Delta E_{\text{vdW}}$	−0.15	0.06	−0.60	0.44	−5.14	0.74
$\Delta\Delta E_{\text{gas}}$	76.67	3.87	−0.660	0.36	−4.92	0.77
$\Delta\Delta G_{\text{PB}}$	−76.38	3.94	0.70	0.50	3.28	1.05
$\Delta\Delta G_{\text{nonpolar}}$	−0.01	0.08	−0.05	0.19	−0.08	0.14
$\Delta\Delta G_{\text{subtotal}}$	0.28	0.55	−0.01	0.52	**−1.71**	1.26

alanine-scanning mutagenesis analysis has been done on the collected trajectory while keeping all interacting components of the complex neutral. The results for the binding free energy of the complex are shown in Table 13 (Supporting Information). Because the counterions were kept during the free energy calculations, all terms have a contribution from the interaction between the counterions during complex formation. The data in Table 13 (Supporting Information) displays that the energy terms exhibit more noise, as standard deviations in the table are larger than those in Tables 1, 7, and 8. This is a direct consequence of the incorporation of counterions in the energy calculations. Particularly affected are the terms of an electrostatic origin: $E_{\text{electrostatic}}$ and G_{PB}. Nevertheless, the contribution from the counterions to the binding free energy terms and the total noise turn out not to have a significant effect due to cancellation of terms. The binding energies determined by this approach should be comparable to experiment because under physiological conditions all interacting molecules exist as neutral species surrounded by counterions, and the measured effective binding constants should in principle correspond to the numbers shown in Table 13B (Supporting Information). The standard deviations in Table 13B are about 30% larger than those in Tables 1C and 8. The alanine scanning results for the hMdm2-p53 complex surrounded by counterions are shown in Table 14 (Supporting Information). The energy terms related to the interactions with

the counterions cancel for all contributions except those introduced by the difference in interactions of the counterions with the original and mutated residue, which should be small because the counterions were added at the distances farther than 15 Å from the peptide. Interestingly, the numbers in Table 14 are in excellent agreement with the experiment and with the data from the computational experiments for xMdm2. Though addition of the counterions to the system brings extra noise, it cancels during the data processing. Therefore, the scanning mutagenesis results are essentially identical for calculations with and without explicitly added counterions (Tables 9 and 14). This test demonstrated that the effect of the remote groups on the binding free energies is negligible. The mutagenesis results are mostly insensitive to remote charges on the interacting molecules and our computational method can be used to produce the reliable data for the cases when the assignment of the protonated states or charged groups is controversial or when parts of the structures remote from the binding interface are not seen in crystallographic maps due to mobility or absence of the electron density.

Conclusions

The development of a computational method which combines explicit molecular mechanical energies and continuum solvation models for calculating protein−peptide interaction free energies has been presented here and proven to be successful in reproducing the experimental tendencies, although difficulties still exist in evaluation of some terms (conformational entropy) and in the large standard deviations. In our view, the least satisfactory aspect in the evaluation of the absolute free energy of complexation between protein and peptide is the application of normal-mode analyses to estimate the change in solute entropy upon complexation. In computational alanine scanning, this solute entropy term is not likely to be critical for the estimates of the relative changes in the binding free energies during alanine scanning because of similar contributions of the $-T\Delta S$ term to the free energy of the binding for the wild type and its mutants. The method reasonably relies on the resemblance of the structures and the assumption about adequacy of the conformational sampling.

The computational experiments with the Trp23Ala mutation and the fact that the $\Delta G_{\text{subtotal}}$ ($\Delta G_{\text{binding}}$ without the TS terms) are very similar for this protein when calculated from either the mutant or the wild-type trajectory have demonstrated that the method can work in a number of cases. Alanine scanning involves substitution of the larger group with the smaller one; however, it will likely be much more difficult to add a larger group than those simulated because of highly unfavorable van der Waals interactions due to atomic position overlap. Thus, future challenges are applications of the scanning mutagenesis experiments considering larger amino acids with both natural and unnatural side chains.

With the advances in molecular biology techniques, surface mutagenesis methods have become widely used to identify the hot spots in binding interfaces. These methods are elaborate and time consuming. The novel approach reported here provides a fast computational mutagenesis method which could be used for computationally inexpensive screening of a vast variety of possible modifications to the binding site and ligands. One could imagine simpler methods that merely employ minimization of different mutants starting with the crystal structure,[31] but these are likely to lead to less reliable and robust free energy estimates than the method we have proposed here. The outcome of such search is a ligand or a target optimized for specific interactions. The major appeal of this technique is that it can be used to evaluate the binding free energies and predict the favorable modifications to interfaces between any interacting molecules to improve binding or stability of the complex.

We have used this approach to study the interactions between the tumor suppressor protein p53 and oncoprotein Mdm2. The method has demonstrated an excellent predictive power and agreement with the experimental data. Interesting insights have been gained into the details of interaction of these important proteins. The success of our computational surface mutagenesis method relied on achieving reliable aqueous trajectories for the system, which has been aided by recent advances in molecular mechanics force fields and calculations of long-range electrostatics by the PME method. The use of continuum solvent models has been essential to accurately estimate the solvation free energies.

In this paper, we have shown that computational alanine scanning can reproduce the qualitative result from amino acid replacement experiments on which residues are essential for the binding. The next logical step is to apply this computational approach to study the "classic" application of experimental alanine scanning, the binding of growth hormone to its receptor.[32] In this case, extensive quantitative free energy data is available for the effect of mutating various residues in both receptor and hormone to alanine.[26,27] In addition to these studies of protein−protein interactions, the computational alanine-scanning mutagenesis approach described here could find important applications in investigation of protein stability, which would necessitate the simulation of a model of the denatured as well as the native states. Finally, it is our view that the method described here for alanine scanning can be extended to considering many other mutations and conformations of either macromolecules (proteins, nucleic acids, lipids, or saccharides) or ligands to optimize the interacting components for binding and stability.

Supporting Information Available: Tables 10−14 giving energy terms and binding free energy data as well as alanine scanning results (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA990935J

(31) Reyes, C. M.; Kollman, P. A. *J. Mol. Biol.* Submitted for publication.
(32) Huo, S.; Massova, I.; Kollman, P. A. Studies in progress.