doi:10.1016/j.jmb.2005.12.083 J. Mol. Biol. (2006) 357, 1039–1049

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

# Kinetic Computational Alanine Scanning: Application to p53 Oligomerization

**Lillian T. Chong[1,2], William C. Swope[2]*, Jed W. Pitera[2] and Vijay S. Pande[1]***

[1]*Departments of Chemistry Structural Biology, Stanford University, Stanford, CA 94305-5447, USA*

[2]*IBM Almaden Research Center 650 Harry Road, San Jose CA 95120, USA*

**Corresponding author*

We have developed a novel computational alanine scanning approach that involves analysis of ensemble unfolding kinetics at high temperature to identify residues that are critical for the stability of a given protein. This approach has been applied to dimerization of the oligomerization domain (residues 326–355) of tumor suppressor p53. As validated by experimental results, our approach has reasonable success in identifying deleterious mutations, including mutations that have been linked to cancer. We discuss a method for determining the effect of mutations on the location of the dimerization transition state.



*Keywords:* tumor suppressor p53; oligomerization; alanine scanning; ensemble unfolding kinetics; molecular dynamics simulation

## Introduction

Tumor suppressor p53 is a "gatekeeper" of the genome, functioning at the center of a network of biological pathways that guard the cell from potential cancer. As a multidomain transcription factor, p53 is comprised of an N-terminal activation domain, a central DNA-binding domain, an oligomerization domain, and a C-terminal regulatory domain. The oligomerization domain enables p53 to adopt its biologically active tetrameric form. More than half of human cancers result from mutations in the p53 gene.[1] Although most of these mutations are in the DNA-binding domain, several studies have identified mutations within the oligomerization domain (p53tet) that are linked to increased incidence of cancer.[2–5] More cancer-associated mutations may yet be identified in p53tet, due to the fact that this region of the p53 gene was not sequenced in studies that searched for mutations in the p53 DNA-binding domain.[3]

Structures of p53tet have been determined by both NMR spectroscopy (Protein Data Bank 1PES[6] and 1SAK[7]) and X-ray crystallography (Protein Data Bank 1C26[8] and 1AIE[9]). This α/β domain, which is one of the smallest known protein oligomerization domains, assembles as a dimer of dimers.[10] Each dimer consists of an antiparallel β-sheet and a pair of antiparallel helices. The two dimers associate with one another to form the active tetramer *via* an extensive hydrophobic surface that is formed by the antiparallel helices of each dimer. Based on kinetics experiments and Φ-value analysis, tetramerization of p53tet is thought to involve induced-fit associations of monomers to form the dimer intermediates followed by "lock-and-key" associations of pre-organized dimers to form the tetramers.[10] A recent study of p53 biogenesis *in vitro* has shown that p53 dimerizes cotranslationally and then forms the tetramer post-translationally.[11] Given the kinetic advantage of dimerization before the p53 chains leave the polysome, mutations that affect the kinetics could play important biological roles. Consistent with the importance of kinetics in p53 function, molecular dynamics (MD) simulations of p53tet dimerization evolving from the rate-limiting transition state ensemble have revealed a nucleation-condensation mechanism in which L330, I332, and F338 from each monomer form a folding nucleus.[12]

Because of the structural symmetry of p53, point mutations in the oligomerization domain can reduce the stability of its tetramers significantly. For example, mutation of L330H in the folding nucleus is thought to disrupt the hydrophobic core of the dimers (Figure 1(a)) and has been linked to

Abbreviations used: p53tet, p53 oligomerization domain; MD, molecular dynamics; RMSD, root-mean-squared deviation.

E-mail addresses of the corresponding authors: swope@almaden.ibm.com; pande@stanford.edu

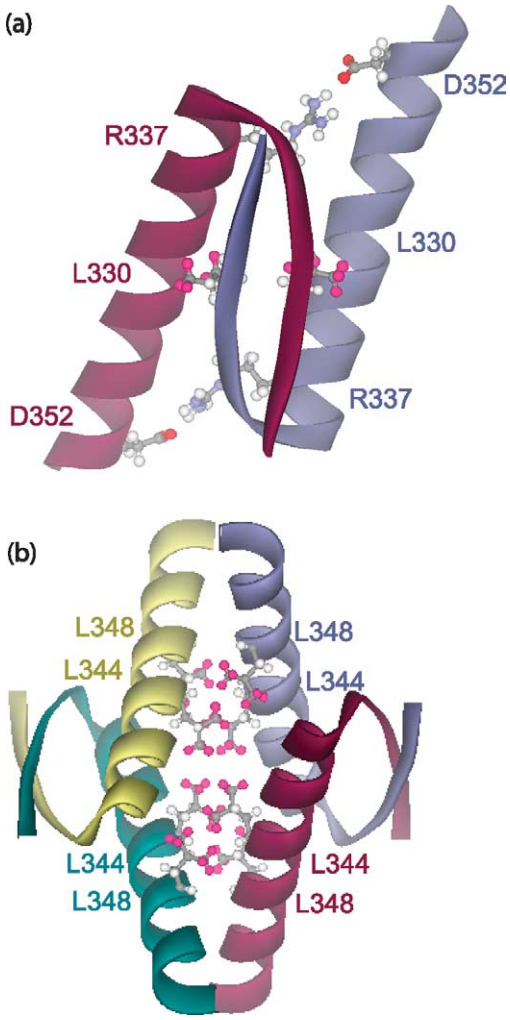


**Figure 1.** Amino acids that are mutated in the test set of mutant p53tet dimers: (a) L330 in the hydrophobic core of the dimer; R337, which forms a salt-bridge with D352 across the monomer–monomer interface; (b) L344 and L348, which are located centrally at the tetramer interface.

hepatocarcinoma and ovarian sarcoma.[1] As another example, an inherited mutation of R337C is associated with Li-Fraumeni-like syndrome (LFLS), in which the individual is predisposed to form a broad spectrum of tumors, including brain tumors, sarcomas, and adrenal cortical tumors.[3] Due to this mutation, the salt-bridges between R337 and D352 across the monomer–monomer interfaces no longer form; four such salt-bridges are lost in the tetramer (Figure 1(a)). The germline mutation of R337H also leads to loss of these salt-bridges and has been associated with pediatric adrenal cortical carcinoma (ACC).[5] Unlike many cancer-associated mutations in p53, this mutation is tumor-specific, leading only to pediatric ACC. Interestingly, malfunction of the p53 R337H mutant is pH-dependent, occurring only when the histidine residue, which has an elevated p$K_a$, is neutral as opposed to positively charged, when the mutant has near wild-type stability.[13]

Residues that are critical for the stability of the p53tet tetramer have been identified by systematically replacing nearly every residue in p53tet by alanine (alanine scanning) and experimentally measuring the effect on the stability of the tetramer.[14] The most critical of these residues are I332, L330, and F341, which lie in the hydrophobic core of the dimer; truncations of I332 in each of four monomers to alanine residues prevent folding while truncations of either L330 or F341 lead to folding at only high concentrations of protein or low temperature. Other critical residues are L344 and L348, which are located centrally at the tetramer interface; truncation of either residue leads to formation of stable dimers instead of tetramers (Figure 1(b)). To aid in the interpretation of folding kinetics data on p53tet, these mutants have been used as experimental models of the transient dimeric intermediates.[10] Residues that are strongly destabilizing when truncated ($\Delta\Delta G_u$ of 8.8–11.7 kcal/mol) lie in the periphery of the core (R337, F328, and F338) and at the tetramer interface (M340). Less critical, but still important ($\Delta\Delta G_u$ of 4.1–5.7 kcal/mol) are residues that are solvent-exposed (T329), involved in intermonomer hydrogen bonds (R333, N345, E349), or at the tetramer interface (A347).

To provide more efficient, alternative strategies for performing alanine scanning, computational approaches have been developed in recent years. The first computational alanine scanning study involved the MM-PBSA approach, in which molecular dynamics simulations with explicit water were used to generate relevant protein conformations and a continuum solvent model was applied to compute the free energies of the conformations.[15] Other computational mutagenesis studies involved the application of energy functions that were parameterized to reproduce experimentally measured changes in stability for a large database of mutations in proteins.[16,17]

While the computational mutagenesis approaches that have been developed thus far are based on thermodynamic analyses, the use of kinetic analyses can be effective as well. In particular, kinetic analyses can be used to identify thermodynamically destabilizing mutations if the mutations are also kinetically destabilizing, reducing the barriers to unfolding by destabilizing primarily the native, folded state of the protein. Indeed, it has been found for cancer-associated mutants of the p53 DNA-binding domain that, the more unstable the mutants, the faster they unfold.[18] With a few exceptions, destabilizing mutations to alanine residues in p53tet have little effect on unfolding rates of the tetramer, since the rate-limiting transition state to tetramer unfolding, which involves unfolding to the dimer intermediate, is suggested by Φ-value analysis to closely resemble the native tetramer.[10] However, these mutations may have a greater effect on unfolding rates of the dimer intermediate, since the transition state to dimer unfolding involves a transition state

with near-zero Φ-values[10] that have been characterized by simulations as having some unstructured regions and native-like regions that exist as residual structure in the unfolded state.[12]

Here, we have probed the contributions of individual residues to the thermodynamic stability of the tetramer using a novel, kinetic computational alanine scanning approach that involves high-temperature MD simulations of dimer unfolding. Distributed computing is used to provide necessary statistical precision by running 100 independent unfolding simulations per mutant dimer. We have tested the robustness of this approach by applying it to a test set that consists of the wild-type dimer and six mutant dimers that are either cancer-associated or silent: L330H, R337C, R337H, R337HIP, L344A, and L348A where H and HIP indicate neutral and positively charged histidine, respectively. Our approach has reasonable success in identifying the main cancer-associated mutations in the test set and yields alanine scanning results that are qualitatively consistent with data from experiments.[14]

## Results

### Test set

A robust protocol for evaluating unfolding rates was developed using the test set described above (see Methods for details). For each p53tet dimer in the test set, an ensemble of unfolding pathways was generated by performing 100 independent simulations at 470 K for 10 ns. As experimental models of the p53tet dimer,[10] the L344A and L348A mutant dimers are likely to have near-wild-type stability and thus, unfolding rates similar to that of the wild-type dimer. On the other hand, the L330H, R337C, and R337H mutant dimers would be expected to have faster unfolding rates, since the mutations disrupt the monomer–monomer interface significantly, leading to a predisposition for cancer.[1,4,13] Finally, since the R337HIP mutant tetramer has near-wild-type stability,[13] the R337HIP mutant dimer is likely to have an unfolding rate that is similar to that of wild-type.

Unfolding rates were found to be sensitive to the starting conformations used for the high-temperature simulations. For example, unfolding simulations for a given dimer that start from a single, equilibrated structure yield unfolding rates that differ depending on the equilibration temperature used (300 K and 470 K in Figure 2(a) and (b), respectively). In particular, both the R337H and R337HIP mutant dimers have much faster unfolding rates relative to the other mutants when an equilibration temperature of 470 K is used instead of 300 K.

To reduce the sensitivity of unfolding rates to the initial structure, 100 independent equilibrations were performed for each dimer, each of which was then used for a single unfolding simulation. Results with this new protocol were much less

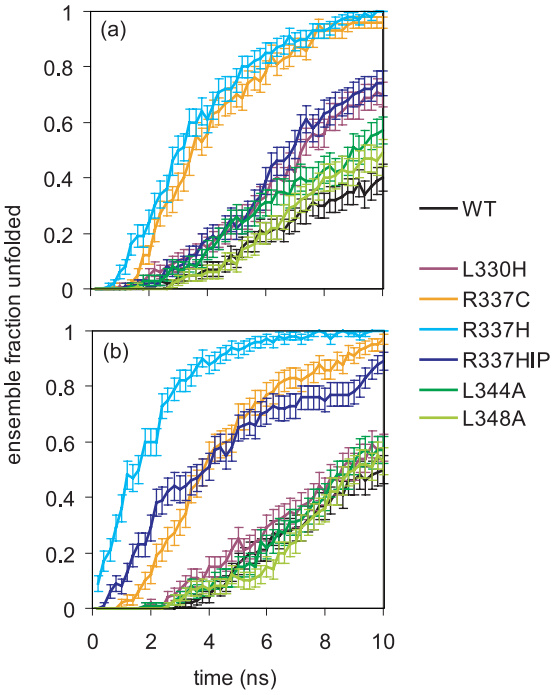


**Figure 2.** Ensemble fraction unfolded $f_u(t)$ as a function of time for each p53tet dimer in the test set evolving from a single structure that was equilibrated at either (a) 300 K or (b) 470 K. Structures are based on the initial F0 dimer model. Uncertainty estimates represent one standard deviation, as described in Methods.

sensitive to the temperature used for the equilibrations (see Figure 3(a) and (b)). Due to partial unfolding during the 1 ns equilibration stage, the unfolding simulations from conformations equilibrated at 470 K started to unfold about 1 ns earlier than simulations from conformations equilibrated at 300 K. However, the relative rates of unfolding in both sets of results are consistent. The R337C and R337H mutants unfold the most rapidly, while the L344A and L348A mutants unfold at rates similar to that of the wild-type dimer. The R337HIP and L330H mutants unfold moderately faster than wild-type. These results are in qualitative agreement with experiment, except for the fact that the L330H mutant, which has a disrupted hydrophobic core, would be expected to be thermodynamically destabilized to a significant extent and therefore unfold significantly faster than the wild-type dimer. It is possible that the thermodynamic instability of the L330H mutant does not result in kinetic instability. Nonetheless, given the robustness of results obtained by unfolding from an ensemble of conformations, the final protocol for computational alanine scanning of the p53tet dimer involves performing unfolding simulations at 470 K from an ensemble of 100 independent conformations that have been equilibrated at 300 K.

Since the model of the dimer used above, which we label the F0 dimer, is an approximate one, the final protocol was repeated for mutants based on a relaxed conformation of the dimer, the F1 dimer,

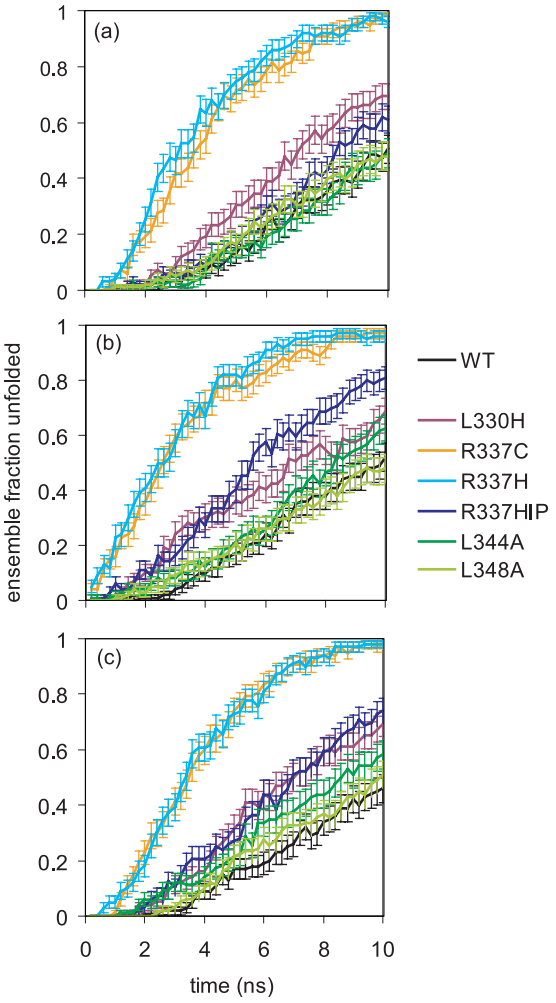


**Figure 3.** Ensemble fraction unfolded $f_u(t)$ as a function of time for each p53tet dimer in the test set evolving from 100 different structures that were obtained after performing 100 independent equilibrations at either (a) 300 K or (b) 470 K. Structures are based on the initial F0 dimer model. The relaxed F1 dimer model was used in (c), where the unfolding simulations evolved from 100 structures equilibrated at 300 K. Uncertainty estimates represent one standard deviation, as described in Methods.

which was taken from room-temperature MD simulations in explicit water that evolved from the F0 dimer.[12] Both the F0 and F1 dimer conformations are likely to be local minima in the folded dimer basin.[12] Results for the test set of mutants based on the F1 dimer (Figure 3(c)) are in agreement with those obtained using the F0 dimer (Figure 3(a)).

## Appropriate simulation lengths

In general, the ensemble unfolding data for mutant dimers in this study show three distinct phases: an initial lag phase, an intermediate non-exponential phase, and a final, single-exponential phase. The lag phase is possibly due to non-equilibrium processes of relaxation or local structural rearrangements of the dimer before unfolding. The significance of the intermediate, non-exponential phase is not clear and awaits further investigation; this phase is absent from the unfolding kinetics of some mutant dimers. Fitting the final phase of the unfolding data to a single-exponential function with a lag time enables us to extract an unfolding rate constant.

For more than half of the mutant dimers (31 in the combined test and alanine scanning sets), 10 ns unfolding simulations were sufficient for determining unfolding rate constants from the fits to a single-exponential function with a lag time. The quality of the fits was judged by visual inspection. Plots of the ensemble fraction unfolded *versus* time and corresponding fits to data from the first 10 ns and entire 20 ns of simulations are included for each mutant dimer in Supplementary Data.

As shown in Figure 4(a), the fits to the 10 ns and 20 ns simulation data for the wild-type p53 dimer were mutually consistent. However, in cases such as the E339A mutant dimer, the fits to the first 10 ns and the entire 20 ns of simulation data were significantly different. As shown in Figure 4(b), the 10 ns simulations had extended beyond the lag time, but had not progressed beyond the intermediate, non-exponential region into the later exponential region of the data that is visible from 10 to 20 ns. It is evident in this case that extending the simulations out to 20 ns was necessary to obtain a more reliable estimate of the unfolding rate constant.

The only case in which 20 ns was not sufficient is the E336A mutant dimer, which has been experimentally determined to be more stable than wild-type[14] with less than half of the ensemble unfolding after 20 ns of high-temperature simulation (Figure 4(c)). As a general rule of thumb, we found that if the ensemble fraction unfolded had not yet reached 50% or the data had not yet reached a plateau and adopted a negative curvature, then the simulation time (e.g. 10 ns) should be doubled to obtain a more reliable single-exponential fit to the data. However, since the goal was to identify destabilizing mutations, it was not necessary to extend the simulations if the mutant did not meet the aforementioned criteria long after the wild-type protein did (e.g. 20 ns). Thus, 20 ns unfolding simulations were sufficient for performing mutagenesis of the p53tet dimer.

## Alanine scanning

To determine whether a mutation is stabilizing or destabilizing based on 20 ns unfolding simulations, the effect of the mutation on the height of the barrier to unfolding was estimated as $\ln(k_{mut}/k_{wt})$. Results from our kinetic computational alanine scanning approach were validated by comparing $\ln(k_{mut}/k_{wt})$ values to thermodynamic data ($\Delta\Delta G_u$) from alanine scanning experiments that were performed using CD spectroscopy and guanidinium hydrochloride as a denaturant.[14] It is not possible to compare the computed and experimentally measured results

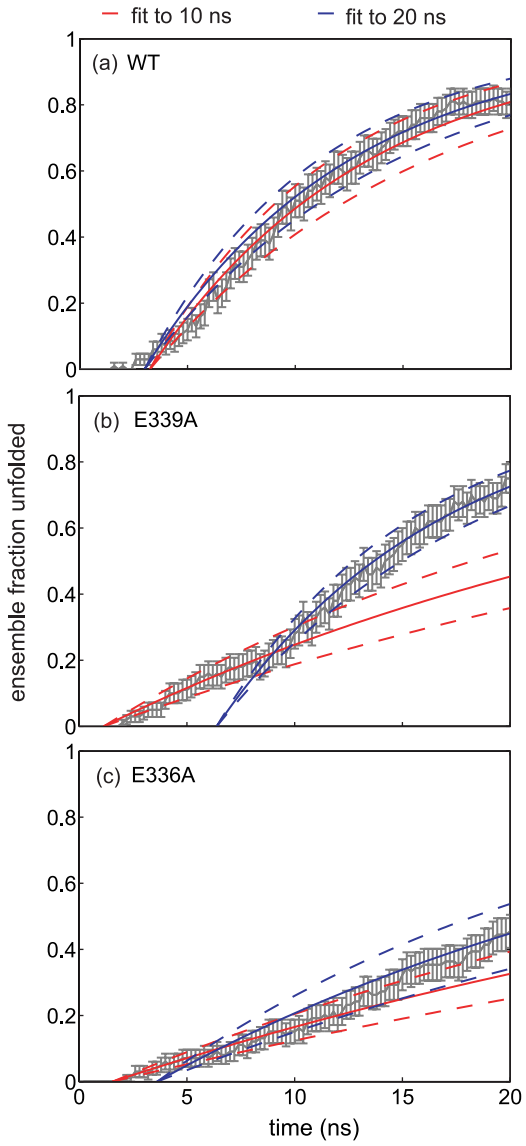


**Figure 4.** Examples of fits using a single-exponential function with a lag time determined using the first 10 ns and entire 20 ns of data from unfolding simulations of the (a) wild-type dimer, (b) E339A mutant dimer, and (c) E336A mutant dimer. Fits corresponding to average unfolding rate constants and time lag are shown as continuous lines while fits representing uncertainties (one standard deviation) in the unfolding rate constant and the average time lag are shown as broken lines. Uncertainties shown for the unfolding data represent one standard deviation.

quantitatively, due to differences in the nature of the analysis (kinetic *versus* thermodynamic), the denaturant used (thermal versus chemical), and the oligomerization state of p53tet being considered (dimer *versus* tetramer). However, as shown in Figure 5, it is clear that the computed results, using both the F0 and F1 dimer models, generally agree with the experimentally measured results in terms of the sign of the mutation-induced change in free energy when the uncertainties are taken into account. Consistent with experiment, most of the mutations were determined by our computational approach to destabilize the dimer, while two mutations, E336A and K351A, actually stabilize the dimer. The former is solvent-exposed and the latter lies at the tetramer interface. Our approach clearly distinguishes between destabilizing and stabilizing mutations.

No experimental values are available for four sets of mutations: (1) Y327A, since solvent-exposed Y327 was used as a reporter for measuring protein concentration and was therefore not mutated; (2) L330A, I332A, and F341A, due to absence of folding under the experimental conditions used; (3) L344A and L348A, due to formation of dimers instead of tetramers; and (4) Q354A, which lies beyond the range of residues evaluated.[14] Our results suggest that Y327A, L330A, and I332A destabilize the dimer. In the case of F341A, definitive conclusions cannot be drawn from the simulations, since the mutation is predicted to be stabilizing by computations based on the F0 dimer and destabilizing by computations based on the F1 dimer. The L344A and L348A mutations are predicted to be moderately destabilizing, while Q354A has no effect on the stability of the dimer.

## Strongly *versus* moderately destabilizing mutations

As discussed in Introduction, mutations that were determined by alanine scanning experiments to destabilize the tetramer significantly are at the tetramer interface (L344A, L348A, M340A, A347G), in the dimer core (L330A, I332A, and F341A), in the periphery of the dimer core (R337A, F328A, F338A), exposed to solvent (T329A), and otherwise involved in intermonomer, water-mediated hydrogen bonds (R333A, N345A, and E349A).[14] In addition, cancer-associated L330H, R337C, and R337H mutations that inactivate p53 function would be expected to be strongly destabilizing.[1,4,13] To distinguish between mutations that kinetically destabilize the dimer to a strong or moderate extent using our simulation data, unfolding rate constants computed for both F0 and F1 mutant dimers are plotted in descending order according to the values for the F0 mutant dimers (Figure 6). The most kinetically destabilizing mutations, which result in unfolding rate constants that are distinct from that of wild-type, are R333A, R337A, R337H, R337C, L330A, F328A, and E349A. All of these mutations have been determined by experiment to destabilize the tetramer thermodynamically to a significant extent.[14]

While the L344A and L348A mutations were experimentally determined to be strongly destabilizing due to the fact that the mutations lead to the formation of dimers instead of tetramers, these mutations would not be expected to strongly destabilize the dimers. Thus, the fact that the unfolding rate constants for the L344A and L348A mutant dimers are similar to that of the wild-type dimer is consistent with experiment. The failure to predict M340A as a detrimental mutation is likely

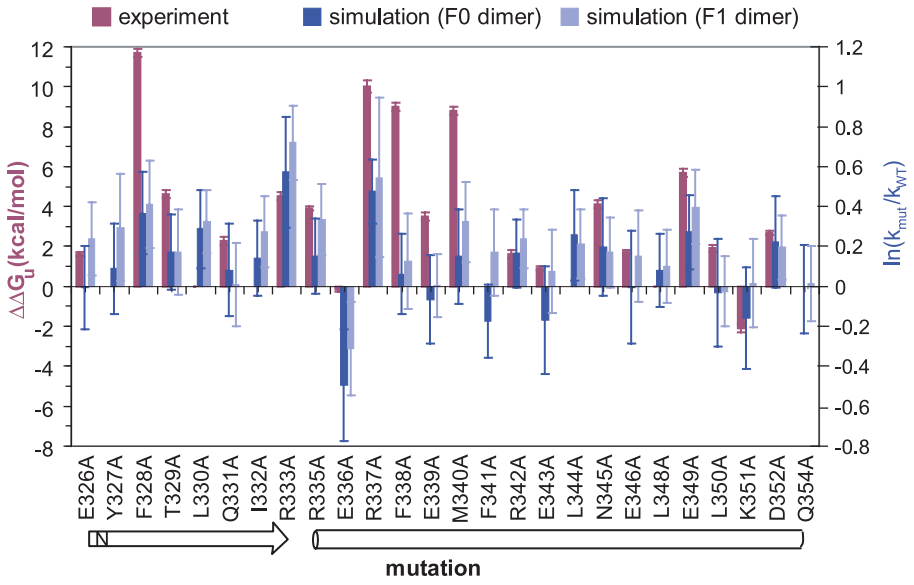


**Figure 5.** Qualitative comparison of alanine scanning results from computational *versus* experimental approaches. Computational results are from 20 ns unfolding simulations and reported as mutation-induced changes in the height of the unfolding barrier as estimated by $\ln(k_{mut}/k_{wt})$ for the p53tet dimer. Experimental results are $\Delta\Delta G_u$ values at a denaturant concentration that leads to ~50% unfolding of the p53tet tetramer.[14] Standard deviations for the simulation results were determined as described in Methods. Uncertainties shown for the experimental results are the standard errors of fitting.[14] For various reasons (see Results) no experimental measurements were available for the following mutations: Y327A, L330A, I332A, F341A, L344A, L348A, and Q354A.

due to the fact that the mutation is at the tetramer interface, destabilizing the tetramer in a way that cannot be detected by simulations of the dimer. The A347G mutation at the tetramer interface was not analyzed in our study.

The inability of our computational approach to predict I332A, F341A, F338A, N345A, T329A, and L330H as the most severely destabilizing mutations could be because these thermodynamically destabilizing mutations are not necessarily kinetically destabilizing. For example, it is conceivable that F338A and L330H, which are in the folding nucleus,[12] destabilize the transition state and folded state to similar extents, such that the resulting barriers to unfolding do not change significantly from that of the wild-type dimer. Interestingly, L330H, unlike L330A, was not found to be strongly kinetically destabilizing. While several false negatives have resulted from our computational alanine scanning approach in terms of identifying deleterious mutations, it is worth emphasizing that there are no false positives: all of the mutations that our approach has predicted to be strongly destabilizing are in experimental agreement. Thus, our approach has the potential to be predictive of deleterious mutations that may be difficult to detect by experimental strategies.

## Discussion

### Comparison to previous methods

We have developed a novel computational alanine scanning approach that reliably identifies mutations that destabilize the p53tet dimer. The approach also has reasonable success in identifying the most strongly destabilizing mutations, and is able to determine the effects of mutations that cannot be measured by experiment. While our kinetic approach does not yield results as quantitatively accurate as prior approaches, which are based on thermodynamic analyses,[15,16,19] it has a number of advantages. First of all, with a modest amount of computation required per mutant (under a week on a 2.8 GHz Intel processor for each of the 100 20 ns simulations), our approach is ideal for identifying kinetically destabilizing mutations, which are often thermodynamically destabilizing as well. The approach can therefore be applied to high-throughput mutagenesis studies. Second, the protocol, by design, incorporates flexibility of the protein backbone, generating a large ensemble of structures for each mutant at room temperature. Finally, our approach uses a general, physics-based energy function that is predictive and does not rely on parameterization to a particular subset of all proteins or database of structures.[16]

### Limitations

An obvious criticism of our approach is the use of high-temperature (470 K) simulations. While it has been argued that high-temperature simulations can be representative of room-temperature folding/unfolding pathways,[20] it is possible also that such simulations are biased toward "fast-track" pathways that are uncharacteristic of pathways at room temperature.[21] Nonetheless, even if the absolute unfolding rates are not directly comparable to

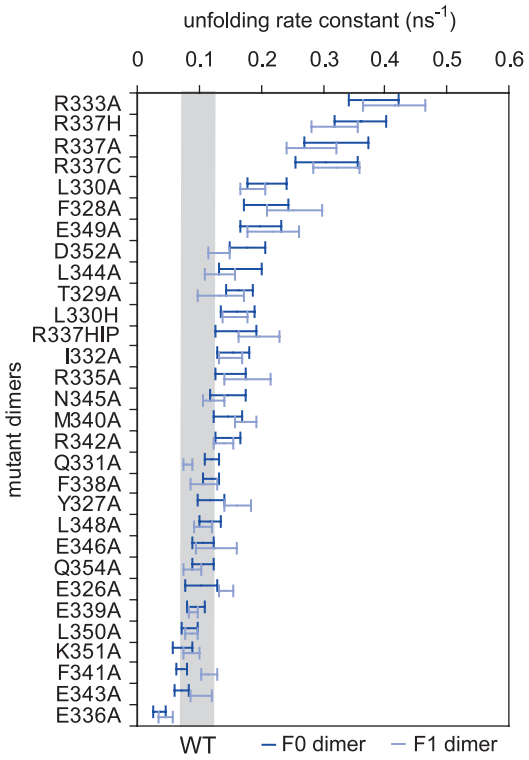


**Figure 6.** Unfolding rate constants determined for mutant dimers in the alanine scanning and test set based on the F0 dimer and the F1 dimer. Ranges shown for the unfolding rate constants represent mean values ± one standard deviation and are plotted in descending order according to mean values for the F0 mutant dimers. The gray region labeled as WT is the range for the unfolding rate constant determined for the wild-type F0 dimer ($0.11(\pm 0.02)\ ns^{-1}$), which has a slightly higher maximum value than that of the wild-type F1 dimer ($0.09(\pm 0.01)\ ns^{-1}$).

experiment, we have shown that $\ln(k_{mut}/k_{wt})$ can be used to predict destabilizing mutations of the p53tet dimer reliably in a qualitative manner.

Another limitation is the use of a dimer model to evaluate the effects of mutations on the stability of the p53tet tetramer. A dimer model was used rather than a tetramer model for a greater likelihood of identifying thermodynamically destabilizing mutations with kinetic analyses. While our approach is unable to detect mutations that destabilize the tetramer, but have little effect on the dimer, destabilizing mutations identified on the basis of the dimer model would certainly destabilize the tetramer. That said, the F0 dimer model is an approximate one, taken from the crystal structure of the tetramer.[9] This issue has been addressed by testing the same protocol on mutants based on an alternative dimer model (F1 dimer). Results from the two dimer models generally agree (see Figures 3(a) and (c), and 5), with a few exceptions, demonstrating the robustness of our protocol.

Finally, our approach to alanine scanning identifies destabilizing mutations only if the folded dimer state is destabilized more than the transition state to unfolding. Since transition state conformations are often relatively unstructured, this differential destabilization of the folded and transition states would generally occur for destabilizing mutations. However, as mentioned above, the breakdown of this assumption may account for the fact that the I332A, F341A, F338A mutations in the alanine scanning set as well as the cancer-linked L330H mutation in the test set do not unfold significantly faster than wild-type, even though the mutations are thermodynamically destabilizing to a significant extent. These mutations, which lie either in the hydrophobic core or its periphery, may destabilize the transition state nearly as much as they destabilize the folded state, thus resulting in only a moderate reduction in the barrier to unfolding.

### Determining shifts in the transition state

In addition to analyzing changes in the stability of the p53tet dimer due to a particular mutation, we tested the effectiveness of an approach for determining changes in the location of the rate-limiting transition state along the dimerization pathway in response to the mutation. This approach involves determining the probability of folding ($p_{fold}$)[22] for a set of protein conformations by performing 100 independent simulations starting from each conformation. Due to the large amount of computational effort required for this approach, it was applied only to mutants in the test set.

As described in Methods, mutant versions of 40 randomly selected putative transition state conformations of the wild-type dimer were created by first mutating the appropriate residues, then re-solvating the resulting mutant conformation with new configurations of water molecules and the appropriate number of neutralizing counter-ions. To test the effect of solvent (and counter-ion) configurations on the $p_{fold}$ values, we removed the original water molecules and counter-ions from the wild-type conformations, added a different configuration of solvent and counter-ions, and then re-evaluated the $p_{fold}$ values. Given the uncertainties of the computations, the resulting $p_{fold}$ values for the re-solvated wild-type conformations ($WT^{solv}$) are in excellent agreement with the values for the original wild-type conformations (WT) (see Figure 7). The dimerization pathway of p53tet is therefore determined solely by the protein conformations and not by the surrounding water configurations. A previous study on the folding of the monomeric protein BBA5 had the same finding, suggesting that water equilibrates much faster than the protein conformation rearranges along the folding pathway.[23]

Since the water dynamics do not affect the dimerization pathway, any differences in the $p_{fold}$ values of mutant dimers relative to those of wild-type can be attributed to the mutations. Correlations of $p_{fold}$ values for wild-type *versus* $p_{fold}$

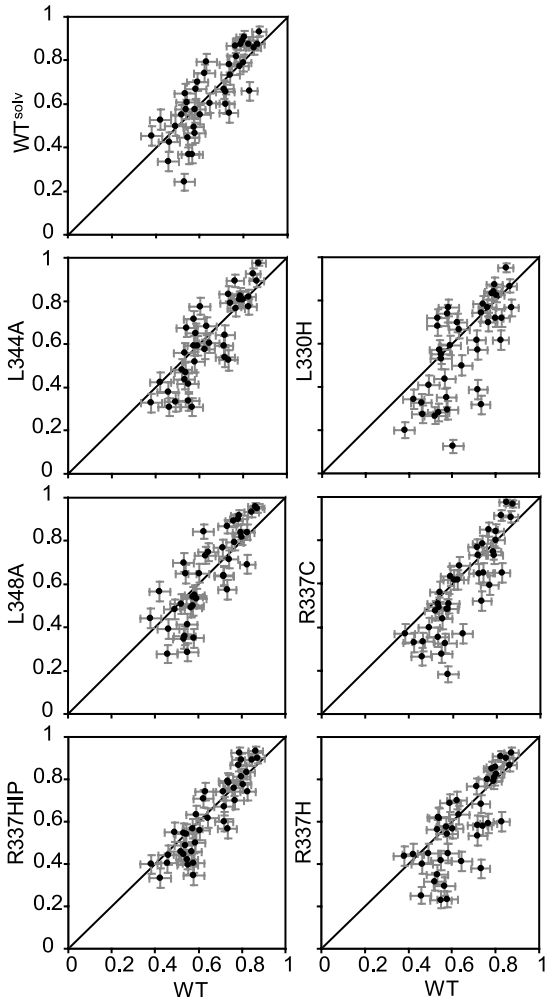


**Figure 7.** Correlations of $p_{fold}$ values for 40 randomly selected wild-type (WT) F0 dimer conformations *versus* $p_{fold}$ values for mutant versions of the conformations, which were each resolvated with different configurations of water molecules and counter-ions (see Methods). As a control, $p_{fold}$ values were determined for resolvated wild-type ($WT^{solv}$) conformations. The $p_{fold}$ values for a given dimer conformation were determined from 100 simulations at 300 K, yielding uncertainties of less than or equal to 0.05. The diagonal line in each plot is included as a guide for visualizing deviations from perfect correlation.

values for mutant conformations are shown in Figure 7. As predicted for the cancer-associated L330H, R337C, and R337H mutant conformations, more than half of the $p_{fold}$ values are lower than those of the wild-type conformations, suggesting that the transition states have moved closer to the folded dimer states, thereby lowering the probabilities of folding. In the context of the Hammond postulate for a unimolecular process, the movement of the transition state along the reaction coordinate, or Hammond effect, is likely caused by destabilization of the folded state.[24] On the other hand, the R337HIP, L344A, and L348A mutant conformations yield $p_{fold}$ values that correlate with those of the wild-type conformations to the same extent as $p_{fold}$ values of the re-solvated wild-type conformations, implying no shifts in the locations of the transition states. The correlations of these mutants with wild-type conformations are consistent with their predicted near-wild-type stabilities.

**Table 1.** Average $p_{fold}$ differences

| Mutant | Average $\Delta p_{fold}$ |
|---|---|
| $WT^{solv}$ | 0.006 |
| L330H | 0.077 |
| R337C | 0.084 |
| R337H | 0.092 |
| R337HIP | 0.027 |
| L344A | 0.026 |
| L348A | −0.001 |

Average $p_{fold}$ differences (average $\Delta p_{fold}$; see Methods) between a total of 40 $p_{fold}$ values for wild-type conformations and $p_{fold}$ values of mutant versions of the conformations, including re-solvated wild-type conformations ($WT^{solv}$) as a control. For comparison, the average uncertainty of $p_{fold}$ values for wild-type and all mutants is ~0.04.

To quantify deviations from a perfect correlation, the average differences between the $p_{fold}$ values of the wild-type conformations and conformations with a particular mutation were computed and are shown in Table 1. Average differences were significant, or greater than the average uncertainty for the $p_{fold}$ values (~0.04 for wild-type and all mutants, as can be seen in Figure 7), for the L330H, R337C, and R337H mutant conformations. As expected, the wild-type conformations with new water and ion configurations as well as the R337HIP, L344A, and L348A mutant conformations did not have significant average differences. Thus, even though unfolding simulations were unable to detect the L330H as being strongly destabilizing, results from $p_{fold}$ simulations show that this mutation is nearly as destabilizing as the R337C and R337H mutations, and certainly more destabilizing than mutations (R337HIP, L344A, and L348A) that either moderately destabilize the dimer or have little effect. These results demonstrate that $p_{fold}$ values, which effectively order conformations along a pathway,[22,25] can be used to determine changes in the location of the transition state due to mutations, and infer effects on the stability of the folded state in the context of the Hammond postulate.

## Conclusions and Future Work

In conclusion, we have developed a novel, kinetic approach for computational alanine scanning that distinguishes rapidly between kinetically stabilizing and destabilizing mutations. With some exceptions, mutations that thermodynamically destabilize the p53tet tetramer appear to kinetically destabilize the dimer, enabling the identification of seven strongly destabilizing mutations (R333A, R337A, R337C, R337H, L330A, F328A, and E349A) with no false positives. Given these encouraging

results, we are in the process of performing more extensive mutagenesis of the p53 oligomerization domain, analyzing the effects of all possible single-nucleotide polymorphisms at each residue position in order to predict cancer-associated mutations. Finally, we have shown that $p_{fold}$ simulations can be used to determine mutation-induced shifts in the rate-limiting transition state for dimerization that are consistent with experiment. Both the computational alanine scanning and $p_{fold}$ simulation approaches can be applied to any system, provided that the folded and unfolded states are clearly defined.

## Methods

### Simulation details

Model building and simulations were performed using the GROMACS MD software package[26] modified for the Folding@Home distributed computing infrastructure.[27] Coordinates of heavy atoms were taken from the crystal structure of p53tet (Protein Data Bank 1AIE)[9] in its active tetrameric form to create a starting model of the wild-type dimer (F0 dimer). An alternative, relaxed model (F1 dimer) of the transient dimer intermediate was obtained from room-temperature MD simulations in explicit water that evolved from the F0 dimer.[12] Heavy atoms for residue mutations, affecting both monomers, were positioned using the SCAP side-chain prediction program in the Jackal 1.5 protein structure modeling software package.[28] Acetyl and *N*-methyl capping groups were added to the N terminus and to the C terminus, respectively, of each monomer (residues 326–355). Hydrogen atoms were added using ionization states present in neutral solution. Neutral histidine residues were protonated at the ε-nitrogen atom, which is the predominant tautomer for free histidine in solution.[29] The dimer was solvated in a cubic box of TIP3P water,[30] with an initial box length of 50 Å, then charge-neutralized by adding counter-ions.

MD simulations were performed using the modified AMBER ff94 force-field developed by Garcia and Sonbanmatsu (AMBER-GS),[31] which is essentially the ff94 force-field[32] with the torsional potentials of ϕ and ψ angles set to zero to achieve agreement with experiment for the helical propensity of peptides.[33] Normally, no scaling of 1–4 van der Waals interactions is used in the AMBER-GS force-field. In this study, 1–4 van der Waals interactions were divided by a factor of 2, which is standard for all AMBER force-fields. A 10 Å cut-off distance was used for coulombic and van der Waals interactions along with a reaction-field treatment[34] of long-range electrostatics and periodic boundary conditions. A dielectric constant of 80 was used beyond the Coulombic cut-off distance. To enable use of a 2 fs time-step, bonds involving hydrogen atoms in the protein and water were constrained to their equilibrium values with the LINCS[35] and SETTLE[36] algorithms, respectively. Constant temperature and pressure (1 atm = 101,325 kPa) were maintained by the Berendsen coupling algorithm[37] with time constants for coupling set to 0.5 ps.

To relieve unfavorable interactions, each initial model was subjected to energy minimization followed by MD equilibration in two stages: first, the solvent and counter ions for 1 ns at 300 K with the protein restrained, then the entire system for 1 ns at 300 K. The second stage of equilibration was carried out in 100 independent trajectories with different initial velocities (selected from a Maxwell–Boltzmann distribution at 300 K) that were run in parallel on the Folding@Home distributed computing network.[27] Each of these independent trajectories was then simulated at 470 K for 20 ns in order to unfold the dimer rapidly, yielding ~2 μs of aggregate simulation time for each mutant dimer.

### Determination of unfolding rates

Definitions for the unfolded and folded manifold were taken from Chong *et al.*[12] These definitions involve the evaluation of the root-mean-squared deviation (RMSD) of the $C^{\alpha}$ atoms in the β-sheet region as well as the RMSD of all $C^{\alpha}$ atoms from the initial dimer model. Fitting for RMSD calculations was performed using the quaternion superposition algorithm,[38] as implemented in the program ProFit†. For each mutant dimer, the increase in the ensemble fraction unfolded ($f_u(t)$) as a function of time ($t$) can be described for $t > t_0$ by a single-exponential function with a lag time ($t_0$):

$$f_u(t) = 1 - e^{-k_u(t-t_0)}$$

where $k_u$ is the rate constant for unfolding at 470 K. The ensemble fraction unfolded was determined every 200 ps from simulations of the mutant dimer at 470 K, yielding 1σ uncertainties $\sigma[f_u(t)]$ that are ~0.05 or less according to a binomial distribution:

$$\sigma[f_u(t)] = \sqrt{\frac{f_u(t)[1-f_u(t)]}{N_{tot}}}$$

where $N_{tot}$ is the total number of unfolding simulations (i.e. 100). To estimate the unfolding rate constant ($k_u$) at 470 K based on the long-timescale events, as opposed to the short-timescale events represented by the lag portion ($t < t_0$) of the data, the natural logarithm of both sides of the equation was taken:

$$\ln[1-f_u(t)] = -k_u t + k_u t_0$$

and the uncertainties $\sigma[f_u(t)]/[1-f_u(t)]$ in the $\ln[1-f_u(t)]$ values were then used to compute the weights in a weighted linear, least-squares fit of the data.[39]

To optimize the accuracy of the fit to the linear portion of the data, which corresponds to single-exponential unfolding kinetics, weighted linear fits were performed for successively larger sets of data starting from the latest time data point (i.e. at 20 ns) and working backwards until a linear fit could no longer pass within two standard deviations of all of the data. Due to the significant amount of noise in $\ln[1-f_u(t)]$ when the ensemble fraction unfolded is 0.85 or greater, this portion of the data was ignored in the fit. While it is straightforward to compute the uncertainty in the unfolding rate constant from a weighted linear fit,[39] this approach was not used, since it underestimates the uncertainty for time-correlated ensemble kinetics data. Instead, the uncertainty in the unfolding rate constant was determined using the bootstrap method,[40] randomly selecting sets of 100 trajectories, with replacement, from the 100 independent trajectories in 20 trials and determining the unfolding rate constant for each set. An average unfolding rate

† http://www.bioinf.org.uk/software/profit/

constant and its uncertainty could then be determined from 20 estimated unfolding rate constants.

#### Determination of $p_{fold}$ values

The probability of entering the folded basin before the unfolded basin ($p_{fold}$)[22] for any particular protein conformation was evaluated by performing a large number of independent simulations at room temperature evolving from that conformation and terminating when either the unfolded or folded basin (described above) is reached. The $p_{fold}$ simulations were performed for 40 protein conformations that were selected randomly from the 799 putative transition state structures identified by Chong *et al*.[12] Mutations were modeled, solvated, and charge-neutralized by adding counter ions as described above. Each model was energy minimized followed by equilibration of the solvent, counter ions, and mutated residues for 1 ns at 300 K. To determine the probability of folding, 100 independent simulations starting from the equilibrated system were performed at 300 K, yielding standard deviations for $p_{fold} \leq 0.05$. Conformations were saved for analysis with a sampling period of 200 ps. Average $p_{fold}$ differences (average signed $\Delta p_{fold}$) between sets of mutant and wild-type conformations were computed as follows:

$$\text{Average } \Delta p_{fold} = \sum_{i=1}^{N} \frac{p_{fold,i}^{mut} - p_{fold,i}^{WT}}{N}$$

where $N$ is the number of conformations (i.e. 40).

## Acknowledgements

We thank the Folding@Home volunteers who made this work possible. We are grateful for helpful discussions with John Chodera. This work was supported by grants from NSF Molecular Biophysics NIH NIGMS, and CPIMA (an NSF MRSEC).

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.12.083

## References

1. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. (1991). p53 mutations in human cancers. *Science*, **253**, 49–53.
2. Varley, J. M., McGown, G., Thorncroft, M., Cochrane, S., Morrison, P., Woll, P. *et al*. (1996). A previously undescribed mutation within the tetramerization domain of TP53 in a family with Li-Fraumeni syndrome. *Oncogene*, **12**, 2437–2442.
3. Lomax, M. E., Barnes, D. M., Gilchrist, R., Picksley, S. M., Varley, J. M. & Camplejohn, R. S. (1997). Two functional assays employed to detect an unusual mutation in the oligomerization domain of p53 in a Li-Fraumeni like family. *Oncogene*, **14**, 1869–1874.
4. Davison, T. S., Yin, P., Nie, E., Kay, C. & Arrowsmith, C. H. (1998). Characterization of the oligomerization defects of two p53 mutants found in families with Li-Fraumeni and Li-Fraumeni-like syndrome. *Oncogene*, **17**, 651–656.
5. Ribeiro, R. C., Sandrini, F., Figueiredo, B., Zambetti, G. P., Michalkiewicz, E., Lafferty, A. R. *et al*. (2001). An inherited p53 mutation that contributes in a tissue-specific manner to pediatric adrenal cortical carcinoma. *Proc. Natl Acad. Sci. USA*, **98**, 9330–9335.
6. Lee, W., Harvey, T. S., Yin, Y., Yau, P., Litchfield, D. & Arrowsmith, C. H. (1994). Solution structure of the tetrameric minimum transforming domain of p53. *Nature Struct. Biol.* **1**, 877–890.
7. Clore, G. M., Ernst, J., Clubb, R., Omichinski, J. G., Kennedy, W. M., Sakaguchi, K. *et al*. (1995). Refined solution structure of the oligomerization domain of the tumour suppressor p53. *Nature Struct. Biol.*, 2.
8. Jeffrey, P. D., Gorina, S. & Pavletich, N. P. (1995). Crystal structure of the tetramerization domain of the tumour suppressor p53. *Science*, **267**, 1498–1502.
9. Mittl, P. R., Chene, P. & Grutter, M. G. (1998). Crystallization and structure solution of p53 (residues 326-356) by molecular replacement using an NMR model as template. *Acta Crystallog. sect. D*, **54**, 86–89.
10. Mateu, M. G., Sanchez Del Pino, M. M. & Fersht, A. R. (1999). Mechanism of folding and assembly of a small tetrameric protein domain from tumor suppressor p53. *Nature Struct. Biol.* **6**, 191–198.
11. Nicholls, C. D., McLure, K. G., Shields, M. A. & Lee, P. W. K. (2002). Biogenesis of p53 involves cotranslational dimerization of monomers and posttranslational dimerization of dimers. *J. Biol. Chem.* **277**, 12937–12945.
12. Chong, L. T., Snow, C. D., Rhee, Y. M. & Pande, V. S. (2005). Dimerization of the p53 oligomerization domain: identification of a folding nucleus by molecular dynamics simulations. *J. Mol. Biol.* **345**, 869–878.
13. DiGiammarino, E. L., Lee, A. S., Cadwell, C., Zhang, W., Bothner, B., Ribeiro, R. C. *et al*. (2002). A novel mechanism of tumorigenesis involving pH-dependent destabilization of a mutant p53 tetramer. *Nature Struct. Biol.* **9**, 12–16.
14. Mateu, M. G. & Fersht, A. R. (1998). Nine hydrophobic side chains are key determinants of the thermodynamic stability and oligomerization status of tumor suppressor p53 tetramerization domain. *EMBO J.* **17**, 2748–2758.
15. Massova, I. & Kollman, P. A. (1999). Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **121**, 8133–8143.
16. Kortemme, T. & Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl Acad. Sci. USA*, **99**, 14116–14121.
17. Guerois, R., Nielsen, J. E. & Serrano, L. (2002). Predicting changes in the stability of protein and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387.
18. Friedler, A., Veprintsev, D. B., Hansson, L. O. & Fersht, A. R. (2003). Kinetic instability of p53 core domain mutants. *J. Biol. Chem.* **278**, 24108–24112.
19. Zhong, H. & Carlson, H. A. (2005). Computational studies and peptidomimetic design for the human p53-MDM2 complex. *Proteins: Struct. Funct. Bioinformat.* **58**, 222–234.

20. Day, R., Bennion, B. J., Ham, S. & Daggett, V. (2002). Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.* **322**, 189–203.
21. Dinner, A. R. & Karplus, M. (1999). Is protein unfolding the reverse of protein folding? A lattice simulation analysis. *J. Mol. Biol.* **292**, 403–419.
22. Du, R., Pande, V. S., Grosberg, A. Y. & Tanaka, T. (1998). On the transition coordinate for protein folding. *J. Chem. Phys.* **108**, 334–350.
23. Rhee, Y. M., Sorin, E. J., Jayachandran, G., Lindahl, E. & Pande, V. S. (2004). Simulations of the role of water in the protein-folding mechanisms. *Proc. Natl Acad. Sci. USA*, **101**, 6456–6471.
24. Fersht, A. R. (1999). *Structure and Mechanism in Protein Science.* 2nd edit., W.H. Freeman and Company, New York.
25. Pande, V. S. & Rokhsar, D. S. (1999). Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein G. *Proc. Natl Acad. Sci. USA*, **96**, 9062–9067.
26. Lindahl, E., Hess, B. & van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317.
27. Pande, V. S., Baker, I., Chapman, J., Elmer, S. P., Khaliq, S., Larson, S. M. *et al*. (2003). Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers*, **1**, 91–109.
28. Xiang, J. Z. (2001). Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol.* **311**, 421–430.
29. Tanokura, M., Tasumi, M. & Miyazawa, T. (1976). $^{1}$H nuclear magnetic resonance studies of histidine-containing di- and tripeptides. Estimation of the effects of charged groups on the $pk_a$ value of the imidazole ring. *Biopolymers*, **15**, 393–401.
30. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. (1983). Comparison of simple potential function for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
31. Garcia, A. E. & Sanbonmatsu, K. Y. (2002). Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl Acad. Sci. USA*, **99**, 2782–2787.
32. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M. *et al*. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
33. Beachy, M., Chasman, D., Murphy, R., Halgren, T. & Friesner, R. (1997). Accurate *ab initio* quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. *J. Am. Chem. Soc.* **119**, 5908–5920.
34. Neumann, M. & Steinhauser, O. (1980). The influence of boundary conditions used in machine simulations on the structure of polar systems. *Mol. Phys.* **39**, 437.
35. Hess, B., Bekker, H., Berendsen, H. & Fraaije, J. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472.
36. Miyamoto, S. & Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962.
37. Berendsen, H., Postma, J., van Gunsteren, W., DiNola, A. & Haak, J. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
38. McLachlan, A. D. (1982). Rapid comparison of protein structures. *Acta Crystallog. sect. A*, **38**, 871–873.
39. Taylor, J. R. (1997). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurement.* 2nd edit., University Science Books, Mill Valley, CA.
40. Chernick, M. R. (1999). Bootstrap methods: a practioner's guide. In *Wiley series in probability and statistics*, Wiley-Interscience, New York.

***Edited by M. Levitt***

*(Received 10 October 2005; received in revised form 15 December 2005; accepted 29 December 2005)*
Available online 17 January 2006