

# Dimerization of the p53 Oligomerization Domain: Identification of a Folding Nucleus by Molecular Dynamics Simulations

Lillian T. Chong<sup>1</sup>, Christopher D. Snow<sup>1,2</sup>, Young Min Rhee<sup>1</sup> and Vijay S. Pande<sup>1\*</sup>

<sup>1</sup>Departments of Chemistry and Structural Biology, Stanford University, Stanford, CA 94305-5080, USA

<sup>2</sup>Graduate Group in Biophysics, Stanford University, Stanford CA 94305, USA

Dimerization of the p53 oligomerization domain involves coupled folding and binding of monomers. To examine the dimerization, we have performed molecular dynamics (MD) simulations of dimer folding from the rate-limiting transition state ensemble (TSE). Among 799 putative transition state structures that were selected from a large ensemble of high-temperature unfolding trajectories, 129 were identified as members of the TSE *via* calculation of a 50% transmission coefficient from at least 20 room-temperature simulations. This study is the first to examine the refolding of a protein dimer using MD simulations in explicit water, revealing a folding nucleus for dimerization. Our atomistic simulations are consistent with experiment and offer insight that was previously unobtainable.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** tumor suppressor p53; oligomerization; protein folding; protein interactions; molecular dynamics simulation

\*Corresponding author

## Introduction

The exquisite reliability of proteins in recognizing their targets has been intensely studied for more than a century. Several paradigms for protein recognition have emerged, varying in the degree to which the proteins undergo conformational changes upon binding. The simplest paradigm involves rigid docking of folded proteins with complementary binding surfaces. A second paradigm is induced-fit binding, where folded proteins rearrange to optimize their interactions upon association. Yet another paradigm involves folding of intrinsically unstructured proteins only upon binding their targets.<sup>1</sup>

Induced-fit and rigid binding may apply consecutively in the folding of the tetrameric p53 tumor suppressor.<sup>2</sup> This protein, known as the “gatekeeper” of cell growth and division, is of great interest to cancer research. In addition, the small,

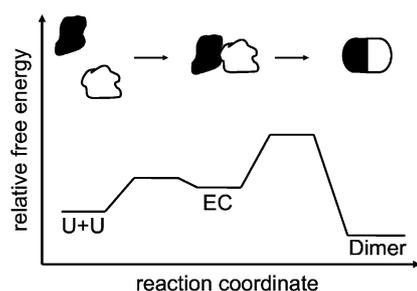
isolated oligomerization domain (p53tet) of p53 forms stable tetramers, providing an ideal system for studying the dynamics of protein–protein interactions and the folding of oligomers. The domain is comprised of a  $\beta$ -strand, a tight turn, and an  $\alpha$ -helix. The native tetramer is a dimer of dimers, where each dimer is formed by an antiparallel  $\beta$ -sheet and two antiparallel helices.<sup>3–6</sup>

Based on kinetics experiments and  $\Phi$ -value analysis, the folding of p53tet is thought to occur in two stages.<sup>2</sup> First, unfolded monomers bind while folding to a transient dimeric intermediate. Structured dimers then bind to form the tetramer. The observed rate constant for folding of p53tet appeared bimolecular and corresponded to the first stage of folding.<sup>2</sup> The rate constant did not vary with solvent viscosity and is therefore not diffusion-limited, yielding a rate that is more than two orders of magnitude slower than the diffusion-limited rate.

A bimolecular rate constant that is not diffusion-limited can result from a pathway in which the monomers bind to form an unstable encounter complex (EC) that dissociates more rapidly than it rearranges to form the dimer (Figure 1). Thus, the unimolecular rearrangement of the EC is rate-limiting while the rate constant still depends on the concentration of unassociated monomers. In analogy to the unfolded ensemble of a monomeric

Abbreviations used: p53tet, p53 oligomerization domain; EC, encounter complex; MD, molecular dynamics; TSE, transition state ensemble; TS, transition state; RMSD, root-mean-squared deviation; dRMS, distance root-mean-square.

E-mail address of the corresponding author: [pande@stanford.edu](mailto:pande@stanford.edu)



**Figure 1.** Proposed qualitative energy profile for dimerization of p53tet. Unfolded monomers (U+U) bind to form an unstable encounter complex (EC), which then undergoes a rate-limiting rearrangement to the dimer.

protein, the EC of a dimeric protein provides an opportunity for the denatured dimer to fold, rather than relying on the monomers to be in the correct orientations upon collision.

Understanding the role of conformational changes in protein–protein interactions has broad implications ranging from drug design to the study of protein function in the larger context of biological pathways in the cell. While Brownian dynamics simulations have been effective for examining binding kinetics of rigid proteins,<sup>7</sup> they are not a means to analyze protein conformational changes. Examining conformational changes using molecular dynamics (MD) simulations is appealing due to the great temporal resolution and atomistic detail. Informative MD (and Monte Carlo) simulations of protein–protein interactions have been performed using models employing Gō potentials in which one approximates the nature of interatomic interactions by setting native interactions to one attractive energy and non-native interactions to one repulsive energy.<sup>8–10</sup> However, if the timescales are accessible, MD simulations employing empirical molecular mechanics potentials (e.g. AMBER, CHARMM with explicit solvation) can provide a more detailed view of the thermodynamics and kinetics of proteins.

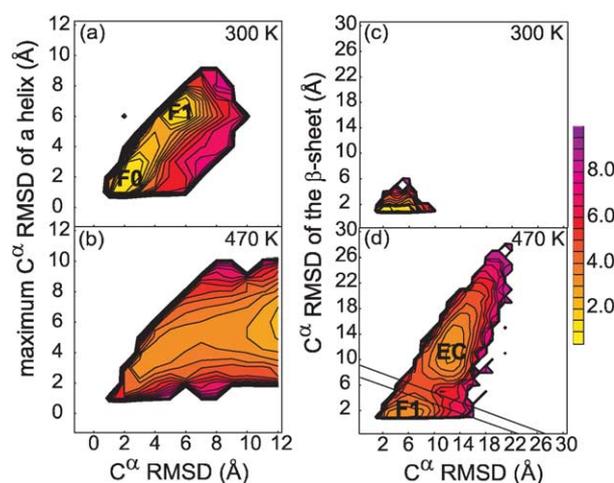
Our goal is to use such MD simulations with extensive sampling from distributed computing<sup>11</sup> to examine induced-fit dimerization of p53tet in explicit water. While micromolar concentrations of monomer were used in the folding kinetics experiments,<sup>2</sup> the lowest concentration afforded by typical MD simulations is  $\sim 0.1$  M monomer (two monomers in a  $50 \text{ \AA}^3$  box of water molecules and counterions). Using the measured rate constant for folding ( $3.11 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ ),<sup>2</sup> we extrapolate the folding time for the dimer at  $0.1$  M monomer to be  $\sim 30 \mu\text{s}$ , which is not easily accessible to MD simulations. Most of this time, however, is spent waiting for a fluctuation to reach the transition state ensemble (TSE), or the highest free energy state along the pathway.<sup>12</sup> Once a configuration reaches the unstable TSE, it will either unfold or fold quickly (likely tens of nanoseconds).<sup>11–15</sup>

We therefore simulated folding of the p53tet dimer from the rate-limiting TSE. TSE structures were taken from high-temperature unfolding simulations of the dimer and identified based on 50% probabilities of refolding at room temperature. This unfolding/refolding strategy has been successful for monomeric proteins using MD simulations;<sup>13–15</sup> it has also been used in Monte Carlo simulation studies of an intrinsically unstructured protein that folds upon binding a rigid, target dimer complex.<sup>16</sup> This study is the first to apply this strategy to dimerization of a protein using MD simulations.

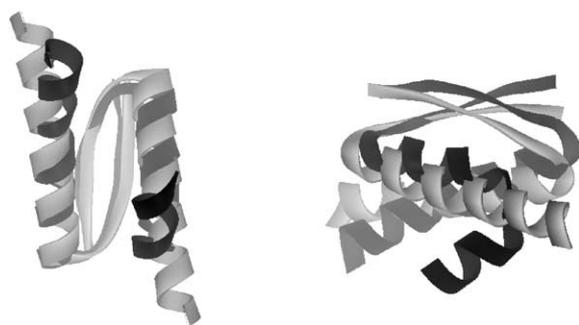
## Results

### Stability of the dimer

Hundreds of independent simulations at 300 K that started from our initial dimer model (see Methods) drifted rapidly to a relaxed ensemble, which has significant root-mean-squared deviation (RMSD) of the  $C^\alpha$  atoms ( $C^\alpha \text{ RMSD} > 4.5 \text{ \AA}$ ) from the original model. By fitting the cumulative probability of reaching the relaxed ensemble with a single exponential function, we computed the rate constant for conversion of the initial to the relaxed ensemble. The rate constant for this conversion is  $0.05 \text{ ns}^{-1}$  at 300 K. At 470 K, this rate constant is 20 times faster, with  $\sim 90\%$  conversion after 5 ns (Figure 2(b)). As revealed in Figure 2(a), the relaxed ensemble (F1) has large deviations from the initial



**Figure 2.** Potential of mean force surfaces for ensembles from plotting the maximum  $C^\alpha$  RMSD of a helix versus  $C^\alpha$  RMSD of the entire structure at (a) 300 K and (b) 470 K, and the  $C^\alpha$  RMSD of the  $\beta$ -sheet versus  $C^\alpha$  RMSD at (c) 300 K and (d) 470 K. Deviations are relative to the initial dimer model (F0). Data shown for each temperature are from structures sampled every nanosecond from more than  $22 \mu\text{s}$  of aggregate simulation time after 5 ns of simulation. Contours are drawn at intervals of the available thermal energy,  $0.5 \text{ RT}$ . The diagonal lines shown in (d) are defined boundaries for the unfolded (EC) and folded basins (F1).



**Figure 3.** Superposition of the relaxed dimer fold (black) and half of the tetramer crystal structure (gray).

model (F0) due to deviations in at least one of the two helices (maximum  $C^\alpha$  RMSD of a helix  $> 5 \text{ \AA}$ ).

To further examine the relaxed F1 ensemble, we clustered structures of the ensembles at 15 ns using a  $1 \text{ \AA}$   $C^\alpha$  distance root-mean-square (dRMS) cutoff. Only one cluster resulted, containing 67% of the 272 structures. The closest ( $0.4 \text{ \AA}$   $C^\alpha$  dRMS) structure to the average  $C^\alpha$ - $C^\alpha$  distance matrix of the cluster was selected to be the representative F1 structure. This structure has a kinked helix, but is otherwise similar to the initial dimer model (F0) with small deviations ( $1.3 \text{ \AA}$   $C^\alpha$  RMSD across residues 326–348) throughout the  $\beta$ -sheet, turn, and N-terminal portions of the helices (Figure 3). As rationalized previously for the NMR structure of a designed p53tet dimer,<sup>17</sup> the more open and flexible helices of the representative F1 structure are likely due to the absence of stabilizing interactions contributed by the opposing dimer in the tetramer structure.

The F1 ensemble is stable. For example, after 20 ns, 771 independent simulations that started from the F1 structure only deviate by an average  $C^\alpha$  RMSD of  $2.1 (\pm 1.1 \text{ standard deviation}) \text{ \AA}$ . Some conversion of F1 to F0 conformations ( $C^\alpha$  RMSD from F0 model  $< 4 \text{ \AA}$ ) occurred with a rate constant of  $0.02 \text{ ns}^{-1}$ , which is more than twice as slow as the rate of F0 to F1 conversion ( $C^\alpha$  RMSD from F0 model  $> 4.5 \text{ \AA}$ ) estimated above. The F0 and F1 dimer ensembles are thus likely to be local minima that are both populated at 300 K.

### Unfolding the dimer

Progress of dimer unfolding simulations at 470 K was monitored *via*  $C^\alpha$  RMSD of the  $\beta$ -sheet and of the overall structure from the starting F0 structure. Only two distinct ensembles were apparent: one folded (F1) and one unfolded (EC) (Figure 2(d)). The unfolded ensemble has large deviations of the overall structure that are coupled to the disappearance of the  $\beta$ -sheet between monomers. In most cases, the monomers remained associated in the unfolded dimers, consistent with EC structures. While the disassociation/association of monomers to form the EC is an integral part of dimer formation, this step will be left for future exploration. As expected, no unfolded

conformations were apparent at 300 K after 5 ns (Figure 2(c)).

### Identifying members of the TSE

A putative transition state (TS) region containing 799 structures was delineated by defining boundaries for the folded (F1) and unfolded basins (EC) (Figure 2(d)). To test these boundaries, at least 100 independent simulations were performed for two trial structures: one from the folded basin and another from the unfolded basin. These simulations remained in the basins from which they started for at least 35 ns. However, since estimation of the TSE can depend strongly on the choice of order parameters, a more robust criterion for identifying TS structures is to determine the probability of folding ( $p_{\text{fold}}$ ) for each of the 799 putative TS structures from at least 20 simulations starting from the structure.<sup>18</sup> Of these 799 structures, 129 structures were more rigorously identified as members of the TSE with  $\sim 50\%$  probabilities of entering the folded basin ( $0.4 < p_{\text{fold}} < 0.6$ ). These structures are diverse with an average pairwise  $C^\alpha$  RMSD that is  $9.2 \text{ \AA}$  (2.8 to  $18.3 \text{ \AA}$ ).

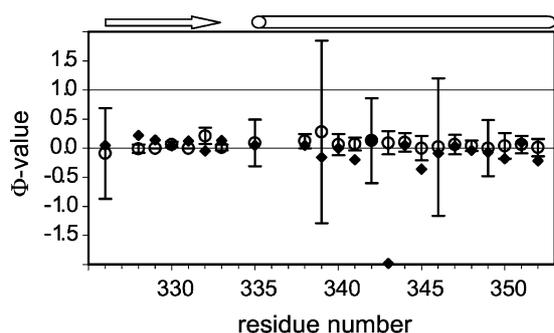
### Refolding the dimer

Simply entering the folded basin does not necessarily mean that the simulation will immediately reach the folded state. One may define the folded state as the ensemble of conformations that are within one standard deviation ( $1.1 \text{ \AA}$ ) greater than the average  $C^\alpha$  RMSD ( $2.1 \text{ \AA}$ ) of the simulated dimer (F1) ensemble at 300 K and 20 ns, or within  $3.2 \text{ \AA}$  ( $C^\alpha$  RMSD) from the starting structure. Using this somewhat generous criterion for being folded, 197 out of more than 3000 simulations starting from various members of the TSE refolded. Using more stringent criteria for folding, such as being within  $2 \text{ \AA}$  or  $1 \text{ \AA}$  from the F1 dimer, significantly decreases the number of folding events to 12 or 2, respectively. The most refolded simulation folds within  $0.8 \text{ \AA}$  of the F1 dimer. While the criterion for being folded has a drastic effect on the number of folding events determined, it does not affect the details of our proposed mechanism of folding presented in Discussion. The strong dependence of the number of folding events on the criterion for being folded suggests that not all of the simulations starting from members of the TSE were carried out long enough to refold completely. Indeed, only  $\sim 37.5\%$  of the simulations have been performed for 20 ns. As mentioned in Introduction, complete folding from the TSE of the p53tet dimer is likely to require tens of nanoseconds.

## Discussion

### Limitations of the strategy

An obvious criticism of our unfolding/refolding

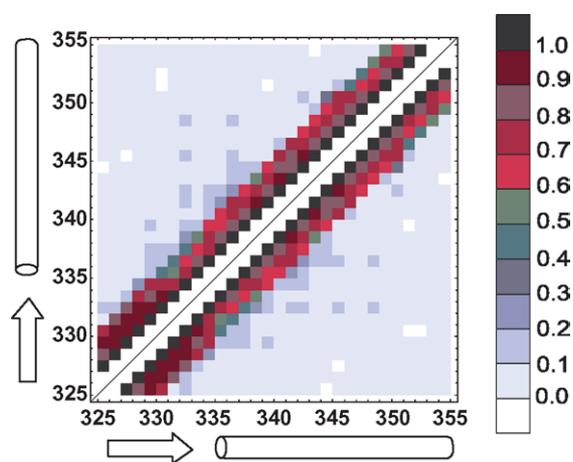


**Figure 4.** Comparison of 23 computed (circles) and experimentally measured (diamonds)  $\Phi$ -values. Standard deviations for the computed values are also shown.

strategy is the selection of TS structures from high-temperature (i.e. 470 K) simulations. While it has been argued that high-temperature simulations can be representative of room-temperature folding/unfolding pathways,<sup>19</sup> such simulations could be biased toward uncharacteristic, “fast-track” pathways.<sup>20</sup> Unfortunately, simulating the unfolding of p53tet at just 100 K lower (370 K) would be impractical, since it is estimated (by Arrhenius extrapolation) to take  $\sim 1 \mu\text{s}$  to unfold at this temperature, requiring more than a year of CPU time on a typical PC. Most importantly, even though our TS structures were sampled by unfolding at 470 K, these structures were verified to be members of the TSE by  $p_{\text{fold}}$  simulations at room temperature and refolded at room temperature.

### Further validation of the TSE

To validate the TSE identified by our  $p_{\text{fold}}$  simulations, we computed  $\Phi$ -values to compare



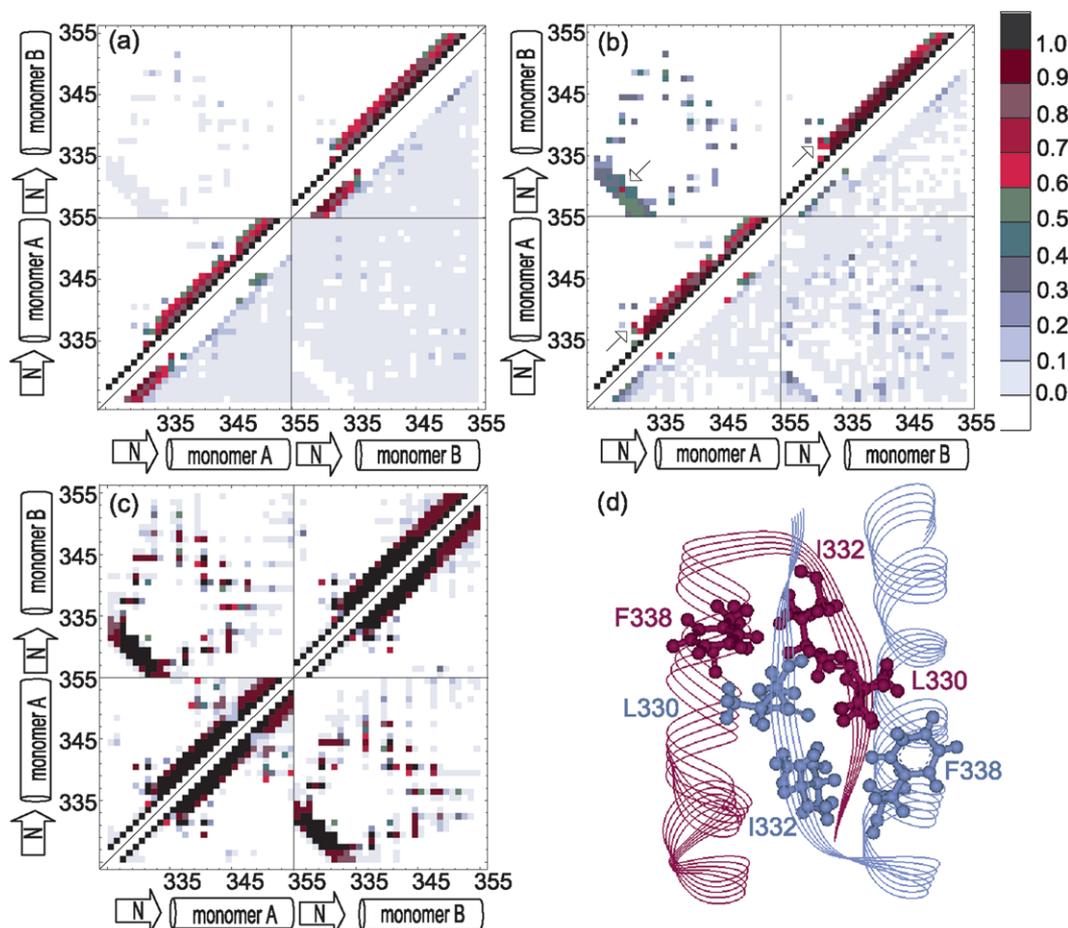
**Figure 5.** Average pairwise residue contact map for unfolded monomer ensemble (described in Methods). The average contact map for the ensemble was generated by averaging the contact maps of each structure in the ensemble. Residues are in contact if they are non-adjacent and any of their constituent heavy atoms are within 6 Å.

with those determined from experimental data.  $\Phi$ -value analysis is a protein engineering method that probes TS structure by making mutations in different positions of proteins and relating the resultant destabilization of the TS to the effect on stability.<sup>21</sup> From the experimental data on p53tet dimer formation, such values were determined using  $\Phi = 2\Delta\Delta G_{\text{TS-U}}/\Delta\Delta G_{\text{N-U}}$ , where  $\Delta\Delta G_{\text{TS-U}}$  is the difference in the free energy of activation between wild-type and mutant TSEs and  $\Delta\Delta G_{\text{N-U}}$  is the difference in the free energy of unfolding between wild-type and mutant tetramers.<sup>2</sup> The corresponding equation for computing  $\Phi$ -values from simulated structures is described in Methods. Such values typically range from 0 (unfolded) to 1 (folded).

The 23 measured  $\Phi$ -values are compared with our computed values in Figure 4. For many of the computed  $\Phi$ -values, large standard deviations arise from the diversity of the TSE and unfolded monomer ensembles. All but one of the computed  $\Phi$ -values are close to the measured values, which are mostly near zero. The exception is the  $\Phi$ -value for E343A, which was computed to be 0.09 instead of the unusually negative value of  $-1.98$  measured by experiment. This discrepancy is mainly due to the fact that our strategy of computing  $\Phi$ -values, which considers only native contacts, is not likely to yield large negative values. Excluding this outlier, the remaining 22 computed and measured values differ by an average of  $0.14(\pm 0.11)$ .

Since virtually all  $\Phi$ -values are near zero, one would not necessarily expect to achieve a strong correlation between computed and measured values. Intrinsic sources of error in the measured  $\Phi$ -values could exceed any fitting errors and rival the deviations observed in the computed values. A potential source of error is the evaluation of  $\Phi$ -values at denaturant concentrations that are not within the range directly measured but dependent on extrapolation. Another source of error is the use of the free energy of tetramer unfolding instead of the free energy of dimer unfolding. While it was not possible to measure the free energy of unfolding the wild-type p53tet dimer,<sup>2</sup> using that of the tetramer assumes that the free energy of unfolding the tetramer is equivalent to the free energies of unfolding two dimers. This assumption of additivity may not be valid for all of the  $\Phi$ -values, especially those corresponding to positions along the dimer-dimer interface.

Near-zero  $\Phi$ -values can result if TS structures are largely unstructured or if the regions probed are similarly structured in the transition and unfolded states. Our TSE obtained from simulations demonstrates both scenarios. The  $\beta$ -strand regions are largely unstructured while the helical regions, which are mostly formed in the unfolded state ( $57(\pm 17)\%$  helical structure among 450 conformations (see Methods)), are equally structured in both the transition and unfolded states. Residual structure in the unfolded monomeric state was previously ruled out based on a sequence-based



**Figure 6.** Average pairwise residue contact maps of the (a) EC, (b) TSE, and (c) F1 ensembles (described in Methods). Residues are in contact if they are non-adjacent and any of their constituent heavy atoms are within 6 Å. The average contact map for an ensemble was generated by averaging the contact maps of each structure in the ensemble. In (a) and (b), native contacts, or contacts formed in the F1 dimer structure, are displayed above the diagonal; non-native contacts are below the diagonal. The locations of folding nucleus residues are shown in (d).

prediction made by the AGADIR program that the monomers have only  $\sim 4\%$  helical structure.<sup>2</sup> While Garcia and Sanbonmatsu have shown that the force field used for our simulations agrees with the experimental helical propensities of peptides,<sup>22</sup> some degree of bias toward helix formation may still be present. However, predictions made by AGADIR may go astray for peptides with tertiary interactions, which can significantly increase the helical content.<sup>23</sup> In our unfolded monomer ensemble,  $\sim 74\%$  of the conformations form at least one tertiary contact between residues that are separated by a minimum of eight residues (Figure 5). Thus, it is plausible that the low  $\Phi$ -values of residues in the helical regions could be due to residual structure in the unfolded state.

Another point of validation for the simulated TSE is computation of the Tanford  $\beta$  value ( $\beta_T$ ), which measures the degree of buried surface area in the TS relative to that of the denatured state from the native state. This value was determined from experimental data to be 0.20, suggesting that the TSE is relatively expanded, burying only 20% of the surface area buried in the native tetramer.<sup>2</sup> The

compact, native-like TSE from our simulations yields a computed average  $\beta_T$  value of  $0.44(\pm 0.25)$  (see Methods), which appears to somewhat overestimate the compactness of the TS, since the experimental value is barely within the computational uncertainty. However, even though the  $\beta_T$  value is thought to be a reliable index of the compactness of the TS for monomeric proteins, it can be misleading for multimeric proteins that have different oligomerization states for the TS and folded ensembles. In the case of p53tet,  $\beta_T$  will never be large even if the TS is highly compact due to the fact that two, unassociated TS dimers do not contact each other along the dimer-dimer interface that is buried in the native tetramer. Ideally, the  $\beta_T$  value for dimerization should be evaluated by computing the amount of buried surface area in the folded dimer intermediate that is buried in the dimeric TS. While such evaluation is not possible by experimental methods, computation of the  $\beta_T$  value using the simulated dimeric intermediate (F1) ensemble instead of the native tetramer ensemble yields a much higher  $\beta_T$  value of  $0.69(\pm 0.37)$ . Thus, the low  $\beta_T$  value determined from experiments does

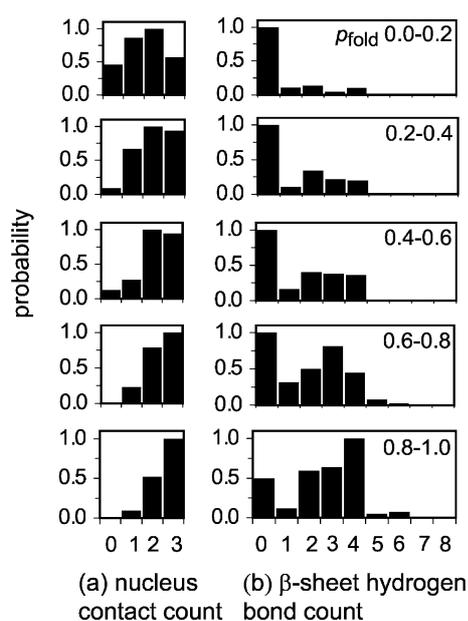
not necessarily mean that the structure of the TS for p53tet dimer formation is expanded. Our identified TSE is therefore consistent with experimental observations,  $\beta_T$  as well as  $\Phi$ -values.

### Mechanism of folding

To identify common features of p53tet dimer folding from our large ensemble of trajectories, we analyzed the average pairwise residue contact maps for the EC, TS, and folded dimer (F1) ensembles. As shown in Figure 6(a), much of the "native" F1 helical structure (Figure 6(c)) is present in the EC ensemble with some non-native contacts within the monomers, particularly in the native  $\beta$ -strand regions. Although a large number of intermonomer contacts (mostly non-native) are formed, none of these contacts is frequent, suggesting non-specific monomer-monomer associations. Once folding progresses from the EC to the TS (Figure 6(b)), most non-native interactions within the monomers disappear while intermonomer contacts generally remain heterogeneous. Despite the diversity of the TSE, more than 70% of the conformations contain the same three native contacts in the hydrophobic core of the dimer: two contacts within each monomer, between I332 and F338, and one contact between L330 of each monomer (Figure 6(d)). With their proximity to each other in the TSE and absence in the EC ensemble, these contacts are characteristic of a "folding nucleus," which has been found for a few small proteins,<sup>24</sup> but not previously for protein-protein interactions.

Residues in the folding nucleus have been shown to be critical for the stability and oligomerization of p53tet.<sup>25</sup> Mutation of F338 to an alanine residue (F338A) yields significant destabilization ( $\Delta\Delta G_{N-U}$  of  $\sim 9$  kcal/mol) and decreases the folding rate by a small amount (within  $2\times$ ),<sup>2</sup> suggesting a modest kinetic role for F338. In contrast, removing a single methyl group in I332V (I332V) or L330 (L330Nva (to norvaline)) leads to a small change in stability ( $\Delta\Delta G_{N-U}$  of  $\sim 3-4$  kcal/mol) and a similar folding rate to the wild-type.<sup>2</sup> However, removing two additional alkyl groups (V332A or Nva330A) prevented folding, resulting in only monomers. Certainly these residues are important for the thermodynamics of folding; it is unclear to what extent they are important for kinetics. The high flux through pathways that involve formation of a folding nucleus in the context of our simulations does not imply that the nucleus must form in the TS; mutation of residues in the nucleus need not drastically perturb kinetics due to structural plasticity and pathway diversity.

To monitor formation of the folding nucleus, we first ordered the 799 putative TS structures by their computed  $p_{fold}$  values and then determined the probabilities of forming contacts in the folding nucleus in various  $p_{fold}$  intervals (Figure 7(a)). Since these structures were selected from unfolding trajectories using the  $C^\alpha$  RMSDs of the  $\beta$ -sheet and

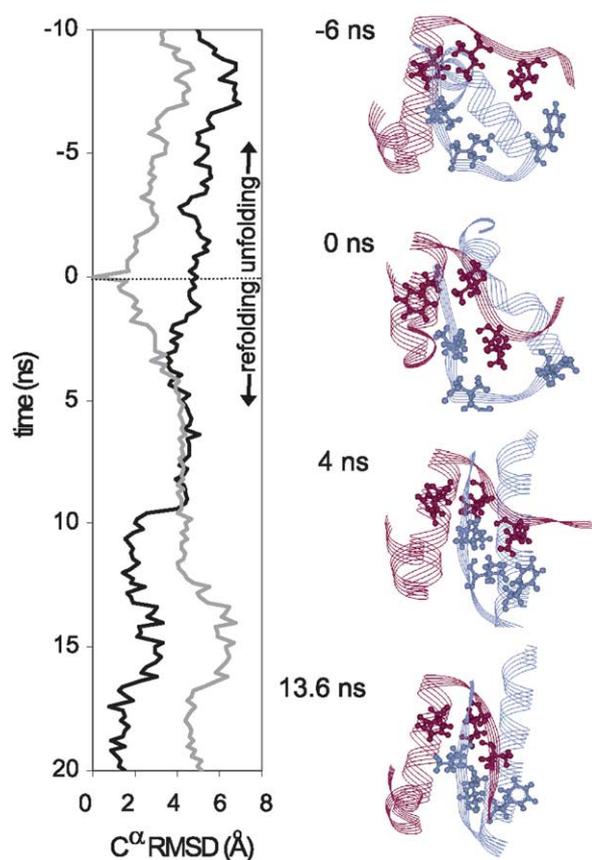


**Figure 7.** Probability distributions of (a) number of folding nucleus contacts and (b) number of  $\beta$ -sheet hydrogen bonds of structures with  $p_{fold}$  values in the following intervals: 0–0.2 (304 structures), 0.2–0.4 (121 structures), 0.4–0.6 (129 structures), 0.6–0.8 (124 structures), and 0.8–1 (121 structures).

of the overall structure, care must be taken in interpreting progress along the  $p_{fold}$  reaction coordinate. That said, either two or all three of the folding nucleus contacts formed before the TS (0.2–0.4  $p_{fold}$  interval). In comparison,  $\beta$ -sheet hydrogen bonds began forming after the TS (0.6–0.8  $p_{fold}$  interval) (Figure 7(b)). Thus, in the most probable pathways, formation of the folding nucleus precedes and is likely necessary for  $\beta$ -sheet formation as well as other intermonomer contacts that exist in the folded dimer (see Figure 6(c)).

In summary, the mechanism of folding for the p53tet dimer begins with non-specific binding of monomers to form an EC. EC conformations that then rearrange to allow formation of the folding nucleus are more likely to proceed beyond the TS and fold. The folding nucleus appears to position the  $\beta$ -strands such that they can more easily form the  $\beta$ -sheet as well as other intermonomer contacts. The rate-limiting rearrangement of the EC to the dimer is therefore consistent with a nucleation-condensation mechanism, as previously proposed by Mateu *et al.*<sup>2</sup> for dimerization of p53tet.

To illustrate the proposed mechanism, we combine the most refolded and unfolded  $p_{fold}$  trajectories at 300 K from an initial TS structure (Figure 8). Due to microscopic reversibility, we may represent pre-TSE folding events from post-TSE unfolding events in reverse. After 6 ns of unfolding (or  $-6$  ns of folding) the TS, an EC is reached that has only one folding nucleus contact and no hydrogen bonds between the  $\beta$ -strands. Gradually,



**Figure 8.** Selected structures from the most refolded trajectory at 300 K. The EC (−6 ns), obtained from unfolding the TS (0 ns), is also shown. Only side-chains of residues in the folding nucleus are displayed. As shown in the plot of time versus  $C^\alpha$  RMSD, the TS refolds to a structure that is within 0.8 Å of the F1 dimer (in black) and 4.9 Å from the TS (in gray).

the  $\beta$ -strands “zipper” from one end, forming one backbone hydrogen bond after another, until four of the eight hydrogen bonds are attained in the TS. By the time the TS is reached, the structure has formed two of the three folding nucleus contacts, but still deviates largely ( $C^\alpha$  RMSD of 4.9 Å) from the folded dimer (F1) due to differences in the halfway completed  $\beta$ -sheet and in the partially unraveled helix that eventually forms an  $\alpha$ -helix at 4 ns. The  $\beta$ -strands continue to zipper until the  $\beta$ -sheet is fully formed at 13.6 ns and stays formed for the remainder of the 20 ns simulation. The overall structure folds within 0.8 Å  $C^\alpha$  RMSD of the folded dimer at 17.4 ns.

Levy *et al.* recently applied simplified Gō simulations to fold 11 homodimers and study the degree to which the proteins rearrange during recognition.<sup>9</sup> Even with a simplified model, which included a covalent link between the monomers, this study was able to identify unfolded intermediates along certain pathways. However, the atomistic, physical simulations that we have used in this study could offer several potential benefits. One

potential advantage is that these detailed simulations can provide a more thorough analysis of transient states. For example, Levy *et al.* have used such simulations to study the unassociated monomers of the HIV-1 protease dimer.<sup>8</sup> Another advantage is that our models incorporate non-specific contacts, which are apparent in the EC ensemble for p53tet. Finally, our models allow the eventual computation of rates for folding and/or binding that can be compared with experimental results.

## Conclusions

The refolding of a protein dimer from the rate-limiting TSE was observed for the first time using MD simulations in explicit water. As recently observed for a monomeric protein,<sup>26</sup> conformations along the folding pathway display a great deal of structural plasticity yet form the same folding nucleus. The folding nucleus appears necessary for the formation of hydrogen bonds in the  $\beta$ -sheet and other intermonomer contacts. The importance of the nucleus is underscored by the thermodynamic effects of mutating the constituent residues.<sup>2,25</sup> Furthermore, malfunction of p53 *via* mutation of one of the residues (L330) to a histidine residue is linked to cancer.<sup>27</sup>

A dominant paradigm in protein–protein interactions has been that of the association of rigid monomers. However, another mode for binding, such as the dimerization observed here for p53tet, could include cases where folding is concomitant with binding. In these cases, paradigms of protein folding, such as the role of a specific nucleus, may play a significant role in binding.

It is also interesting to consider the biological impact of the folding kinetics of p53tet. A recent study has suggested that p53tet dimerizes cotranslationally and then forms the tetramer posttranslationally.<sup>28</sup> Completing dimerization before the p53 chains leave the polysome has a significant kinetic advantage due to the higher local concentration. Thus, even a subtle mutation that affects the kinetics could play an important biological role.

## Methods

### Simulation details

Model building and simulations were performed using the GROMACS MD software package<sup>29</sup> modified for the Folding@Home distributed computing infrastructure.<sup>11</sup> Coordinates of heavy atoms were taken from the crystal structure of p53tet (1AIE in the Protein Data Bank)<sup>6</sup> in its active tetrameric form to create starting models of the monomer, dimer, and tetramer. Acetyl and *N*-methyl capping groups were added to the N terminus and C terminus, respectively, of each monomer (residues 326–355). Hydrogen atoms were added using ionization states present in neutral solution. Models of the monomer,

dimer, and tetramer were solvated in cubic boxes of TIP3P water<sup>30</sup> with initial box lengths of 40 Å, 50 Å, and 50 Å, respectively, 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),<sup>22</sup> which is essentially the ff94 force field<sup>31</sup> with the torsional potentials of  $\phi$  and  $\psi$  angles set to zero to achieve agreement with experiment for the helical propensity of peptides.<sup>32</sup> A 10 Å cut-off distance was used for Coulombic and van der Waals interactions along with a reaction-field treatment<sup>33</sup> 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 were constrained to their equilibrium values with the LINCS algorithm.<sup>34</sup> Constant temperature and pressure (1 atm) were maintained by the Berendsen coupling algorithm<sup>35</sup> with time constants for coupling set to 0.5 ps.

To relieve unfavorable interactions, each initial model was subjected to energy minimization followed by equilibration in two stages: first, the solvent and counter-ions for 1 ns at 300 K, then the entire system for 1 ns at either 300 K or 470 K. For each temperature and system, more than 10  $\mu$ s of aggregate simulation time was accumulated from 1000 independent trajectories with different initial velocities (selected from a Maxwell distribution) that were run in parallel on the Folding@Home distributed computing network.<sup>11</sup>

### Evaluating structural similarity

Fitting for RMSD calculations was performed using the McLachlan algorithm<sup>36</sup> as implemented in the program ProFit (Martin, A.C.R.).<sup>†</sup> Distance root-mean-square (dRMS) values were calculated as described by Zagrovic *et al.*<sup>37</sup> Hydrogen bonds and the percentage of helical structure were determined by the DSSP program.<sup>38</sup> Structures were clustered using the Jarvis-Patrick method,<sup>39</sup> adding a structure to a cluster if this structure and a structure in the cluster are neighbors and they shared at least three other neighbors. Neighbors of a structure were all structures within a dRMS cutoff.

### Defining ensembles

All ensembles consisted of structures taken after 20 ns of simulation at the appropriate temperature: 733 structures for the native tetramer ensemble at 300 K ( $C^\alpha$  RMSD 1.3( $\pm$ 0.6) Å from the starting tetramer model), 638 structures for the dimer ensemble at 300 K ( $C^\alpha$  RMSD 2.1( $\pm$ 1.1) Å from the starting relaxed dimer structure), 450 structures for the unfolded monomer ensemble at 470 K ( $C^\alpha$  RMSD 6.9( $\pm$ 1.3) Å from the starting monomer model), and 476 structures of unfolded dimers for the EC ensemble at 470 K ( $C^\alpha$  RMSD 12.9( $\pm$ 1.9) Å from the starting dimer model), excluding unassociated monomers (minimum  $C^\alpha$ - $C^\alpha$  intermonomer distance  $\geq$  15 Å).

Members of the TSE were identified by measuring the transmission coefficient, or probability of entering the folded basin ( $p_{\text{fold}}$ ).<sup>18</sup> A  $p_{\text{fold}}$  of  $\sim$ 0.5 defines the TS. To determine this probability, an average of 26( $\pm$ 2) independent  $p_{\text{fold}}$  simulations were run for each putative TS structure at 300 K, yielding standard deviations within 0.1. Structures were analyzed every 200 ps. Simulations

were terminated when either the unfolded or folded basin was reached. The TSE used for subsequent analysis consists of 129 structures with  $p_{\text{fold}}$  values ranging from 0.4 to 0.6.

### Computing $\Phi$ -values

$\Phi$ -values ( $\Phi_i^{\text{MD}}$ ) were computed using a strategy similar to those used in other studies:<sup>14,40,41</sup>

$$\Phi_i^{\text{MD}} = (2\langle N_i \rangle_{\text{TSE}} - 4\langle N_i \rangle_{\text{U}}) / (\langle N_i \rangle_{\text{N}} - 4\langle N_i \rangle_{\text{U}})$$

where  $N_i$  is the number of native contacts formed by residue  $i$ ,  $\langle N_i \rangle_{\text{TSE}}$  is over all conformations in the dimeric TSE,  $\langle N_i \rangle_{\text{N}}$  is over all conformations in the native tetramer state, and  $\langle N_i \rangle_{\text{U}}$  is over all conformations in the unfolded monomer state. As defined by Gsponer and Caflisch,<sup>14</sup> contacts between non-adjacent residues were counted for each instance where their side-chain heavy atoms are closer than 6 Å. Following Ding *et al.*,<sup>42</sup> the unfolded state has been incorporated to account for residual structure in that state.

### Computing the Tanford $\beta$ value

The Tanford  $\beta$  value ( $\beta_{\text{T}}$ ) for the TS to dimerization was computed using the following equation:

$$\beta_{\text{T}} = (2\langle \text{SASA} \rangle_{\text{TSE}} - 4\langle \text{SASA} \rangle_{\text{U}}) / (\langle \text{SASA} \rangle_{\text{N}} - 4\langle \text{SASA} \rangle_{\text{U}})$$

where SASA is the solvent-accessible surface area of a given conformation,  $\langle \text{SASA} \rangle_{\text{TSE}}$  is the average SASA for all conformations in the TSE,  $\langle \text{SASA} \rangle_{\text{N}}$  is over all conformations of the native tetramer ensemble, and  $\langle \text{SASA} \rangle_{\text{U}}$  is over all conformations of the unfolded monomer ensemble. The equation approximates the ratio  $\text{RT}(m_{\text{kf}}/m_{\text{eq}})$  from experimental data, where  $m_{\text{kf}}$  is the sensitivity of the rate constant for refolding to the concentration of denaturant and  $m_{\text{eq}}$  is the sensitivity of the free energy of unfolding the native tetramer to the concentration of denaturant.<sup>2</sup> SASA values were determined by the Alphasurf program (Koehl, P.A.‡).

## Acknowledgements

We thank the Folding@Home volunteers who made this work possible. We are also grateful for helpful discussions with Kevin Plaxco and Joerg Gsponer. This work was supported by a grant from NSF Molecular Biophysics and a predoctoral fellowship from the Howard Hughes Medical Institute (to C.D.S.).

## References

1. Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60.
2. 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.
3. Clore, G. M., Ernst, J., Clubb, R., Omichinski, J. G.,

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

‡ <http://csb.stanford.edu/koehl/ProShape>

- Kennedy, W. M., Sakaguchi, K. *et al.* (1995). Refined solution structure of the oligomerization domain of the tumour suppressor p53. *Nature Struct. Biol.* 1995;, 2.
4. 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.
  5. Jeffrey, P. D., Gorina, S. & Pavletich, N. P. (1995). Crystal structure of the tetramerization domain of the tumour suppressor p53. *Science*, 267, 1498–1502.
  6. 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.
  7. Gabdouliline, R. R. & Wade, R. C. (2002). Biomolecular diffusional association. *Curr. Opin. Struct. Biol.* 12, 204–213.
  8. Levy, Y., Caflich, A., Onuchic, J. N. & Wolynes, P. G. (2004). The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations. *J. Mol. Biol.* 340, 67–79.
  9. Levy, Y., Wolynes, P. G. & Onuchic, J. N. (2004). Protein topology determines binding mechanism. *Proc. Natl Acad. Sci. USA*, 101, 511–516.
  10. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2002). Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. *J. Mol. Biol.* 324, 851–857.
  11. 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.
  12. Chandler, D. (1987). *Introduction to Modern Statistical Mechanics*, Oxford University Press, Oxford.
  13. 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.
  14. Gsponer, J. & Caflich, A. (2002). Molecular dynamics simulations of protein folding from the transition state. *Proc. Natl Acad. Sci. USA*, 99, 6719–6724.
  15. Alonso, D. O. & Daggett, V. (1995). Molecular dynamics simulations of protein unfolding and limited refolding: characterization of partially unfolded states of ubiquitin in 60% methanol and in water. *J. Mol. Biol.* 247, 501–520.
  16. Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Freer, S. T. & Rose, P. W. (2003). Simulating disorder–order transitions in molecular recognition of unstructured proteins: where folding meets binding. *Proc. Natl Acad. Sci. USA*, 100, 5148–5153.
  17. Davison, T. S., Nie, X., Ma, W., Lin, Y., Kay, C., Benchimol, S. & Arrowsmith, C. H. (2001). Structure and functionality of a designed p53 dimer. *J. Mol. Biol.* 307, 605–617.
  18. Du, R., Pande, V. S., Grosberg, A. Y. & Tanaka, T. (1998). On the transition coordinate for protein folding. *J. Chem. Phys.* 108, 334–350.
  19. 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.
  20. Dinner, A. R. & Karplus, M. (1999). Is protein unfolding the reverse of protein folding? A lattice simulation analysis. *J. Mol. Biol.* 292, 403–419.
  21. Fersht, A. R. & Sato, S. (2004). Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl Acad. Sci. USA*, 101, 7976–7981.
  22. 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.
  23. Lacroix, E., Viguera, A. R. & Serrano, L. (1998). Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* 284, 173–191.
  24. Fersht, A. R. (1997). Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* 7, 3–9.
  25. 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.
  26. Hubner, I. A., Oliveberg, M. & Shakhnovich, E. I. (2004). Simulation, experiment, and evolution: understanding nucleation in protein S6 folding. *Proc. Natl Acad. Sci. USA*, 101, 8354–8359.
  27. Cariello, N., Cui, L., Beroud, C. & Soussi, T. (1994). Database and software for the analysis of mutations in the human p53 gene. *Cancer Res.* 54, 4454–4460.
  28. 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.
  29. 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.
  30. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935.
  31. 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.
  32. 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.
  33. Neumann, M. & Steinhauser, O. (1980). The influence of boundary conditions used in machine simulations on the structure of polar systems. *Mol. Phys.* 39.
  34. Hess, B., Bekker, H., Berendsen, H. & Fraaije, J. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472.
  35. Berendsen, H., Postma, J., van Gunsteren, W., DiNola, A. & Haak, J. (1984). Molecular dynamics with coupling to an external bath. *J. Comput. Phys.* 81, 3684–3690.
  36. McLachlan, A. D. (1982). Rapid comparison of protein structures. *Acta Crystallog. A*, 38, 871–873.
  37. Zagrovic, B., Snow, C. D., Khaliq, S., Shirts, M. R. & Pande, V. S. (2002). Native-like mean structure in the unfolded ensemble of small proteins. *J. Mol. Biol.* 323, 153–164.
  38. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22, 2577–2637.
  39. Downs, G. M., Willett, P. & Fisanick, W. (1994).

- Similarity searching and clustering of chemical-structure databases using molecular property data. *J. Chem. Inf. Comput. Sci.* **34**, 1094–1102.
40. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. *Nature*, **409**, 641–645.
41. Li, A. & Daggett, V. (1994). Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl Acad. Sci. USA*, **91**, 10430–10434.
42. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2002). Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys. J.* **83**, 3525–3532.

*Edited by C. R. Matthews*

*(Received 30 August 2004; received in revised form 26 October 2004; accepted 27 October 2004)*