Solvation in protein folding analysis: Combination of theoretical and experimental approaches

A. M. Fernández-Escamilla*, M. S. Cheung†, M. C. Vega‡, M. Wilmanns‡, J. N. Onuchic†, and L. Serrano*§

*European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117, Heidelberg, Germany; †Center for Theoretical Biological Physics and Department of Physics, University of California at San Diego, La Jolla, CA 92093-0319; and ‡European Molecular Biology Laboratory, Notkestrasse 85, Gebäude 25A, D-22603, Hamburg, Germany

Edited by Peter G. Wolynes, University of California at San Diego, La Jolla, CA, and approved December 19, 2003 (received for review July 6, 2003)

An effort to combine theoretical analyses and protein engineering methods has been made to probe the folding mechanism of SH3 by using Energy Landscape Theory and a ϕ-value analysis. Particular emphasis was given to core residues and the effect of desolvation during the folding event by replacing the core valines with isosteric threonines. These mutations have the advantage of keeping the core structurally invariant while affecting core stability relative to the unfolded state. Although the valines that form the core appear spatially invariant, the folding kinetics of their threonine mutants varies, indicating their different extent of solvation in the transition-state ensemble. Theoretical studies predicted the distribution of folding kinetics of threonine mutants without previous knowledge of the measured rates. This initial success encourages further investigations of the molecular details behind these macroscopic phenomena and of the role of solvation in the folding mechanism.

A large body of recent data suggests that proteins, especially small fast-folding (submillisecond) proteins (1–3), have sequences with a level of energetic frustration sufficiently reduced that their overall energy landscape (4–8) resembles a moderately rough funnel (9). The landscape roughness corresponds to local free energy minima arising from geometric (topological) and energetic traps (10). Geometric or topological traps are associated with chain connectivity and the shape of the native fold and occur when correct contacts form prematurely (10). Because the energetic roughness is minimal, the structural heterogeneity observed in the transition-state ensemble (TSE) is strongly influenced by the topological effects, which, to a large extent, may be inferred from the native structure (11–16). This explains why simple energetically unfrustrated (Gō-like) models (17) reproduce nearly all experimental results for the global geometrical features of the TSE and/or intermediates of a large number of real proteins, which are two- or three-state folders (13). These models have been recently generalized to include the ability to capture effects such as the desolvation of the hydrophobic core (14).

Using this theoretical framework, the funnel landscape coupled with microscopic desolvation can be pictured by using two ordering parameters: Q, a fraction of the total native contact formation, and pseudo Q, a fraction of the total single-water separated native contact formation (14). The latter is a direct consequence of using a pairwise potential with a characteristic desolvation barrier, which accounts for entropy costs to expel a single water molecule between two interacting apolar groups before a full contact can be made (14, 18–20). Although no explicit water molecule has been included in the simulations, the emergence of this desolvation barrier emulates the granular properties of solvents. By using this model, we have recently predicted that the folding of SH3 is controlled by a structural-search collapse followed by desolvation of the hydrophobic core (14).

Experimentally, ϕ-value analysis has been the best tool to explore the nature of the TSE (21). In principle, mutations used to probe the TSE of the protein involved mutation into Ala or Gly. The reason behind this is that new interactions are avoided, and solvation changes are relatively minimized (21), probably because hydrophobic residues are partly buried in the denatured state, explaining the inverse hydrophobic effect (22). Several studies using these types of mutations have provided a wealth of information that has shown good agreement with theoretically predicted global geometrical features of the TSE of a large number of real proteins (12). A more interesting mutation, Val→Thr, does not disrupt the core of the protein packing but stabilizes the unfolded state and therefore destabilizes the protein (23). Thus for computation simulations, by replacing the core hydrophobic residue Val by the polar Thr, the net stability of this contact will be reduced; although the van der Waals interactions at the native structure are about the same for both, Thr is much more stable in the unfolded ensemble.

In this paper, we compare theoretical predictions of folding rate variations for several Val→Thr mutations in the core of α-spectrin SH3 with experimentally measured results. These predictions were made with no previous knowledge of the measured rates. From a practical perspective, α-spectrin SH3 is a doable platform for this analysis, because: (*i*) It is a simple two-state folder (24) whose TSE has been studied exhaustively by various mutations (25, 26). (*ii*) It has a significant number of hydrophobic residues at different positions, especially in the core region (five of nine hydrophobic core residues are valines, which can be isosterically replaced by threonines). (*iii*) Its size is manageable for carrying out folding simulations within a reasonable wall time scale (a 62-residue polypeptide that folds into an orthogonally β-sandwich globular protein) (27).

Although a simple protein model, described above, is used, the qualitative agreement between theory and experiments is already excellent. This initial success encourages further investigation of the molecular details behind these macroscopic phenomena.

Methods

A minimalist Cα model with Gō-like interactions generalized to include a desolvation potential is used (for model details, see ref. 14). A construct for the desolvation potential is based on earlier theoretical studies (20, 28) that quantified hydration for small hydrophobes in water. Its importance has been connected to computational simulations of protein folding and dynamics by various groups (10, 14, 18, 19, 29–32). To fully investigate the importance of solvent-mediated interactions in core-packing, simulations are also performed for native interactions represented simply by attractive Lennard–Jones (LJ)-like interactions (12). Nonnative contacts are treated as a hardcore repulsive interactions (parameters are described in ref. 14).

The experimental point mutation, where an apolar residue (valine) is replaced by a polar residue (threonine), is mimicked by modifying the aforementioned Gō-like interactions in the

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: TSE, transition-state ensemble; LJ, Lennard–Jones.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 1UUE).

§To whom correspondence should be addressed. E-mail: serrano@embl.de.

www.pnas.org/cgi/doi/10.1073/pnas.0304180101

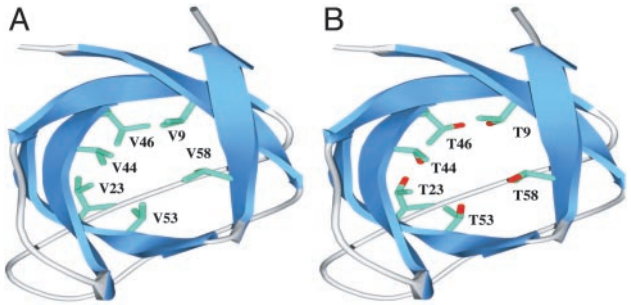

Fig. 1. The structure of the D48G mutant of α-spectrin SH3 domain (1BK2). (*A*) A ribbon diagram of this domain and valines at the hydrophobic core are displayed. (*B*) The same perspective as in *A*, except that valines are mutated to threonines.

following way: given a native attractive contact pair (e.g., 44 in V44T), the mutated interaction is replaced by a hard-core repulsion; i.e., it accounts only for the excluded volume effect (parameters are described in ref. 14). This type of interaction is similar to an HP model (33), in which renormalized energetic contributions between an H (hydrophobic) and a P (hydrophilic) residue are not attractive. Presumably, this modification reflects how a newly replaced Thr interacts with chunks of hydrophobes in the core, provided the native structure is unperturbed. To connect with experiments, computer mutants were created for measured mutants (V9T, V23T, V44T, V53T, and V58T).

Results and Discussion

Exploring TSE with the Aid of Val→Thr Mutations. Because we are interested in the change of stability in response to solvation, a strategy based on mutations of a buried apolar residue to a polar one (that favors unfolded states) is appealing (21). Because both residues have similar volumes and therefore similar core-packing structure, the approach just described that systematically mutates buried valines to isosteric threonines is implemented to investigate the dynamics of the hydrophobic core.

These mutated residues (9, 11, 23, 44, 46, 53, 58, distributed at various β-strands) point inward to the core (except Val-46) (Fig. 1 *A* and *B*) and are presumed to not make any new interactions in the folded state, with perhaps the exception of position 58. At position 58, possibly a new hydrogen bonding to the CO group of Ala 55 (Fig. 1*B*) could be detected. Computationally, we can also mimic mutations of Val→Thr by a logical design of interactions that do not disrupt the native fold and change only the net stability of this contact (see *Supporting Text*, which is published as supporting information on the PNAS web site, for details).

To interpret the kinetic analysis of Val→Thr mutations, it is useful to include the standard analysis of Val→Ala mutation that mainly destabilizes the native folded state on the same protein (21), for two reasons. First, the cost of making a cavity in the protein is greater than the effect due to the difference in the hydrophobic burial of an Ala with respect to a Val (34). Second, it is well known that hydrophobic residues, on average, are not fully exposed in the unfolded state, explaining the inverse hydrophobic effect (22), thus minimizing the solvation difference between a Val and an Ala.

Thus, these two kinds of mutations, which address different energetic considerations, can provide complementary information in revealing kinetics details associated with the TSE from an experimental perspective.

X-Ray Analysis of V44T Mutation. One important assumption in the interpretation of data is that no conformational change takes place on mutating valine residues into threonine ones. To determine whether this is the case, we have selected the critical V44 position, which is at the center of the hydrophobic core and of the folding nucleus. We have determined the crystal structure of V44T (Fig. 2) at 2.6-Å resolution (for methods and parameters details, see *Table 2*, which is published as supporting information on the PNAS web site.

The rms deviation (rmsd) of backbone atoms between the D48G reference mutant and the V44TD48G mutant was calculated with LSQKAB (35) to be merely 0.28 Å. Other than the N or C terminus, the second most deviating spot comprises residues 39–46 (thus including the V44T mutation, which had the highest rmsd value of 0.45 Å), followed by two neighboring β-strands, reflecting minor structural changes to accommodate the Thr side chain.

Thus superimposition of two structures shows essentially no

Table 1. Kinetic parameters for Val→Thr and Val→Ala mutants of D48G-SH3

Proteins	$k_{\ddagger-U}$*, s^{-1}	$m_{\ddagger-U}$†, kcal·mol^{-1}·M^{-1}	$k_{\ddagger-F}$,* s^{-1}	$m_{\ddagger-F}$†, kcal·mol^{-1}·M^{-1}	$\Delta\Delta G_{\ddagger-U}$§, kcal·mol^{-1}	$\Delta\Delta G_{\ddagger-F}$§ kcal·mol^{-1}	$\Delta\Delta G_{F-U}$ kcal·mol^{-1}	$\phi_{\ddagger-U}$¶
D48G‖	63.63 ± 1.20	−0.78 ± 0.01	0.010 ± 0.000	0.47 ± 0.00	–	–	–	–
V9A	51.59 ± 1.84	−0.76 ± .01	0.009 ± 0.009	0.43 ± 0.12	–	–	–	–
V9T	52.52 ± 3.08	−0.73 ± 0.03	0.297 ± 0.054	0.45 ± 0.02	−0.11 ± 0.04	2.00 ± 0.11	−2.12 ± 0.11	0.05 ± 0.05
V23A	16.45 ± 0.67	−0.77 ± 0.02	0.055 ± 0.008	0.46 ± 0.02	−0.80 ± 0.03	1.01 ± 0.09	−1.81 ± 0.09	0.44 ± 0.05
V23T	21.87 ± 0.63	−0.78 ± 0.01	0.042 ± 0.006	0.45 ± 0.02	−0.63 ± 0.02	0.85 ± 0.08	−1.48 ± 0.09	0.43 ± 0.06
V44A	2.75 ± 0.16	−0.91 ± 0.04	0.117 ± 0.005	0.37 ± 0.01	−1.86 ± 0.04	1.46 ± 0.02	−3.32 ± 0.04	0.56 ± 0.02
V44T	1.11 ± 0.08	−0.83 ± 0.05	0.065 ± 0.004	0.48 ± 0.01	−2.40 ± 0.04	1.11 ± 0.04	−3.51 ± 0.06	0.68 ± 0.02
V46A	11.41 ± 0.36	−0.89 ± 0.01	0.008 ± 0.002	0.37 ± 0.04	−1.02 ± 0.02	−0.13 ± 0.15	−0.89 ± 0.15	1.15 ± 0.26
V46T	8.98 ± 0.27	−0.86 ± 0.01	0.002 ± 0.000	0.46 ± 0.02	−1.16 ± 0.02	−0.90 ± 0.00	−0.20 ± 0.02	5.62 ± 0.58
V53A	8.53 ± 0.62	−0.97 ± 0.04	0.054 ± 0.007	0.43 ± 0.02	−1.19 ± 0.04	1.00 ± 0.08	−2.19 ± 0.09	0.54 ± 0.05
V53T	2.79 ± 0.18	−0.86 ± 0.04	0.079 ± 0.006	0.47 ± 0.01	−1.85 ± 0.04	1.22 ± 0.05	−3.07 ± 0.06	0.60 ± 0.02
V58A	40.05 ± 1.32	−0.98 ± 0.02	0.404 ± 0.027	0.41 ± 0.01	−0.27 ± 0.02	2.19 ± 0.04	−2.47 ± 0.05	0.11 ± 0.02
V58T	19.06 ± 0.72	−0.76 ± 0.02	0.061 ± 0.007	0.49 ± 0.01	−0.71 ± 0.03	1.07 ± 0.07	−1.79 ± 0.07	0.40 ± 0.04

The experimental conditions and analysis are described in *Supporting Text*. Errors correspond to fitting errors. 1 kcal = 4.18 kJ.

*Refolding and unfolding rate constant in water.

†Dependence of the natural unfolding-refolding logarithm on urea.

§Gibbs free energy of refolding and unfolding of the mutants respect to D48G Gibbs free energy.

¶ϕ values for refolding.

‖Value determined by A. R. Viguera, personal communication, and fixed for the fitting of the data by using Eq. **1** (*Supporting Text*). This mutant is the reference, because all Val mutations have been performed on it to counteract the expected large destabilization of the Thr and Ala mutations. D48G stabilizes the WT protein by 1.4 kcal·mol^{-1} and is located at the tip of a loop, thus not interfering with the mutations done here.

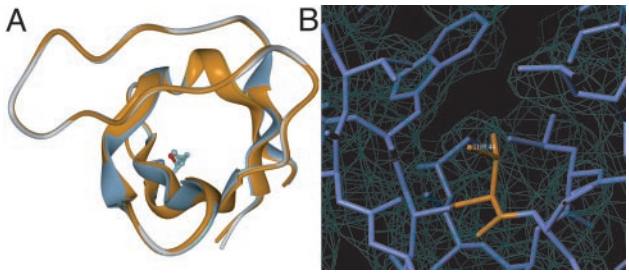

Fig. 2. Crystal structure of V44T. (*A*) Overall topologies of V44TD48G and D48G are superimposed. Secondary structures are colored as follows: for V44TD48G, β-strands, light blue; the 3_{10}-helix, gold (residues 56–58); turns and random coil, white. The backbone of D48G is shown in orange. Side chains at position 44 are presented as colored sticks: OG2 atom of T44, red; CG2 atom of V44, orange. (*B*) A close-up of T44 in a Fourier electron density map calculated with CNS and displayed with TURBO at 1.0 contour level.

conformational changes apart from differences due to comparison with different crystal structures and tight packing of the Thr side chain. This validates our hypothesis that Val→Thr mutation does not perturb the folded state.

Mutation Effect on Folding and Unfolding Events. Kinetic parameters of the unfolding and refolding reactions of all mutants (V9, V23, V44, V46, V53, and V58) based on D48G are shown in Table 1 and their corresponding data in Fig. 3. By comparing folding ($k_{\ddagger\text{-F}}$) and unfolding ($k_{\ddagger\text{-U}}$) rate constants to the D48G reference, we calculate changes in free energy of the TSE (‡) with respect to the unfolded ($\Delta\Delta G_{\ddagger\text{-U}}$) and folded states ($\Delta\Delta G_{\ddagger\text{-F}}$) for each mutant (see *Supporting Text*), for the determination of kinetic and thermodynamic parameters and methods of kinetic measurements). These parameters quantify the change in folding and unfolding barriers, respectively. In addition, differences in free energy of unfolding ($\Delta\Delta G_{\text{F-U}}$) can also be calculated in accordance with the kinetic parameters. For instance, $m_{\ddagger\text{-F}}$ values are related to the difference in solvent accessibility between the folded and TSE states, whereas $m_{\ddagger\text{-U}}$ values relate to the difference between the unfolded and TSE states. Chevron plots for each pair of V/A and V/T mutants are shown in Fig. 3. According to these kinetic parameters, $\phi_{\ddagger\text{-U}}$ values are obtained to acquire knowledge of interaction formations in the TSE by means of mutagenetic studies (21).

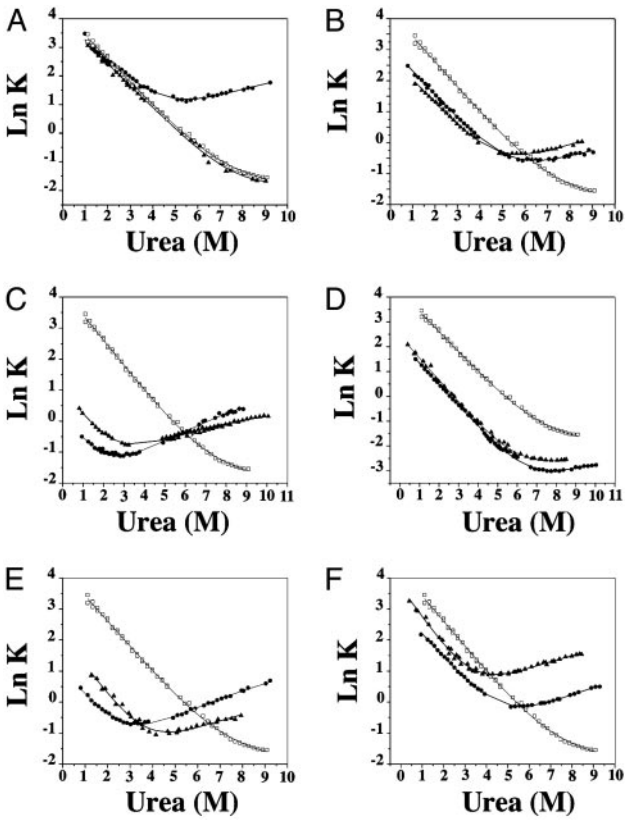

Fig. 3. Chevron curves of Val→Thr and Val→Ala mutants at positions V9 (*A*), V23 (*B*), V44 (*C*), V46 (*D*), V53 (*E*), and V58 (*F*). □, D48G background; ▲, Val→Ala mutant; ●, Val→Thr mutant.

Interestingly, there are different patterns in the analysis of $m_{\ddagger\text{-U}}$ and $m_{\ddagger\text{-F}}$ values between Val→Ala and Val→Thr mutations. For Val→Ala mutations, an overall value of $m_{\ddagger\text{-F}}$ is always lower than the corresponding Val→Thr, whereas $m_{\ddagger\text{-U}}$ values show the opposite (the only exceptions are for $m_{\ddagger\text{-F}}$ and $m_{\ddagger\text{-U}}$ values at positions 9 and 23, where both mutants behave in a similar way). For Val→Thr mutations, generally speaking, both $m_{\ddagger\text{-U}}$ and $m_{\ddagger\text{-F}}$ values are more similar to the reference than for Val→Ala mutations. Such observations could either imply a displacement of TSE moving toward to a more compact conformation of Thr→Ala mutants, or they could merely justify a difference in solvation of Thr and Ala. In any case, these changes are not large enough to invalidate the "protein engineering" analysis.

It is particularly interesting to compare folding kinetics of Thr and Ala mutants, because they can be connected to the folding simulations in this study (for simulation results, please see Table 3 and Fig. 7, which are published as supporting information on the PNAS web site). Comparing the refolding kinetics of Thr mutants to those of Ala mutants, there is a significant decrease in rates observed at positions 44 and 53, whereas other positions, 9, 23, and 46, behave similarly. Furthermore, the opposite is also true for unfolding events, when Thr mutations are done at positions 44 and 53. However, we cannot compare the results of Thr and Ala mutants at position 58, because a nonnative hydrogen bond could be introduced.

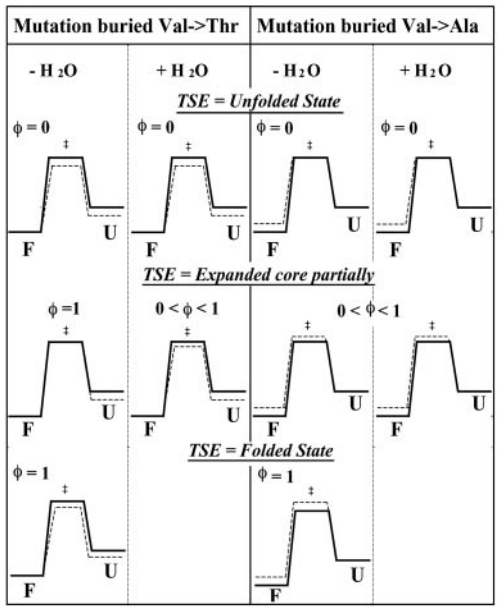

Fig. 4. Free energy profiles of mutating a buried Val→Thr and Val→Ala. Two extreme cases of $\phi_{\ddagger\text{-U}} = 0$ and $\phi_{\ddagger\text{-U}} = 1$ are argued. The effect of the Thr mutation stems from stabilizing the unfolded state, and the effect of Ala mutation stems from destabilizing the folded state.

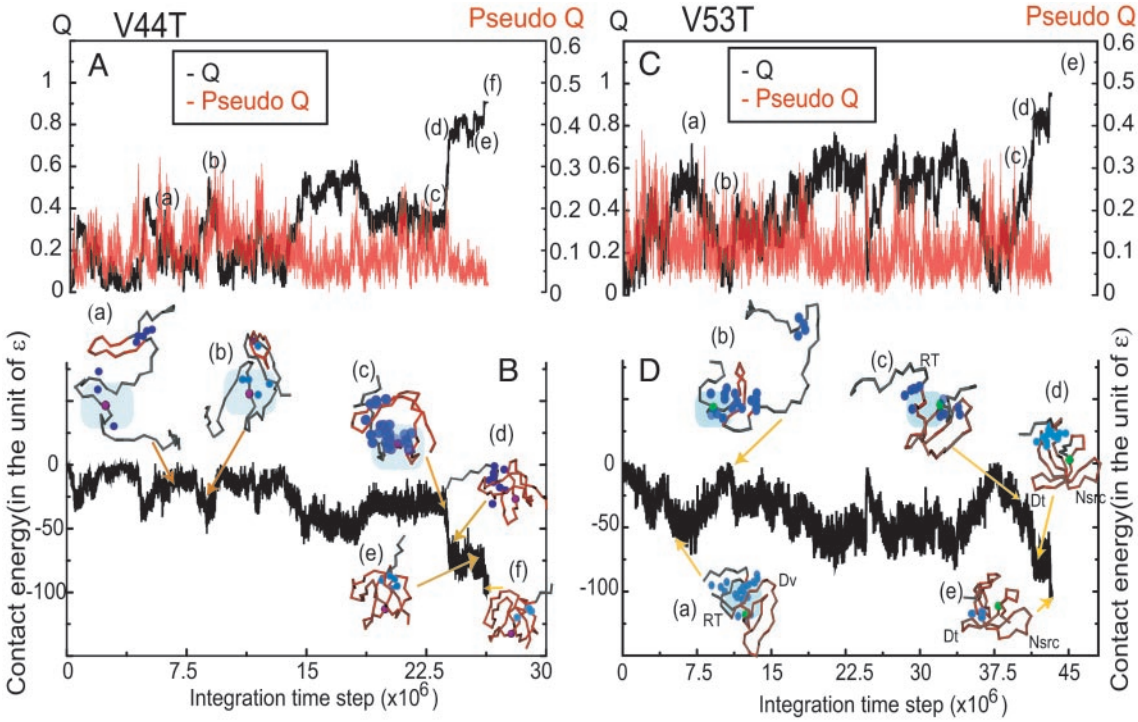

Fig. 5. Typical folding trajectories for Q (the fraction of native contact formation), pseudo Q (the fraction of native single-water separated contact formation), and contact energies as a function of integration time steps, providing molecular details of kinetic traps controlling folding of V44T and V53T. Folding simulations of V44T are shown in *A* and *B*, and those of V53T are shown in *C* and *D*. Both calculations are carried out at 0.9 T_f, where T_f is the folding temperature shown at Table 3 (*Supporting Text*). In the snapshots at both *B* and *D* (where contact energies are plotted as a function of time steps), blue spheres are used to identify single-water separated contacts (notice that unfolded residues without blue spheres are fully solvated). We color residues at native conformations in red to identify folded regions and the unstructured residues in gray. (*B*) For V44T, position 44 is colored purple for visual guidance, and snapshot (a) shows a typical unfolded configuration where short-range contacts are formed. Snapshot b shows a typical kinetic trap where the removal of favorable hydrophobic contacts at this position sabotages nascent native contact formation at the β2, β3, and β4 regions. These β sheets are frayed, and their pairing distances to position 44 are beyond the size of a single water (i.e., fully solvated, highlighted in a cyan box). As a consequence, the distribution of contact formation shifts toward the RT loop at the early stage of folding; an alignment for structural-search collapse is impeded, and kinetic traps are formed. Snapshots c and d correspond to the structural-search collapse where abundant water molecules are expelled from β2-, β3-, and β4-strands, and the diverging turn. Afterward, position 44 is neatly packed at the center of the aforementioned β sheets. Snapshots e and f depict the final stage of folding where residual water molecules are expelled from the hydrophobic core as terminal β-strands pack against the rest of the protein. (*D*) Next we discuss the folding event of V53T mutant. Position 53 is colored in green to provide visual guidance. Snapshot a is a typical trap state, which shows a collapsed yet seemingly flat conformation with short-range contacts mostly formed at the turn and loop regions. A cyan box highlights position 53, which is exposed to a fully solvated yet unformed hydrophobic core. Snapshot b shows a typical unfolded state where position 53 is fully solvated. Snapshots (c and d) indicate a structural-search collapse in which position 53 is exposed to the hydrated hydrophobic core (as illustrated by a cyan box) where water molecules are found between β4 and the RT loop. At the last stage of folding, the desolvation of hydrophobic core takes place; water molecules are expelled from this core, and V53T folds to its native state (e). Simulations suggest that these two positions play distinct topological roles, accounting for the decreases of folding rates observed experimentally.

Protein Engineering Analysis. An interpretation of $\phi_{\ddagger\text{-}U}$ values can be revealed by two simple cases: $\phi_{\ddagger\text{-}U} = 1$ and $\phi_{\ddagger\text{-}U} = 0$ (36–39). $\phi_{\ddagger\text{-}U} = 1$ means that interactions are already formed at TSE, whereas $\phi_{\ddagger\text{-}U} = 0$ represents otherwise. However, it is difficult to interpret intermediate $\phi_{\ddagger\text{-}U}$ values, and several reporters probing the same region are needed to give an overall understanding of the TSE. In this regard, we describe a schematic free energy diagram (Fig. 4) by using two mutations, such as a Val→Thr and Val→Ala mutations. As discussed before, mutating a Val to a Thr residue should introduce a minor energetic perturbation in the folded state, whereas the denatured state is relatively stabilized. The opposite applies to the Ala mutants, where the destabilization should be larger in the folded state following a standard protocol (21) in which the interpretation of data is straightforward.

More attention is needed when discussing the Val→Thr mutation, provided the perturbation is now applied to the unfolded state instead. Together with the aforementioned argument on the Val→Ala mutation, an analysis of $\phi_{\ddagger\text{-}U} = 1$ and $\phi_{\ddagger\text{-}U} = 0$ can now be argued as the following: Imagine a void is formed in the hydrophobic core at the TSE. The core is causatively expanded by this cavity, and interactions inside the core are either weakened or lost, by which the $\phi_{\ddagger\text{-}U}$ values of a Val→Ala mutation can be scaled responsibly between 1 and 0. However, under the same assumption of no water molecules being found inside the core, the Val→Thr mutation should not introduce as much of a perturbation to the TSE as that of a Val→Ala one, because Val and Thr are isosteric, such that both make a similar amount of contacts. Therefore, if the expanded core at the TSE is indeed a void or a cavity (i.e., an expanded dry core), the $\phi_{\ddagger\text{-}U}$ value of a Val→Thr mutation must approach 1 (the only difference in energy between the unfolded and TSE states).

If, on the other hand, there are water molecules accessing the interior region of the protein, causing an expansion and loss of contact in the hydrophobic core, we could expect a different un/folding behavior between a Val and a Thr. In this case, a Thr should stabilize the TSE by forming hydrogen bonds to the buried water molecules. If there are water molecules in the core region (i.e., an expanded wet core) at the TSE, the $\phi_{\ddagger\text{-}U}$ value of the Val→Thr mutation will become closer to 0. In the extreme case when the solvation shell around the Thr side chain is similar to that in the unfolded state, it will be 0. In addition, this value will fall between 0 and 1 instead, if a Thr is only partially solvated. Recall that in the previous paragraph, a Val→Ala mutation can also present intermediate $\phi_{\ddagger\text{-}U}$ values, regardless of the assump-

tion that there is either a cavity (i.e., dry) or water molecules (i.e., wet) inside the expanded core region. Nevertheless, because the philosophy behind the design of these two mutations is different, their immediate $\phi_{\ddagger\text{-U}}$ values are not obligated to converge to the same value. In Fig. 4, we summarize the above argument in a schematic way.

An integrated interpretation of data generated by multiple protein engineering is now quite clear: (*i*) A combination of the $\phi_{\ddagger\text{-U}}$ value of 1 for a Val→Thr mutation and an intermediate $\phi_{\ddagger\text{-U}}$ value for a Val→Ala mutation means an expanded hydrophobic core with a void inside (i.e., an expanded dry core). (*ii*) The $\phi_{\ddagger\text{-U}}$ value of 0 for both mutations means that the position is fully solvent-exposed. (*iii*) Intermediate $\phi_{\ddagger\text{-U}}$ values of both mutations at the same position should address a certain amount of solvation, indicated by partially stabilized denatured conformations in a Val→Thr mutation as well as by a loss of interactions in a Val→Ala mutation.

By comparing the coupled $\phi_{\ddagger\text{-U}}$ values for each mutant pair (V/A, V/T) in Table 1, we found there are noticeably large differences in the Ala and Thr mutants at positions 46 and 58. The $\phi_{\ddagger\text{-U}}$ values of the V46A and V46T mutants are >1. The $\phi_{\ddagger\text{-U}}$ value of the former is quite close to 1 (within the experimental error), implying this region is fully folded in the TSE, whereas the $\phi_{\ddagger\text{-U}}$ value of the latter is quite high as an indication of nonnative interactions made by the Thr side chain in the TSE. As for V58, a replacement of Thr introduces a new hydrogen bonding to the main chain of A55, which provides an explanation for apparent differences between mutations V58A and V58T in their $\phi_{\ddagger\text{-U}}$ values. Although the valines that comprise the hydrophobic core appear to be spatially indistinguishable in the folded state, they have a distinct distribution of $\phi_{\ddagger\text{-U}}$ values, which implies different local solvating conditions in the TSE (excluding V46T that forms nonnative contacts in the TSE).

For the $\phi_{\ddagger\text{-U}}$ values of Ala mutants (which are the standard mutations in the protein engineering analysis), we find that both 44 and 53 positions have slightly higher intermediate $\phi_{\ddagger\text{-U}}$ values, implying they are more structured at the TSE. Their neighboring 23 and 58 positions that deviate from the center have slightly lower intermediate $\phi_{\ddagger\text{-U}}$ values, indicating more loosened regions. This is in perfect agreement with the overall view of src (40) and spectrin SH3 TSE (25). Comparing the $\phi_{\ddagger\text{-U}}$ values of Thr mutants, which probe the solvation of the TSE, we find that position 9 is fully solvent-exposed [in a good agreement with previous data showing this region of the protein to be unfolded (25, 26)]. Position 58 is difficult to interpret due to an extra hydrogen bond made by the Thr side chain and thus cannot be analyzed. A similar argument can be applied to position 46, which indicates a nonnative interaction made by Thr in the TSE. Interestingly, with the analyses on both the Ala and Thr mutations done for the remaining positions (9, 23, 44, and 53), a solvated hydrophobic core is probed by these valines that are exposed to the solvent differentially in accordance with their relative positions in a polymeric chain.

Simulations and Theoretical Analysis. Folding simulations of various SH3 mutants that undergo selective Val→Thr perturbations were performed concurrently, but without any knowledge, with the experimental studies (simulations using minimalist models are published as supporting information on the PNAS web site). Only after these simulations were completed were they compared to the experimental results. Both the standard LJ and desolvation potentials were used to examine the kinetic consequences of the core mutations. Simulations using only the LJ potential do not show sufficient changes when compared to the WT, indicating that a model able to describe how water affects the core packing is needed. Thus, when the desolvation potentials were used to describe core residue–residue contacts, the folding rates of V44T and V53T decreased dramatically, whereas

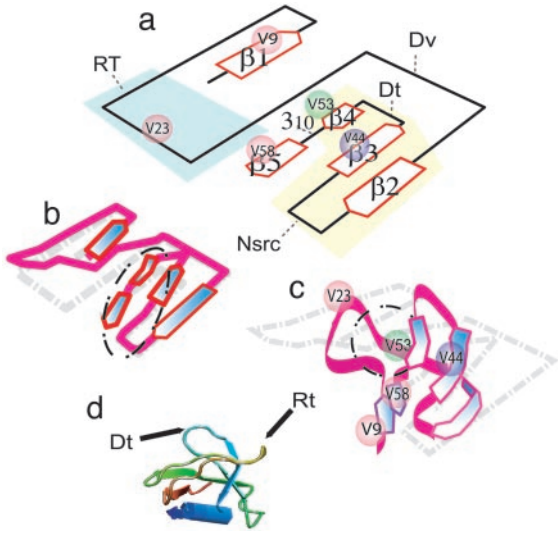

Fig. 6. A schematic folding route that underlies the folding mechanism of SH3. This study supports a general explanation for the folding mechanism in which the structural collapse is followed by desolvation of the hydrophobic core. The former is dominated by the formation of short-range contacts (*a*), whereas the latter is controlled by the closure of interacting β sheets (*b–d*). Replacing valines with threonines in the hydrophobic core region one at a time probes whether SH3 folding dynamics is relevant to a lubricated hydrophobic core. Those valines of interest are represented by colored spheres: V44, purple; V53, green; others, pink. Decreases of folding rates in V44T and V53T can be explained by using this simple diagram: For V44T, removal of favorable contacts to position 44, which is essential to stabilize contacts between β2, β3, and β4 (shown in a yellow block in *a*) results in shifting the distribution of contact formation toward the RT loop (shown in a cyan block); kinetic traps are observed. For V53T, position 53 is involved with many long-range contacts such that it plays an important role in stabilizing the core and terminal β-sheets in *c*. Removal of favorable hydrophobic contacts at this position leads to kinetic traps that appear at a later stage (than a V44T), where desolvation of the hydrophobic core becomes relevant.

kinetic rates of other Thr mutants remain basically unchanged. A quantitative description of these effects is shown in Table 3 and Fig. 7, which are published as supporting information on the PNAS web site. Both rates are substantially slow for both mutants compared to WT, in particular for the V44T. In addition, the calculated folding times are only a lower bound of the full result, because for both of these mutants, many runs are unable to fold in the maximum allowed folding time (runs do not fold 58% of the time for V44T and 44% for V53T). This prediction is in excellent agreement with the experimental observations, as shown in Fig. 3, where the folding of both V44T and V53T is much slower than in other mutants.

Molecular Details Behind Possible Kinetic Traps in V44T and V53T. It is encouraging to learn that the experimental results already support theoretical predictions. More challenging and detailed questions can now be asked: How does this solvated hydrophobic core look in a molecular resolution? What roles do these valines play in the dynamics of the hydrophobic core? Can we obtain a microscopic understanding of why a point mutation of Val→Thr can account for such substantially distinct macroscopic folding behavior? Looking at simulations in detail, molecular interpretations of the cause of kinetic traps inherent in V44T and V53T mutants are now possible. Simulations suggest that, although both show similar dramatic decreases in folding rates, the physical nature of two kinetic traps is very different (Fig. 5).

Folding simulations of V44T mutant (Fig. 5) show that contacts associated with V44 are essential for the structural collapse initiated

by $\beta 2$, $\beta 3$, and $\beta 4$ at the early stage of folding (defined in the yellowish block in Fig. 6*a*). For the V44T mutant, removal of favorable hydrophobic contacts associated with position 44 not only sabotages the structural formation of the aforementioned β sheets (i.e., $\beta 2$, $\beta 3$, and $\beta 4$) at the early stage of folding but also shifts the population of contact formation toward the RT loop (defined in a cyan box in Fig. 6*a*). (Val and Thr should have the same favorable contacts; however, their hydrophobic effect is different. The same nomenclature that differentiates van der Waals interactions from the hydrophobic effects should be contained.) As a consequence, proper alignment of experimental high ϕ-value residues responsible for the following structural-search collapse becomes less favorable, which gives rise to energetic traps causing a decrease of folding rate. In the folding simulations of the V53T mutant, contacts associated with position 53 at $\beta 4$ are mostly long-range and essential for forming the hydrophobic core at the late stage of the folding (position 53 is shown in a green sphere in Fig. 6*c*). Removal of these favorable long-range hydrophobic contacts results in stabilizing local contacts between turns and intrahairpins, causing kinetic traps that lower the folding rates with respect to the WT.

The successful prediction of the kinetic observations for the Thr mutants supports our simple idea of how the Val→Thr mutation perturbs the folding mechanism. Although all five valines of interest are located in the core (Fig. 1), perturbing each residue by a reduction of the hydrophobic interactions (while keeping the van der Waals distance the same), they can lead to distinct kinetic behaviors. This behavior is not simply a consequence of topological frustration resulting from the self-avoidance of the polymeric chain, but it is also related to the extent of solvation, which distinguishes this mutation from a Val→Ala one. If perturbations are applied to some core valines (V44 and V53) that are substantially involved in the main folding route (i.e., structural-search collapse followed by desolvation of the hydrophobic core) in Fig. 6, there are significant changes in the folding kinetics. Otherwise, it is not surprising to learn that other core valines (V9, V23, and V58), which are less involved in the main folding route, have no significant change in kinetics on the same type of mutation.

Given the same native pairing interactions for folding simulations, the reason why the profile of desolvation potential can better represent the hydrophobic core dynamics than the LJ one falls in its salient feature of the desolvation barrier that gives a clear cutoff to separate folded and unfolded configurations. This characteristic is particularly important when perturbations done to the system affect the ensemble of unfolded states more than folded ones. Therefore, even though the standard LJ potential has been successfully applied to describe the distribution of experimental ϕ-values for Val→Ala mutations (12), it was not sufficient to address the distribution of folding rates of Val→Thr mutations, where this kind of mutation emphasizes perturbing unfolded states while keeping the structure of core packing sterically unchanged.

Conclusion

By using a combined theoretical and experimental approach, we have managed to obtain an approximation of the description at the microscopic level of the folding process of an SH3 domain. Our results show that desolvation of the hydrophobic core in this protein is a crucial step to overcome the TSE barrier. Although very simple and therefore unable to include all of the molecular details, these energetically minimally frustrated Gō models already capture most of the geometrical features of the folding mechanism. In addition, this model of simplicity leads to a clear understanding of the underlying mechanism, as long as essential physical interpretations are sufficient to capture the global folding behavior.

M.S.C. thanks Dr. Chinlin Guo for stimulating discussions. A.M.F.-E. is extremely grateful to Dr. S. Casares-Atienza for providing technical advice on protein purification. M.C.V. also thanks F. J. Fernandez for helpful discussions. This work was financed by postdoctoral Marie Curie Individual Fellowship HPMF-CT-2000-01068 and by European Union Grant CT-0013 (to A.M.F.-E.). Work in San Diego was funded by the National Science Foundation (Grant MCB-0084797, with additional support from Grants PHY-0216576 and -0225630). Additional computational support was provided by the W. M. Keck Foundation.

1. Eaton, W. A., Muñoz, V., Thompson, P. A., Chan, C. K. & Hofrichter, J. (1997) *Curr. Opin. Struct. Biol.* **7,** 10–14.
2. Telford, J. R., Wittung-Stafshede P., Gray, H. B. & Winkler, J. R. (1998) *Acc. Chem. Res.* **31,** 755–763.
3. Gruebele, M. (1999) *Annu. Rev. Phys. Chem.* **50,** 485–516.
4. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7524–7528.
5. Bryngelson, J. D. & Wolynes, P. G. (1989) *J. Phys. Chem.* **93,** 6902–6915.
6. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997) *Annu. Rev. Phys. Chem.* **48,** 545–600.
7. Shea, J. E. & Brooks, C. L., 3rd (2001) *Annu. Rev. Phys. Chem.* **52,** 499–535.
8. Onuchic, J. N., Nymeyer, H., Garcia, A. E., Chahine, J. & Socci, N. D. (2000) *Adv. Protein Chem.* **53,** 87–152.
9. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8721–8725.
10. Shea, J. E., Onuchic, J. N. & Brooks, C. L., 3rd (2002) *Proc. Natl. Acad. Sci. USA* **99,** 16064–16068.
11. Portman, J. J., Takada, S. & Wolynes, P. G. (2001) *J. Chem. Phys.* **114,** 5069–5081.
12. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000) *J. Mol. Biol.* **298,** 937–953.
13. Guerois, R. & Serrano, L. (2000) *J. Mol. Biol.* **304,** 967–982.
14. Cheung, M. S., Garcia, A. E. & Onuchic, J. N. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 685–690.
15. Klimov, D. K. & Thirumalai, D. (2002) *J. Mol. Biol.* **315,** 721–737.
16. Takada, S., Luthey-Schulten, Z. & Wolynes, P. G. (1999) *J. Chem. Phys.* **110,** 11616–11629.
17. Ueda, Y., Taketomi, H. & Gō, N. (1978) *Biopolymers* **17,** 1531–1548.
18. Hillson, N., Onuchic, J. N. & Garcia, A. E. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 14848–14853.
19. Garcia, A. E., Hillson, N. & Onuchic, J. N. (2000) *J. Prog. Theor. Phys.* **138,** 282–291.
20. Hummer, G., Garde, S., Garcia, A. E., Paulaitis, M. E. & Pratt, L. R. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 1552–1555.
21. Fersht, A. R., Matouschek, A. & Serrano, L. (1992) *J. Mol. Biol.* **224,** 771–782.
22. Pakula, A. A. & Sauer, R. T. (1990) *Nature* **344,** 363–364.
23. Serrano, L., Kellis, J., Cann, P., Matouschek, A. & Fersht, A. R. (1992) *J. Mol. Biol.* **224,** 783–804.
24. Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L. & Serrano, L. (1994) *Biochemistry* **33,** 2142–2150.
25. Martinez, J. C. & Serrano, L. (1999) *Nat. Struct. Biol.* **6,** 1010–1016.
26. Cobos, E. S., Filimonov, V. V., Vega, M. C., Mateo, P. L., Serrano, L. & Martinez, J. C. (2003) *J. Mol. Biol.* **328,** 221–233.
27. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. (1992) *Nature* **359,** 851–855.
28. Pratt, L. R. & Chandler, D. (1980) *J. Chem. Phys.* **73,** 3434–3441.
29. Sheinerman, F. B. & Brooks, C. L., 3rd (1998) *J. Mol. Biol.* **278,** 439–456.
30. Sorenson, J. M., Hura, G., Soper, A. K., Pertsemlidis, A. & Head-Gordon, T. (1999) **103,** 5413–5426.
31. Pande, V. S. & Rokhsar, D. S. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9062–9067.
32. Guo, C., Cheung, M. S., Levine, H. & Kessler, D. A. (2002) *J. Chem. Phys.* **116,** 4353–4365.
33. Lau, D. F. & Dill, K. A. (1989) *Macromolecules* **22,** 3986–3997.
34. Baase, W. A., Eriksson, A. E., Zhang, X. J., Heinz, D. W., Sauer, U., Blaber, M., Baldwin, E. P., Wozniak, J. A. & Matthews, B. W. (1992) *Faraday Discuss. Chem. Soc.* **93,** 173–181.
35. Kabsch, W. (1976) *Acta Crystallogr. A* **32,** 922–923.
36. Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. (1989) *Nature* **340,** 122–126.
37. Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M. & Fersht, A. R. (1990) *Nature* **346,** 440–445.
38. Matouschek, A., Serrano, L. & Fersht, A. R. (1992) *J. Mol. Biol.* **224,** 819–835.
39. Matouschek, A., Serrano, L., Meiering, E. M., Bycroft, M. & Fersht, A. R. (1992) *J. Mol. Biol.* **224,** 837–845.
40. Grantcharova, V. P. & Baker, D. (1997) *Biochemistry* **36,** 15685–15692.