Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding

Arash Zarrine-Afsar*†, Stefan Wallin*†§¶, A. Mirela Neculai*‖, Philipp Neudecker*§**, P. Lynne Howell*‖, Alan R. Davidson*§††, and Hue Sun Chan*§††

Departments of *Biochemistry, §Molecular Genetics, and **Chemistry, University of Toronto, Toronto, ON, Canada M5S 1A8; and ‖Program in Molecular Structure and Function, Hospital for Sick Children, Toronto, ON, Canada M5G 1X8

Edited by José N. Onuchic, University of California at San Diego, La Jolla, CA, and approved May 1, 2008 (received for review February 27, 2008)

Many experimental and theoretical studies have suggested a significant role for nonnative interactions in protein folding pathways, but the energetic contributions of these interactions are not well understood. We have addressed the energetics and the position specificity of nonnative hydrophobic interactions by developing a continuum coarse-grained chain model with a native-centric potential augmented by sequence-dependent hydrophobic interactions. By modeling the effect of different hydrophobicity values at various positions in the Fyn SH3 domain, we predicted energetically significant nonnative interactions that led to acceleration or deceleration of the folding rate depending on whether they were more populated in the transition state or unfolded state. These nonnative contacts were centered on position 53 in the Fyn SH3 domain, which lies in an exposed position in a 3_{10}-helix. The energetic importance of the predicted nonnative interactions was confirmed experimentally by folding kinetics studies combined with double mutant thermodynamic cycles. By attaining agreement of theoretical and experimental investigations, this study provides a compelling demonstration that specific nonnative interactions can significantly influence folding energetics. Moreover, we show that a coarse-grained model with a simple consideration of hydrophobicity is sufficient for the accurate prediction of kinetically important nonnative interactions.

double mutant cycles | Fyn SH3 domain | Gō models | HP model | Langevin dynamics

Accounting for the physical interactions governing protein behaviors is key to a fundamental understanding of biological processes. In this endeavor, studies on the folding/unfolding kinetics have proven to be a powerful means for revealing transient yet important conformational states and interactions that are not accessible by equilibrium measurements (1, 2). Empirically, folding of many small, single-domain proteins is well approximated by a two-state (3), cooperative (4) process. This observation suggests that nonnative interactions are not prominent during folding because the presence of strong nonnative interactions would be expected to result in significantly populated intermediate conformations. Accordingly, the folding of many natural single-domain proteins is seen as predominantly driven by native interactions, a process that can be described by a funnel-like energy landscape (5, 6) in which kinetic traps (local minima) are shallow or practically nonexistent along the main folding routes. For these proteins, it appears that evolution has largely designed out nonnative interactions and nonnative topology during folding (7), a feat that has not yet been matched by artificial protein design (8). The predominance of native interactions is the basis of Φ-value analyses for structural characterization of folding/unfolding transition states (9). Taken together with the discovery of a strong correlation between folding rate and native topology (10), these considerations served to justify the application of native-centric, or Gō (11), modeling to protein folding (12, 13). In these models, only native interactions are favored, whereas nonnative interactions are taken to be either neutral or repulsive.

Although this perspective is effective as a first approximation, there are physicochemical limitations on native-centric design because proteins are made up of amino acid residues from a finite alphabet and thus their energetic possibilities are restricted. The existence of nonnative interactions in folding is inevitable because residues that do not contact each other in the folded state are bound to come into contact in the variety of disordered conformations present during the folding process. It seems probable that some fraction of these nonnative contacts, such as those between hydrophobic residues, should be favorable and that they cannot be simply "turned off" in reality as in a Gō model. Native-centric models are ill equipped to capture these effects. Indeed, nonnative interactions have been identified experimentally in the folding pathways of mutants of the α-spectrin SH3 (14) and Fyn SH3 (15) domains as well as that of the bacterial immunity protein Im9 (16). In a similar vein, local propensity for a nonnative backbone conformation was found to be a strong kinetic driving force in the folding of Fyn SH3 domain, indicating that its folding transition state contains a specific nonnative conformation (17). These experimental findings confirmed general predictions from simulations (18–22) and analytical theory (20, 23) that nonnative interactions can accelerate folding under certain conditions and led to the concept of "localized frustration" as a possible basis for the nonnative interactions in the folding intermediate of the Im7 protein (24). However, despite these advances, a direct comparison between theory and experiment on the folding effects of specific nonnative interactions has not been performed.

In this work, we sought to better understand protein energetics and the physical constraints on sequence design by developing a coarse-grained chain model to probe the role of specific, position-dependent nonnative hydrophobic interactions in cooperative folding. For this purpose, we performed simulations of the folding of the Fyn SH3 domain using a model based on a

Author contributions: A.Z.-A., S.W., A.M.N., P.N., P.L.H., A.R.D., and H.S.C. designed research; A.Z.-A., S.W., A.M.N., P.N., P.L.H., A.R.D., and H.S.C. performed research; A.Z.-A., S.W., A.M.N., P.N., P.L.H., A.R.D., and H.S.C. analyzed data; and A.Z.-A., S.W., A.R.D., and H.S.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 3CQT).

†A.Z.-A. and S.W. contributed equally to this work

¶Present address: Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138.

††To whom correspondence may be addressed. E-mail: alan.davidson@utoronto.ca or chan@arrhenius.med.toronto.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0801874105/DCSupplemental.

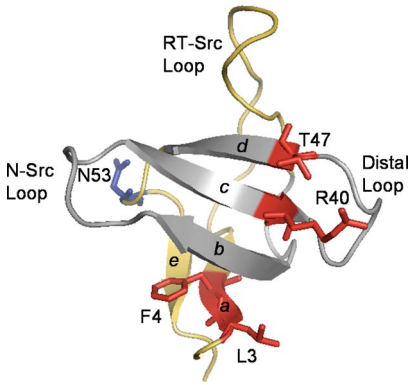

Fig. 1. A ribbon drawing of the WT Fyn SH3 domain (PDB structure 1SHF). The structured region of the folding transition state of the domain as deduced from Φ-value analyses is highlighted in gray. The side chains of N53 (blue) and several other residues (red) investigated in this study are shown.

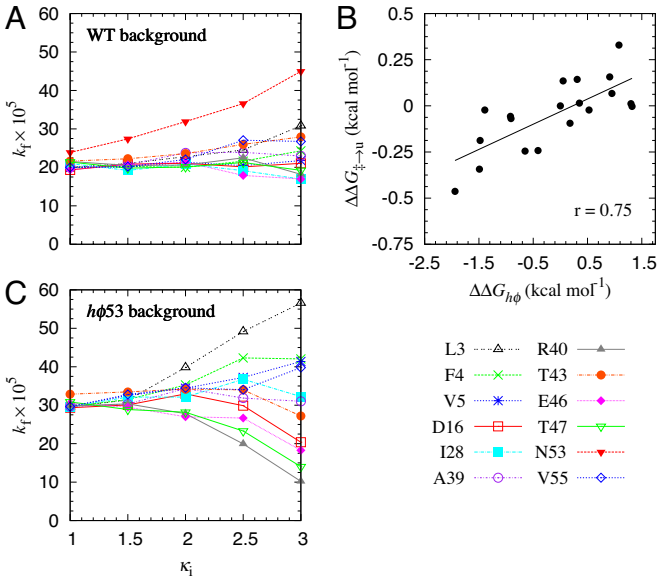

Fig. 2. Effects of increasing the hydrophobicity of specific sites on the folding rate of the Fyn SH3 domain. (*A*) Simulated folding rates k_f in units of reciprocal number of Langevin time steps, as functions of individual hydrophobicity κ_i for 12 amino acid positions. The folding rates here correspond to that of single mutants of WT Fyn SH3 with a hydrophobic residue $h\phi$ of variable κ_i substituted for the WT amino acid residue at the sequence positions indicated (e.g., the N53 curve is for an N53$h\phi$ mutant of WT with variable κ_{53}). (*B*) The correlation between folding transition state stability of N53 mutants (mut) calculated by using the experimental rates (Table S1) and side-chain hydrophobicity. The plotted quantities are normalized by that of the alanine (Ala) mutant, namely, $\Delta\Delta G_{‡\to u} = -RT \ln(k_f^{mut}/k_f^{Ala})$, and the corresponding change in side-chain hydrophobicity $\Delta\Delta G_{h\phi} = \Delta G_{h\phi}^{mut} - \Delta G_{h\phi}^{Ala}$, where $\Delta G_{h\phi}$ is the hydrophobicity values of Fauchère and Pliška (33) and the meanings of the superscripts are as noted above. $\Delta G_{‡\to u}$ corresponds to the height (>0) of the free energy barrier to folding in our notation (17, 30), although only mutational changes of $\Delta G_{‡\to u}$ were considered in this study. (*C*) Simulated folding rates as functions of individual hydrophobicity κ_i for 10 amino acid positions (as indicated) in a $h\phi$53 background with $\kappa_{53} = 2.0$ (e.g., the N53$h\phi$-D16 curve here is for an N53$h\phi$, D16$h\phi$ double mutant of WT with a constant $\kappa_{53} = 2.0$ and variable κ_{16}).

common native-centric construct (13, 25, 26) augmented by simple pairwise hydrophobic interactions similar to that of the hydrophobic polar (HP) model (27), wherein the hydrophobic residues are assigned according to the specific amino acid sequence of the real protein being studied. In this way, we superposed the HP-like component as a "perturbation" on the predominant native-centric Gō model background to capture the physics of sequence-dependent hydrophobic effects, which naturally allow for the possibility of favorable nonnative interactions as in the HP model (27). To test the results of our simulations, we performed *in vitro* folding kinetics studies on mutant versions of the Fyn SH3 domain (Fig. 1). This domain, which consists of two β-sheets packed orthogonally against one another, has proven to be very tractable by *in vitro* protein folding kinetics studies (17, 28–30), and the structure of its folding transition state has been extensively characterized. Using this combined theoretical and experimental approach, we have been able to identify specific nonnative interactions that have significant effects on the folding process. Our findings provide strong support for the proposal that specific nonnative interactions can play a role in both accelerating and decelerating protein folding.

Results

Increased Hydrophobicity at Position 53 Accelerates Protein Folding. To delineate the energetic principles of position-specific nonnative interactions in folding, we used a common Gō-type, native-centric potential for a coarse-grained protein chain model to partially capture the cooperativity achieved by evolutionary design of certain natural proteins and augmented it with an HP-like component that allows for sequence-dependent nonnative hydrophobic interactions (*Materials and Methods*). In this consideration, it is noteworthy that models with only an HP-like potential without a Gō-like component would not be adequate for our purpose because such models lead to noncooperative folding (31, 32).

We used our model to simulate the folding kinetics of WT and mutants of the Fyn SH3 domain. The potential for the formation of nonnative interactions was investigated by simulating the effect of increasing the hydrophobicity of a variety of individual surface and buried positions across the domain (Fig. 2). Strikingly, we found that increasing hydrophobicity (κ_i) of almost all of the positions examined had little effect on the simulated folding rate of the protein (Fig. 2*A*). However, position 53 displayed a dramatic deviation from this trend because the simulated folding rate of the domain increased markedly as a function of increased hydrophobicity at this position. Position 53 is occupied by an Asn residue in the WT Fyn SH3 domain, and it lies at the second (i.e., i+1) position of a three-residue 3_{10}-helix. Because this position is highly exposed (80% solvent accessibility) in the native state of the domain, it is surprising that hydrophobic residues here would increase the folding rate. Thus, it is likely that nonnative interactions are responsible for this phenomenon.

To corroborate the results of our simulations, we substituted position 53 with all of the naturally occurring amino acids and measured the folding rates (k_f) and unfolding rates (k_u) of each mutant in various concentrations of GuHCl using stopped-flow Trp fluorescence experiments [see supporting information (SI) Table S1 for the complete set of data]. The kinetic chevron plots of Fyn SH3 domain mutants (Fig. S1 *A* and *B*) exhibit linear folding and unfolding arms with no signs of rollover, indicating that the mutants still followed a two-state folding mechanism. The only exception was the N53C mutant, which displayed poor-quality biexponential folding and unfolding traces. Because of this anomalous folding behavior, this mutant was excluded from our analyses. Remarkably, substitutions of large hydrophobic residues at position 53 accelerated the folding rate of the domain, most noticeably with the N53W mutant exhibiting a 3-fold increase and the N53I mutant a >2-fold increase in folding rate. Overall, the $\Delta\Delta G_{‡\to u}$ values of all mutants at position 53, which are calculated by comparing mutant folding rates with that of the WT domain, correlated well (Pearson coefficient $r = 0.75$) with changes in side-chain hydrophobicity ($\Delta\Delta G_{h\phi}$) determined

from an experimental hydrophobicity scale for the amino acids (33) (Fig. 2*B*). Thus, consistent with our simulations (Fig. 2*A*), folding kinetics experiments confirmed that hydrophobic substitutions of position 53 increase the folding rate of the domain. It is notable that the hydrophobic substitutions at position 53 accelerate both the folding and unfolding rates (Table S1) resulting in negative Φ-values for these mutants; and negative Φ-values have been implicated in many studies as an indication of nonnative interactions in the folding transition state.

NMR and X-Ray Crystallographic Analysis of the Structure of Fyn SH3 Domain with an N53I Substitution. To determine whether the mechanism of folding rate acceleration by hydrophobic substitutions at position 53 was due to nonnative interactions, it was necessary to ensure that hydrophobic substitutions at position 53 do not significantly alter the structure of the native state. For this reason, we recorded [^{1}H,^{15}N] HSQC spectra of the WT and N53I mutant Fyn SH3 domain. As shown in Fig. S2*A*, the [^{1}H,^{15}N] HSQC spectrum of N53I mutant is minimally perturbed compared with WT, with significant resonance frequency shifts being highly localized to the site of mutation (position 53), as well as the neighboring residues. Furthermore, no peak shifts were observed for the residues of the hydrophobic core, strongly implying that the I53 side chain is not buried in the core of the protein. Thus, we conclude that no dramatic change in the native state structure of the domain is caused by introducing a hydrophobic substitution at position 53. In addition, an x-ray crystal structure of a Fyn SH3 domain mutant containing the N53I substitution (PDB structure 3CQT) shows no significant alteration of the native structure (rmsd value of 0.63 Å for 57 C_α residues aligned); the Ile side chain lies in the identical exposed position as the Asn side chain in the WT structure (Fig. S2 *B* and *C* and Table S2).

Identification of Nonnative Contacts with Position 53. To determine the mechanism by which hydrophobic substitutions at position 53 accelerated folding, we sought to identify residue-specific interactions with this position in our folding simulations. Our first approach involved using a previously described method (25) to identify a subset of simulated conformations around the peak of a model protein's free energy profile that constituted the folding transition ensembles (TSE) for domains containing low hydrophobicity ($l\phi$53) or high hydrophobicity ($h\phi$53) at position 53 (Fig. 3*A* and *B*). A comparison of the residue–residue contact probability maps for these two TSEs shows that increasing hydrophobic strength at position 53 led to enhanced probabilities of contacts between this position and residues 2–6 (strand *a*), 25–28 (strand *b*), and 37–39 (strand *c*) (Fig. 3*C*). As is seen in Fig. 1, these contacts are absent in the native state structure. We also performed further simulations similar to those shown in Fig. 2*A* in which we tested the effects of increased hydrophobicity at a variety of positions in the context of a $h\phi$53 domain. As expected from the contact map result, we observed that increased hydrophobicity at positions 3, 4, and 5 combined with $h\phi$53 led to increased folding rates (Fig. 2*C*), thus predicting that interactions between these positions and position 53, which are not seen in the native structure, would accelerate folding. This contrasts with the situation in a $l\phi$53 background where increasing hydrophobicity at positions 3 and 4 did not affect the folding rate (Fig. 2*A*). A surprising result of our simulations in the $h\phi$53 background was that increased hydrophobicity at positions 16, 40, 43, and 47 caused a marked decrease in the folding rate, suggesting that nonnative interactions (these positions do not contact position 53 in the native state) between these positions and position 53 would lead to a deceleration of the folding rate.

To confirm our predictions of energetically important nonnative contacts with position 53, we carried out a series of experiments on the N53I mutant of the Fyn SH3 domain as a representative for the $h\phi$53 background used in the simulations.

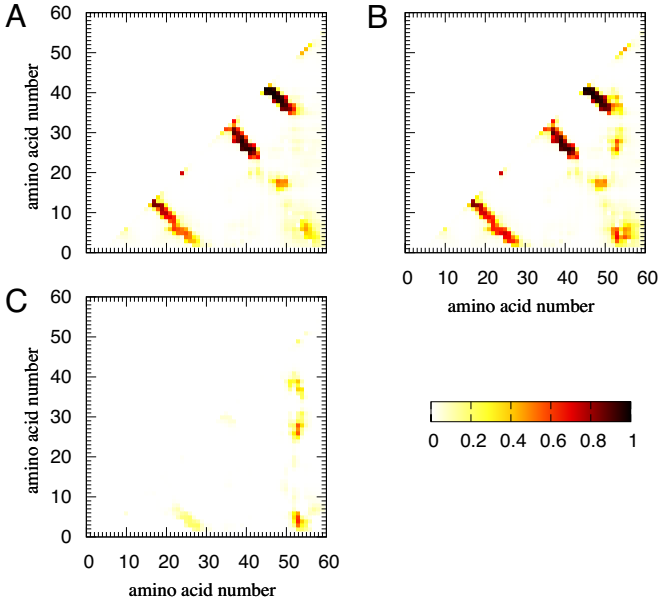

Fig. 3. Simulated contact probability maps of the folding transition state ensemble of N53 mutants of Fyn SH3 domain. (*A*) Contact probabilities for $\kappa_{53} = 1.0$. (*B*) Contact probabilities for $\kappa_{53} = 3.0$. (*C*) Differences between *A* and *B*, i.e., contact probabilities for $\kappa_{53} = 3.0$ minus that for $\kappa_{53} = 1.0$, highlighting amino acid positions interacting more favorably with $h\phi$53 as its hydrophobicity (κ_{53}) increases. In these maps, two residues, *ij*, are considered to be in contact if $r_{ij} < 8$ Å, irrespective of whether the contact is native or nonnative.

We investigated the interactions between I53 and positions 3, 4, 40, and 47 using double mutant thermodynamic cycles based on folding rates, which probe the energy gap between the unfolded state and the folding transition state structure ($\Delta\Delta G_{\ddagger\to u}$). As illustrated in Fig. 4*A*, substitution of Leu-3 with Ala in the presence of Ile at position 53 caused a 0.28 kcal·mol^{-1} destabilization of the transition state structure with respect to the

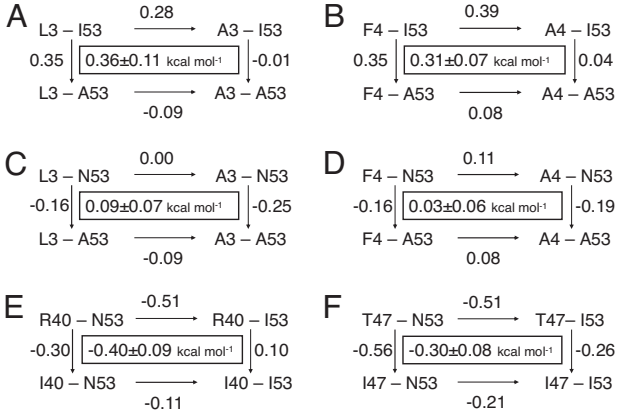

Fig. 4. Thermodynamic cycles to demonstrate nonnative interactions in the folding transition state of N53 mutants. The number along each arrow is the experimentally measured change in the transition-to-unfolded free energy difference, $\Delta\Delta G_{\ddagger\to u}$, in units of kcal·mol^{-1}, resulting from the mutation indicated by the arrow (see Fig. 2 legend). Each boxed value corresponds to the interaction energy (ΔG_{int}), which is equal to the $\Delta\Delta G_{\ddagger\to u}$ value for the vertical arrow on the left minus that on the right, with uncertainty estimated by using the method in the text and *SI Methods*. $\Delta G_{int} > 0$ or $\Delta G_{int} < 0$ is indicative, respectively, of stabilization or destabilization of the folding transition state. The double-mutant cycles here probe interactions between L3 and I53 (*A*), F4 and I53 (*B*), L3 and N53 (*C*), F4 and N53 (*D*), positions 40 and 53 (*E*), and positions 47 and 53 (*F*).

BIOPHYSICS

unfolded state whereas the same substitution tested in a background with Ala at position 53 caused a small increase in the transition state stability. These data indicate that an interaction between L3 and I53 stabilizes the transition state by >0.3 kcal·mol^{-1}. By the same type of analysis, an interaction between F4 and I53 was also seen to stabilize the transition state by a similar amount (Fig. 4*B*). In contrast, cycles probing the N53–L3 and N53–F4 interactions demonstrate that little interaction energy exists between these residues in the transition state structure (Fig. 4 *C* and *D*).

The prediction that interactions between hydrophobic residues at position 53 and position 40 or 47 decelerate the folding rate was tested by using Ile substitutions at these positions. As can be seen in Fig. 4*E*, substitution of the WT Arg residue at position 40 with Ile stabilizes the transition state by 0.3 kcal·mol^{-1} in the context of the WT Asn residue at position 53. This stabilization can be accounted for by the increase in β-strand propensity caused by this substitution, which is in a position that adopts a native β-strand conformation in the transition state (30). Strikingly, when the same R40I substitution was made in combination with the N53I substitution, a destabilization of 0.1 kcal·mol^{-1} was observed, indicating a negative synergy between these positions of 0.4 kcal·mol^{-1} after β-strand propensity effects (a feature not modeled in our simulations) were factored out. In other words, the presence of Ile at position 53 negates the stabilization caused by the R40I substitution in the WT context and adds an additional 0.1 kcal·mol^{-1} of destabilization, strongly supporting a decelerating effect on folding caused by the interaction of hydrophobic residues at these two positions. Negative synergy of a similar magnitude of 0.3 kcal·mol^{-1} is seen between Ile substitutions at positions 47 and 53 (Fig. 4*F*), again consistent with our model prediction (Fig. 2*C*).

Although the interaction energies detected in our thermodynamic cycles are not large (0.3–0.4 kcal·mol^{-1}), these values are substantial relative to the total amount of ≈0.5 kcal·mol^{-1} stabilization of the transition state caused by the N53I substitution, and rigorous analysis of the errors in our measurements and calculations shows that these values are highly significant (*SI Methods*). Buttressing this point is the fact that for folding of the WT protein lacking a hydrophobic residue at 53, the extracted interaction energies between N53 and L3 or F4 were ≈0 (Fig. 4 *C* and *D*). Indeed, thermodynamic cycles for other noninteracting residues we previously investigated have also resulted in interaction energy values that were very close to zero (30), providing further support for the ability of these cycles to detect small interaction energies. To corroborate the kinetics data, Fig. S3 shows that the $\Delta\Delta G_{f\rightarrow u}$ values of mutants used in thermodynamic cycles are strongly correlated ($r = 0.96$) with their ΔT_m values. Analysis of the kinetic m_{kf} values reported in Table S1 shows that the kinetic m_{kf} values of mutants used in double mutant thermodynamic cycles deviated only slightly from WT value with an average deviation from WT value of 6%, with most deviating by 5% and none deviating by >12%. Therefore, none of the mutants possessed grossly altered transition state structures, at least in terms of compactness (34).

Discussion

In this study, we have used a combination of computer simulations and *in vitro* folding studies to identify specific nonnative interactions that significantly affect the folding rate of the Fyn SH3 domain. In particular, we have shown that Leu at position 3 or Phe at position 4 and Ile at position 53 interact to narrow the gap between the unfolded state and transition state by ≈0.3 kcal·mol^{-1} (Fig. 4*A*) even though the β-carbon atoms of these residues are separated by >15 Å in the native state of the domain. On the other hand, a specific interaction between Ile at position 40 or 47 and Ile at position 53, which are separated by ≈20 Å in the native structure, widened the energetic gap between the unfolded state and the folding transition state by 0.3–0.4 kcal·mol^{-1} (Fig. 4*E*). Because of the extreme distance between these positions, only the invocation of nonnative interactions during the folding process can explain these results.

An intriguing aspect of our data is that position 53 appears to be unique among positions tested in mediating nonnative interactions when substituted with hydrophobic residues. Although we have not experimentally tested every position in the Fyn SH3 domain to detect similar effects, extensive mutagenesis studies on positions 22 (35), positions 41 and 24 (29), and positions 29 and 43 (unpublished results) did not uncover any correlation between increased hydrophobicity and the folding rate. A study on positions 40 and 47 did show some correlation between hydrophobicity and folding rate, but these effects likely represented native interactions because these positions are in β-strands where hydrophobic residues have a higher propensity for this secondary structure (30). This is also a likely reason why the increases in folding rate caused by single hydrophobic substitutions at positions 40 and 47 (R40I and T47I data in Table S1) were not captured by the simulation results in Fig. 2*A* (see above discussion of the nonnative $h\phi53$–$h\phi40$ and $h\phi53$–$h\phi47$ interactions). Therefore, unlike a previous study concluding that hydrophobic substitutions of the surface residues of a protein accelerate the rate of folding in an "unspecific" manner (14), our analysis demonstrates that nonnative hydrophobic stabilization of a folding transition state structure is position-specific. Further illustrating a specific hydrophobic effect, the N53I mutant folded almost two times faster than the N53L substitution, suggesting that specific side chain packing plays a role in the nonnative mechanism of folding acceleration. It should be noted that another possible mechanism by which substitutions at position 53 could increase the folding rate is by increasing the local propensity for 3_{10}-helix formation. However, hydrophobic residues have a low propensity for this secondary structure (36, 37); thus, this explanation is unlikely.

Because position 53 appears to possess unique features in our model for SH3 domain folding, it is worth considering potential explanations for this behavior. Numerous Φ-value analyses and related experimental protein folding kinetics studies (28, 38–40) have defined the structure of the folding transition state of the SH3 domain as being highly polarized with strands *b*, *c*, and *d* (the "front sheet" in Fig. 1) adopting a highly native-like structure whereas strands *a* and *e* (the "back sheet" in Fig. 1) are mostly unstructured. Position 53 lies in a 3_{10}-helix, which connects these two sheets, and can be seen as a hinge point in the topology of this domain, likely possessing some degree of flexibility in the folding transition state. Positions 3 and 4 definitely do not make stabilizing native interactions in the folding transition state as demonstrated in this study (i.e., these positions have low Φ-values, see Table S1) and others (28, 41). Thus, it can be imagined that, in the transition state, residues at positions 3 and 4 in the unstructured strand *a* could come into contact with a flexible position 53. Furthermore, because the twist of the backbone required to bring strand *e* into contact with strand *a* in the native state may be a particularly tricky topological feat, nonnative hydrophobic contacts between strand *a* and position 53 could serve to accelerate this difficult process, which may be partially completed before passage through the transition state. To illustrate this point, an example of a simulated structure of a molecule possessing high hydrophobicity at positions 3 and 53 is shown (Fig. 5*A*). In this structure, which may resemble the transition state, strands *b*, *c*, and *d* adopt a native-like structure, but positions 3 and 53 are able to come into close contact because of the lack of native structure in strands *a* and *e*. Consistent with a role for the 3–53 contact in stabilizing the transition state, contact probability distribution plots indicate that the highest probability for contact between these positions occurs close to the transition state as defined by the free energy profiles also shown in these plots (Fig. 5*A*). Our detection of stabilizing nonnative interactions be-

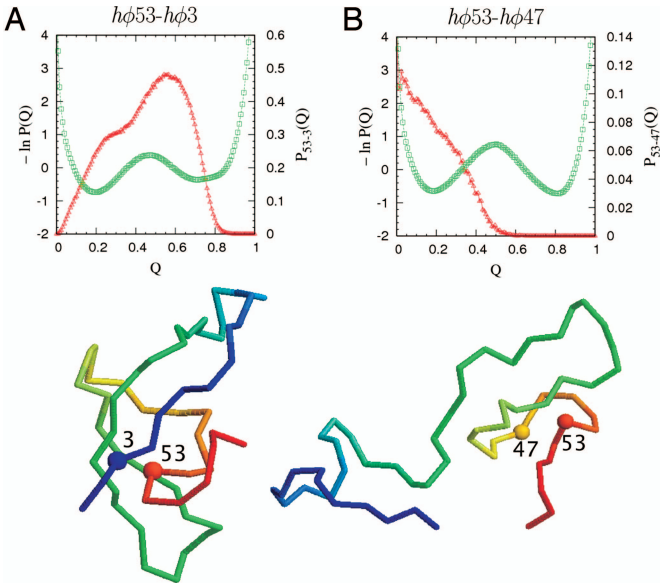

Fig. 5. Populations of specific nonnative hydrophobic interactions along the folding pathway of Fyn SH3 domain. Results shown here are for the $h\phi53$–$h\phi3$ (*A*) and $h\phi53$–$h\phi47$ (*B*) interactions, obtained by simulating the behaviors of two double mutants (e.g., N53I-L3I and N53I-T47I) with $\kappa_{53} = \kappa_3 = 2.0$ (*A*) and $\kappa_{53} = \kappa_{47} = 2.0$ (*B*), respectively, whereas other κ_i values are kept identical to that of WT. The progress variable Q is the total number of native contacts in a conformation normalized by the corresponding number in the fully folded PDB structure. In the determination of Q during simulations (13, 25), two residues, *ij*, are in contact if $r_{ij} < 1.2\ r_{ij}^{n}$, where r_{ij}^{n} is their C_α–C_α distance in the PDB structure. (*Upper*) The probabilities $P_{53-3}(Q)$ and $P_{53-47}(Q)$ (red curves, right vertical scales) of forming the 53–3 (*A*) and 53–47 (*B*) nonnative contacts, respectively, among conformations of a given Q are superposed with free energy profiles of the corresponding proteins at approximately their transition midpoints (green curves, left vertical scales). Here $P(Q)$ denotes the probability distribution of native contact; hence $-\ln P(Q)$ is free energy in units of k_BT, where k_BT is Boltzmann constant times absolute temperature (25). (*Lower*) Example structures in the conformational ensembles containing the given nonnative interactions; positions for the contacting nonnative pairs are marked. The Q values for these two example conformations are 0.56 (*A*) and 0.21 (*B*).

tween strand *a* and position 53 is consistent with NMR relaxation dispersion studies on a very low populated intermediate species in the Fyn SH3 domain, which also detected nonnative interactions between these regions (15, 41).

Because of the highly structured nature of positions 40 and 47 in the transition state, it is unlikely that the decelerating effects on folding caused by contacts between these positions and position 53 occur in this state. In accord with this idea, simulation of a domain containing high hydrophobicity at positions 47 and 53 shows that the highest probability of contact between these positions occurs in the unfolded state where this nonnative interaction could slow folding (Fig. 5*B*). These data suggest that interactions with position 53 that decelerate folding are mostly formed in the unfolded state, whereas those interactions that accelerate folding are mostly formed in the transition state.

The model used here is a folding kinetics model designed to capture the essential physics of nonnative hydrophobic interactions in disordered (unfolded) and partially ordered (transition-state) conformations without substantial formation of tight native packing. Experimental differences in mutational (28) and temperature (42) effects on folding and unfolding rates suggest strongly that the driving forces for folding and unfolding kinetics can be significantly different (43). Evidently, folding kinetics is more influenced by "classical" hydrophobic effects, as characterized typically by solubility measurement of nonpolar compounds (which is the basis for the pairwise HP potential), whereas unfolding kinetics is rate-limited by the disruption of tight native packing (28, 43). For these reasons, the pairwise hydrophobic interactions in our model should provide predictive information about the kinetic process leading up to the folding transition state, i.e., before tight packing becomes overwhelmingly important. For the same reasons, however, our model may be less adequate for unfolding kinetics and overall folding cooperativity (which are not the focus of this study) because the present approach does not consider many-body interactions that are likely important for modeling tight packing and highly cooperative behaviors (44–48). The simulations and experimental work presented here have therefore focused only on folding rates and not unfolding rates or overall thermodynamic stability.

Our observations are an important addition to the growing list of theoretical and experimental studies that have suggested an important role for nonnative interactions in folding pathways (16). What distinguishes this work from other published studies reporting nonnative interactions in folding is that we have shown that a coarse-grained folding model of the domain can be simply modified to correctly predict position-specific nonnative interactions in a folding transition state, without either implicating the knowledge of a protein transition state structure as restraints or jeopardizing drastically the native-centricity of the model. We have also demonstrated that the energetic consequence of nonnative interactions in a folding pathway can be complex, decelerating or accelerating the folding rate of a protein. Our results provide a contrast to the common belief that nonnative interactions generally stall the folding process and support the idea that negative Φ-values may reliably report on formation of nonnative contacts and are not always a result of experimental artifacts (49). In light of our observations that nonnative interactions can assist in folding, the widely accepted notion that folding pathways are exclusively driven by formation of native interactions needs to be revisited. Of particular interest will be future studies of other small proteins to ascertain the generality of nonnative interactions in protein folding pathways.

Materials and Methods

An HP Addition to the Gō-Type Potential. The native-centric part of the energy function for our coarse-grained C_α chain model, denoted as E_0 here, follows that of refs. 13 and 25. To account for sequence-dependent nonnative hydrophobic interactions, we introduce an additional "hydrophobicity" term:

$$E_{\mathrm{HP}} = -\sum_i \sum_j \kappa_i \kappa_j \exp[-(r_{ij} - \sigma_{h\varphi})^2/2] \qquad [1]$$

where r_{ij} is the C_α–C_α distance between residues *i* and *j* and $\sigma_{h\phi} = 5.0$ Å. With this choice of $\sigma_{h\phi}$, each attractive term in Eq. **1** combines with a corresponding excluded-volume repulsion term in E_0 to create a smooth Lennard–Jones-like interaction with optimal attraction at distances $r_{ij} \approx 5.0$ Å. The summation is over all hydrophobic amino acid pairs *ij* satisfying $i < j - 3$ (as for the pairwise terms in E_0) but excludes the native pairs because native hydrophobic interactions have already been favored by E_0. Here κ_i is the hydrophobicity strength of amino acids at position *i*. The complete energy function of our model is

$$E = E_0 + K_{\mathrm{HP}} E_{\mathrm{HP}} \qquad [2]$$

where K_{HP} is the overall strength of the hydrophobic forces. In the present study, alanine, valine, leucine, isoleucine, methionine, tryptophan, phenylalanine, and tyrosine are considered to be hydrophobic. Thermodynamic sampling and folding kinetic simulations were conducted by using Langevin dynamics with the same model parameters as that in Wallin and Chan (25). Further details are provided in *SI Methods* and Fig. S4.

Experimental Studies. Mutagenesis and protein purification are as described previously (50). Details of folding kinetics experiments, NMR studies, and x-ray crystallography and folding simulations are given in *SI Methods*. To ensure that the small interaction energies ($\Delta\Delta G_{\ddagger\to u}$) extracted through thermodynamic cycles are statistically significant, we have carried out analyses of error propagation by using Monte Carlo simulations. Based on the argument provided in *SI Methods*, we are confident that the error margins for the fitted

parameters k_f, m_{kf}, k_u, and m_{ku} in Table S1 accurately reflect the experimental uncertainties associated with our folding kinetic data.

ACKNOWLEDGMENTS. We thank Prof. Lewis Kay for discussion and a critical reading of the manuscript, Drs. Ranjith Muhandiram and Flemming Hansen for recording NMR experiments, and Faiyaz Notta and Lisa Pell for assistance and advice in setting up crystallization conditions and screens. A.Z.-A. is supported by a doctoral Canada Graduate Scholarship (CGS-D3) from the Natural Sciences and Engineering Research Council of Canada. S.W. acknowledges stipend support from the Swedish Research Council. A.M.N. and P.N. are supported by postdoctoral fellowships from the Canadian Institutes of Health Research. P.L.H. and H.S.C. are Canada Research Chair holders. The research reported here was supported by Canadian Institutes of Health Research grants to P.L.H., A.R.D., and H.S.C.

1. Daggett V, Fersht A (2003) The present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol* 4:497–502.
2. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding ''speed limit.'' *Curr Opin Struct Biol* 14:76–88.
3. Jackson SE (1998) How do small single-domain proteins fold? *Folding Des* 3:R81–R91.
4. Chan HS, Shimizu S, Kaya H (2004) Cooperativity principles in protein folding. *Methods Enzymol* 380:350–379.
5. Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol* 4:10–19.
6. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21:167–195.
7. Kaya H, Chan HS (2003) Origins of chevron rollovers in non-two-state protein folding kinetics. *Phys Rev Lett* 90:258104.
8. Watters AL, *et al.* (2007) The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* 128:613–624.
9. Matouschek A, Kellis JT, Serrano L, Fersht AR (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340:122–126.
10. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277:985–994.
11. Taketomi H, Ueda Y, Gō N (1975) Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int J Pept Protein Res* 7:445–459.
12. Micheletti C, Banavar JR, Maritan A, Seno F (1999) Protein structures and optimal folding from a geometrical variational principle. *Phys Rev Lett* 82:3372–3375.
13. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and ''en-route'' intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298:937–953.
14. Viguera AR, Vega C, Serrano L (2002) Unspecific hydrophobic stabilization of folding transition states. *Proc Natl Acad Sci USA* 99:5349–5354.
15. Neudecker P, *et al.* (2006) Identification of a collapsed intermediate with non-native long-range interactions on the folding pathway of a pair of Fyn SH3 domain mutants by NMR relaxation dispersion spectroscopy. *J Mol Biol* 363:958–976.
16. Morton VL, Friel CT, Allen LR, Paci E, Radford SE (2007) The effect of increasing the stability of non-native interactions on the folding landscape of the bacterial immunity protein Im9. *J Mol Biol* 371:554–568.
17. Di Nardo AA, *et al.* (2004) Dramatic acceleration of protein folding by stabilization of a nonnative backbone conformation. *Proc Natl Acad Sci USA* 101:7954–7959.
18. Li L, Mirny LA, Shakhnovich EI (2000) Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. *Nat Struct Biol* 7:336–342.
19. Treptow WL, Barbosa MA, Garcia LG, Pereira de Araujo AF (2002) Non-native interactions, effective contact order, and protein folding: A mutational investigation with the energetically frustrated hydrophobic model. *Proteins* 49:167–180.
20. Clementi C, Plotkin SS (2004) The effects of nonnative interactions on protein folding rates: Theory and simulation. *Protein Sci* 13:1750–1766.
21. Fan K, Wang J, Wang W (2002) Folding of lattice protein chains with modified Gō potential. *Eur Phys J B* 30:381–391.
22. Cieplak M, Hoang TX (2002) The range of the contact interactions and the kinetics of the Go models of proteins. *Int J Mod Phys C* 13:1231–1242.
23. Plotkin SS (2001) Speeding protein folding beyond the Gō model: How a little frustration sometimes helps. *Proteins* 45:337–345.
24. Sutto L, Latzer J, Hegler JA, Ferreiro DU, Wolynes PG (2007) Consequences of localized frustration for the folding mechanism of the IM7 protein. *Proc Natl Acad Sci USA* 104:19825–19830.
25. Wallin S, Chan HS (2006) Conformational entropic barriers in topology-dependent protein folding: Perspectives from a simple native-centric polymer model. *J Phys Condens Matter* 18:S307–S328.
26. Kaya H, Chan HS (2003) Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: How adequate is native-centric topological modeling? *J Mol Biol* 326:911–931.
27. Dill KA, *et al.* (1995) Principles of protein folding—A perspective from simple exact models. *Protein Sci* 4:561–602.
28. Northey JG, Di Nardo AA, Davidson AR (2002) Hydrophobic core packing in the SH3 domain folding transition state. *Nat Struct Biol* 9:126–130.
29. Northey JG, Maxwell KL, Davidson AR (2002) Protein folding kinetics beyond the Φ value: Using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state. *J Mol Biol* 320:389–402.
30. Zarrine-Afsar A, Dahesh S, Davidson AR (2007) Protein folding kinetics provides a context-independent assessment of β-strand propensity in the Fyn SH3 domain. *J Mol Biol* 373:764–774.
31. Chan HS (2000) Modeling protein density of states: Additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. *Proteins* 40:543–571.
32. Knott M, Chan HS (2006) Criteria for downhill protein folding: Calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity. *Proteins* 65:373–391.
33. Fauchère JL, Pliška V (1983) Hydrophobic parameters Π of amino-acid side-chains from the partitioning of N-acetyl-amino-acid amides. *Eur J Med Chem* 18:369–375.
34. Myers JK, Pace CN, Scholtz JM (1995) Denaturant m-values and heat capacity changes—Relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 4:2138–2148.
35. Mok YK, Elisseeva EL, Davidson AR, Forman-Kay JD (2001) Dramatic stabilization of an SH3 domain by a single substitution: Roles of the folded and unfolded states. *J Mol Biol* 307:913–928.
36. Pal L, Basu G, Chakrabarti P (2002) Variants of 3_{10}-helices in proteins. *Proteins* 48:571–579.
37. Karpen ME, de Haseth PL, Neet KE (1992) Differences in the amino acid distributions of 3_{10}-helices and α-helices. *Protein Sci* 1:1333–1342.
38. Riddle DS, *et al.* (1999) Experiment and theory highlight role of native state topology in SH3 folding. *Nat Struct Biol* 6:1016–1024.
39. Grantcharova VP, Riddle DS, Baker D (2000) Long-range order in the src SH3 folding transition state. *Proc Natl Acad Sci USA* 97:7084–7089.
40. Martinez JC, Serrano L (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* 6:1010–1016.
41. Neudecker P, Zarrine-Afsar A, Davidson AR, Kay LE (2007) Φ-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc Natl Acad Sci USA* 104:15717–15722.
42. Segawa S, Sugihara M (1984) Characterization of the transition state of lysozyme unfolding. II. Effects of the intrachain crosslinking and the inhibitor binding on the transition state. *Biopolymers* 23:2489–2498.
43. Kaya H, Chan HS (2003) Simple two-state protein folding kinetics requires near-Levinthal thermodynamic cooperativity. *Proteins* 52:510–523.
44. Kaya H, Chan HS (2003) Contact order dependent protein folding rates: Kinetic consequences of a cooperative interplay between favorable nonlocal interactions and local conformational preferences. *Proteins* 52:524–533.
45. Ejtehadi MR, Avall SP, Plotkin SS (2004) Three-body interactions improve the prediction of rate and mechanism in protein folding models. *Proc Natl Acad Sci USA* 101:15088–15093.
46. MacCallum JL, Moghaddam MS, Chan HS, Tieleman DP (2007) Hydrophobic association of α-helices, steric dewetting, and enthalpic barriers to protein folding. *Proc Natl Acad Sci USA* 104:6206–6210.
47. Qi X, Portman JJ (2007) Excluded volume, local structural cooperativity, and the polymer physics of protein folding rates. *Proc Natl Acad Sci USA* 104:10841–10846.
48. Wang J, Lee C, Stell G (2005) The cooperative nature of hydrophobic forces and protein folding kinetics. *Chem Phys* 316:53–60.
49. de los Rios MA, Daneshi M, Plaxco KW (2005) Experimental investigation of the frequency and substitution dependence of negative Φ-values in two-state proteins. *Biochemistry* 44:12160–12167.
50. Maxwell KL, Davidson AR (1998) Mutagenesis of a buried polar interaction in an SH3 domain: Sequence conservation provides the best prediction of stability effects. *Biochemistry* 37:16172–16182.