Dramatic acceleration of protein folding by stabilization of a nonnative backbone conformation

Ariel A. Di Nardo[†§], Dmitry M. Korzhnev[†¶‖], Peter J. Stogios[†,††], Arash Zarrine-Afsar[†], Lewis E. Kay[†¶‖], and Alan R. Davidson[†¶§§]

Departments of [†]Biochemistry, [¶]Molecular and Medical Genetics, and [‖]Chemistry, University of Toronto, Toronto, ON, Canada M5S 1A8

Edited by Michael Levitt, Stanford University School of Medicine, Stanford, CA, and approved April 12, 2004 (received for review January 24, 2004)

Through a mutagenic investigation of Gly-48, a highly conserved position in the Src homology 3 domain, we have discovered a series of amino acid substitutions that are highly destabilizing, yet dramatically accelerate protein folding, some up to 10-fold compared with the wild-type rate. The unique folding properties of these mutants allowed for accurate measurement of their folding and unfolding rates in water with no denaturant by using an NMR spin relaxation dispersion technique. A strong correlation was found between β-sheet propensity and the folding rates of the Gly-48 mutants, even though Gly-48 lies in an unusual non-β-strand backbone conformation in the native state. This finding indicates that the accelerated folding rates of the Gly-48 mutants are the result of stabilization of a nonnative β-strand conformation in the transition-state structure at this position, thus providing the first, to our knowledge, experimentally elucidated example of a mechanism by which folding can occur fastest through a nonnative conformation. We also demonstrate that residues that are most stabilizing in the transition-state structure are most destabilizing in the native state, and also cause the greatest reductions in *in vitro* functional activity. These data indicate that the unusual native conformation of the Gly-48 position is important for function, and that evolutionary selection for function can result in a domain that folds at a rate far below the maximum possible.

The energetics involved as proteins progress from a broad ensemble of unfolded conformations to a single native structure are still not thoroughly understood. A widely used approach to study folding is the protein engineering method, which measures the thermodynamic and kinetic effects of single amino acid substitutions (usually with Ala) at many different sites in a protein as a means to define the most structured regions of the folding transition state (1, 2). Mapping the transition state identifies regions that fold more quickly than others, and such knowledge has led to important theoretical advances in our understanding of the folding process (3, 4). Although this method provides a general picture of folding pathways, it does not define the mechanisms by which individual residues can facilitate or retard the folding process. Our approach to study these mechanisms is to investigate the effects of multiple amino acid substitutions at single positions (5–7). In the present study, this approach is used to investigate the role in folding and function of position 48 of the Src homology 3 (SH3) domain, which is one of most conserved positions in this domain (8).

The SH3 domain is well suited for protein-folding studies due to its small size and amenability to biophysical analysis (9–11). Composed of two three-stranded β-sheets packed orthogonally against one another (Fig. 1*A*), SH3 domains function as protein–protein interaction modules in a wide variety of eukaryotic proteins. The structure of the SH3 domain transition state has been extensively characterized by protein engineering studies and other experiments (5, 12–16). Striking features of this structure are its relatively well ordered conformation and distinct polarization. Substitution of residues at the N and C termini of the domain has little effect on the folding rate, whereas substitutions in the central β-sheet and the distal loop β-turn cause large decreases (Fig. 1*A*). The SH3 transition state is also conserved among homologues, even those sharing as little as 33% sequence identity (5, 16), suggesting it must be fairly tolerant to amino acid substitutions. This observation provides a rationale for investigating a variety of substitutions at single positions in the SH3 domain because even drastic substitutions at one site would not be expected to change the overall structure of the transition state (5).

Position 48 of the SH3 domain is occupied by Gly in ≈95% of SH3 sequences even though it lies in a β-sheet, where Gly is normally disfavored (Fig. 1*B* and refs. 17 and 18). Surprisingly, the ϕ and ψ angles at this position averaged from 12 different high-resolution SH3 domain structures are $169 \pm 8°$ and $-178 \pm 7°$, respectively, which lie outside the β-strand region of the Ramachandran plot. The consistent occurrence of these unusual ϕ and ψ angles may provide an explanation for the high frequency of Gly at this position, because this amino acid is most energetically favorable in this backbone conformation. Alternatively, Gly-48 may be conserved because side chains at this position can clash with residues in the nearby functionally important RT-Src loop and can potentially disrupt hydrophobic core packing, or because it could be playing some specific role in the folding pathway.

To determine the cause of the high sequence and structural conservation seen at the Gly-48 position, we analyzed the effects of amino acid substitutions on thermodynamic stability, folding kinetics, and peptide-binding function. We were surprised to discover that Gly-48 mutants were highly destabilized, yet folded at dramatically accelerated rates compared with the WT domain. The behavior of these mutants is unusual, and suggests that a nonnative backbone conformation can stabilize the folding transition state of the SH3 domain.

Materials and Methods

Mutagenesis and Purification. Mutant Fyn SH3 proteins were constructed by using PCR-mediated site-directed mutagenesis and were expressed and purified as described (19). For all experiments except the NMR analysis, proteins were buffered in 10 mM Tris·HCl, pH 8.0/0.2 mM EDTA/250 mM KCl.

Stability, Functional, and Stopped-Flow Kinetic Analysis. Urea-induced denaturation of WT and mutant proteins was monitored by both tryptophan (Trp) fluorescence [Aviv Associates (Lakewood, NJ) model ATF 105 spectrofluorometer] and CD (Aviv Associates 62A DS CD spectrometer) spectroscopy. Data generated in these experiments were fit to standard equations by

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

Abbreviations: SH3, Src homology 3; Trp, tryptophan; CPMG, Carr–Purcell–Meiboom–Gill.

[§]Present address: Centre National de la Recherche Scientifique Unité Mixte de Recherche 8542, Ecole Normale Supérieure, 46 Rue d'Ulm, 75230 Paris Cedex 05, France.

[††]Present address: Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada M5G 2M9.

[§§]To whom correspondence should be addressed at: Medical Sciences Building, Room 4176, 1 King's College Circle, University of Toronto, Toronto, ON, Canada M5S 1A8. E-mail: alan.davidson@utoronto.ca.

www.pnas.org/cgi/doi/10.1073/pnas.0400550101

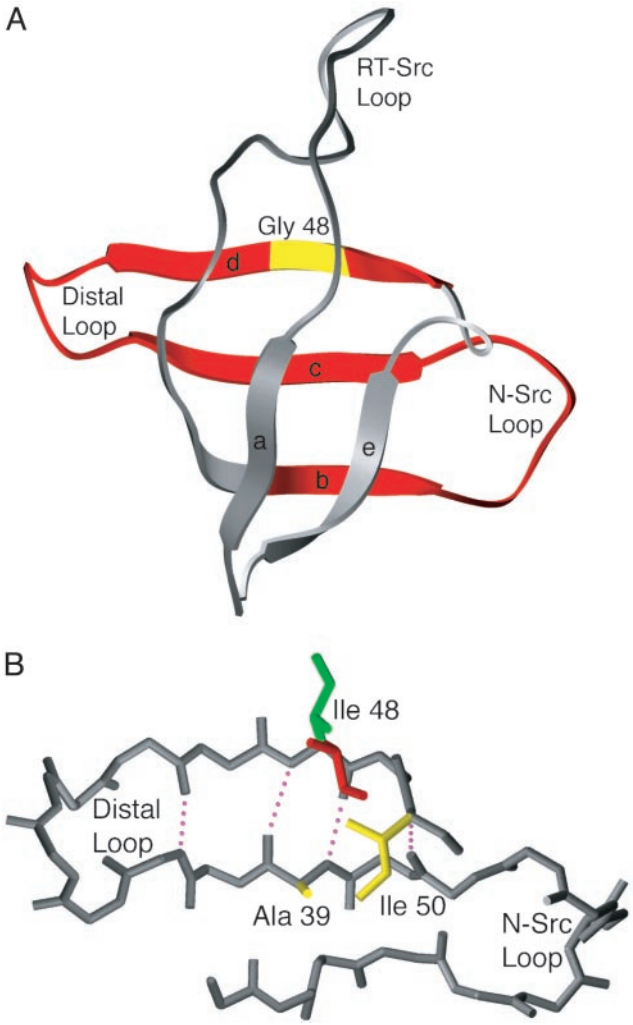

Fig. 1. Location of Gly-48 in the SH3 domain structure. (*A*) Ribbon view of the Fyn SH3 domain [PDB ID code 1SHF (46)] is shown with the Gly-48 position highlighted in yellow, and the β-sheet and loop regions of the domain that are highly structured in the folding transition state in red. The conserved β-strands making up the SH3 domain fold are indicates with the letters a–e. (*B*) View of β-strands b–d of the Fyn SH3 domain. The backbone–backbone H-bonds between β-strands c and d are pink. An Ile side chain has been modeled at the Gly-48 position to show possible orientations for side chains substituted at this position. The direction that the side chain at position 48 would project if the backbone were in its native conformation ($\phi = +168°$ and $\psi = +180°$) is green, and direction that the side chain would project if the backbone at position 48 were in a relaxed β-strand conformation ($\phi = -120°$ and $\psi = +120°$) is red. It can be seen that the red side chain could interact with Ala-39 and Ile-50, which are structured hydrophobic core residues in the folding transition state. Structure figures were created by using SETOR (47).

using the program KALEIDAGRAPH (Synergy Software) as described (19). In some cases where folded or unfolded baselines were not well defined, values were confirmed by manual fitting.

Measurements of peptide binding to VSLARRPLPPLP, a tight binding sequence for the Fyn SH3 domain isolated by phage display (20) were carried out as described (19). The change in the free energy of binding, $\Delta\Delta G_{\text{bind}}$, was calculated as follows:

$$\Delta\Delta G_{\text{bind}} = -RT \ln [K_d(\text{WT})/K_d(\text{mutant})]. \quad [1]$$

The kinetics of folding and unfolding were monitored by Trp fluorescence using a Bio-Logic SFM-4 stopped-flow instrument as described (7). All experiments were performed at 298 K and urea was used as the denaturant. The folding and unfolding data for each mutant at various denaturant concentrations were fit to the following equation by using KALEIDAGRAPH:

$$\ln k_{\text{obs}} = \ln k_f - (m_{kf}[\text{urea}]) + \ln k_u + (m_{ku}[\text{urea}]), \quad [2]$$

where $k_f^{H_2O}$ and $k_u^{H_2O}$ are the folding and unfolding rates in water, respectively, and m_{kf} and m_{ku} are the dependence of $\ln k_f$ and $\ln k_u$, respectively, on the concentration of urea.

NMR Measurements and Data Analysis. Constant relaxation time relaxation compensated Carr–Purcell–Meiboom–Gill (CPMG) experiments (21, 22) were performed for the backbone ^{15}N nuclei of Fyn SH3 domain mutants by using described methodology (23). Data sets were recorded at 25°C at spectrometer field strengths of 600 and 800 MHz (^1H frequency) by using ^{15}N-labeled samples that ranged in protein concentration from 0.5 to 1.0 mM, 10% D_2O, 50 mM sodium phosphate, 0.05% NaN_3, 0.2 mM EDTA, pH 7. The resulting relaxation dispersion profiles $R_2^*(\nu_{\text{CPMG}})$ comprise 14 points for each magnetic field recorded with ν_{CPMG} values ranging from 50 to 1,000 Hz. For each magnetic field, uncertainties in R_2^* were estimated by using spectra recorded twice for at least two values of ν_{CPMG}. To account for possible systematic biases affecting peak intensities, uncertainties in R_2^* were set to at least 2%. Dispersion curves were analyzed assuming a two-state process by using a described procedure (23). For further details see *Supporting Methods*, which is published as supporting information on the PNAS web site.

Derivation of Pseudoenergy Terms from Propensity and Volume Parameters. A β-sheet propensity pseudoenergy term ($\Delta\Delta G_{P\beta}$) for each mutant was derived by using the following formula:

$$\Delta\Delta G_{P\beta} = -RT \ln (P_\beta^{\text{MUT}}/P_\beta^{\text{WT}}), \quad [3]$$

where P_β^{MUT} is the β-sheet propensity value of the substituted residue, and P_β^{WT} is the β-sheet propensity value of the WT residue, which is Gly in this case. The P_β values used here were updated Chou and Fasman type propensities (24) derived from a nonredundant protein database by using DSSP (25) to identify residues in β-sheet regions (26). The relevant P_β values are as follows: Ala (0.75), Gly (0.67), Ile (1.74), Met (1.09), Arg (0.91), Ser (0.87), Thr (1.20), and Val (1.84). A pseudoenergy term ($\Delta\Delta G_{\text{VOL}}$) based on the buried volumes of substituted residues was also derived by using Eq. **3**, except that volume values were used. The relevant side chain volume values in Å^3 are as follows: Ala (92), Gly (66), Ile (169), Met (171), Arg (202), Ser (99), Thr (122), and Val (142) (27).

Results

Equilibrium and Kinetic Folding Studies on the Gly-48 Mutants. To examine its role in the stability and function of the SH3 domain, Gly-48 in the Fyn SH3 domain was substituted with seven different amino acids. The CD spectra of all of the mutants were similar to the unusual spectrum seen for the WT domain (19) with a prominent maximum at 220 nm (data not shown). The thermodynamic stability of these mutants at equilibrium was characterized by determining the free energy of unfolding ($\Delta G_{f\to u}$) by using urea-induced unfolding experiments monitored by Trp fluorescence and by CD (Fig. 5, which is published as supporting information on the PNAS web site). All of the mutants were greatly destabilized compared with WT with $\Delta G_{f\to u}$ value reductions ranging from 2 to 3.5 kcal·mol^{-1} (Table 1). Surprisingly, substitutions with the small Ala and Ser residues caused as much destabilization as substitutions with the larger Val, Ile, and Met residues. The dependence of $\Delta G_{f\to u}$ on denaturant concentration for each mutant ($m_{f\to u}$) was un-

Table 1. Folding and peptide-binding properties of the Gly-48 mutants

	Equilibrium data			Stopped-flow folding kinetics data					
	$\Delta G_{f\to u}$, kcal·mol^{-1}	$m_{f\to u}$, kcal·mol^{-1}·M^{-1}	K_d, μM	$k_f^{H_2O}$, sec^{-1}	m_{kf}*	$k_u^{H_2O}$, sec^{-1}	m_{ku}	$\Delta G_{f\to u}$†, kcal·mol^{-1}	$m_{f\to u}$‡, kcal·mol^{-1}·M^{-1}
WT§	4.43 ± 0.17	0.73 ± 0.01	0.43 ± 0.08	78.0 ± 3.6	0.95 ± 0.02	0.06 ± 0.01	0.30 ± 0.03	4.23 ± 0.13	0.74 ± 0.02
G48V	1.49 ± 0.25	0.68 ± 0.06	17.5 ± 2.7	1038 ± 275	0.94 ± 0.23	55.9 ± 21.1	0.38 ± 0.07	1.72 ± 0.27	0.77 ± 0.14
G48M	2.59 ± 0.32	0.92 ± 0.04	3.0 ± 2.3	453 ± 101	0.80 ± 0.17	17.8 ± 7.8	0.40 ± 0.08	1.91 ± 0.29	0.71 ± 0.11
G48I	1.39 ± 0.09	0.79 ± 0.14	10.1 ± 3.7	585 ± 61	0.51 ± 0.12	35.9 ± 14.7	0.55 ± 0.08	1.65 ± 0.25	0.62 ± 0.08
G48A	1.54 ± 0.01	0.75 ± 0.24	1.7 ± 0.4	126 ± 8.0	0.81 ± 0.05	5.4 ± 0.7	0.36 ± 0.02	1.86 ± 0.08	0.69 ± 0.03
G48R	1.40 ± 0.05	0.73 ± 0.10	1.6 ± 0.5	460 ± 91	0.92 ± 0.21	43.3 ± 10.3	0.38 ± 0.05	1.39 ± 0.18	0.77 ± 0.13
G48S	0.84 ± 0.02	0.74 ± 0.09	3.6 ± 0.3	181 ± 21	0.88 ± 0.13	20.5 ± 3.1	0.36 ± 0.03	1.28 ± 0.11	0.73 ± 0.08
G48T	1.04 ± 0.02	0.72 ± 0.04	7.1 ± 0.1	319 ± 41	0.84 ± 0.18	37.4 ± 9.5	0.43 ± 0.05	1.26 ± 0.17	0.75 ± 0.11

All values are averaged from at least two experimental repetitions. Errors reported for equilibrium data are the SD between the experimental values. The errors reported for $k_f^{H_2O}$, $k_u^{H_2O}$, m_{kf}, and m_{ku} are the fitting errors. Errors shown for other calculated values were propagated by using standard equations.

*m_{kf} and m_{ku} are the dependence of ln(k_f) and ln(k_u), respectively, on the concentration of urea.

†$\Delta G_{f\to u} = -RT\ln(k_u^{H_2O}/k_f^{H_2O})$.

‡$m_{f\to u} = RT(m_{kf} + m_{ku})$

§Our previously reported thermodynamic and kinetic values for the WT Fyn SH3 differ from those reported here (7, 19). This discrepancy is due to a significant stabilization of this domain by GuHCl (A.Z.-A. and A.R.D., unpublished observations), which was the denaturant used in these previous studies.

changed compared with WT except for the G48M mutant, which was more stable than the other mutants and also had a significantly higher m value.

To gain further insight into the mechanisms by which substitutions at the Gly-48 position cause destabilization, the folding and unfolding rates of the Gly-48 mutants were determined using stopped-flow Trp fluorescence. The logarithms of the observed rate constants at a variety of urea concentrations were plotted to determine the folding and unfolding rates in zero denaturant ($k_f^{H_2O}$ and $k_u^{H_2O}$, respectively) for each mutant (Fig. 2*A*). To our great surprise, all of these mutants folded considerably faster than the WT domain even though their overall stabilities were much lower (Table 1). The most dramatic effects were caused by the G48V substitution, which increased the folding rate >10-fold and yet the protein was destabilized by 2.5 kcal·mol^{-1}. The fast folding rates of the Gly-48 mutants were accompanied by even more striking increases in their unfolding rates, as exemplified by the >900-fold increase in the unfolding rate of the G48V mutant. Although the kinetics of the faster folding mutants approached the limit of detection of our stopped-flow instrument, the data are validated by the good agreement between $\Delta G_{f\to u}$ values calculated from both equilibrium and kinetic experiments (a 15% average deviation was observed; Table 1). G48M is the only mutant that shows a large deviation between its equilibrium and kinetic $\Delta G_{f\to u}$ values.

Measurement of Folding and Unfolding Rates by NMR Relaxation Dispersion Experiments. Due to their very fast folding and low stability, the folding rates of the Gly-48 mutants proved difficult to measure using stopped-flow fluorescence experiments. Consequently, the errors in folding rate determinations were quite large, especially for the fastest folding mutants (Table 1). For this reason, NMR relaxation dispersion experiments (21, 22) were also used to determine the folding and unfolding rates of the Gly-48 mutants. By measuring the contributions to ^{15}N line widths in ^{1}H-^{15}N correlation spectra as a function of the applied ^{15}N radio frequency field strength (ν_{CPMG}), the rates of interconversion between folded and unfolded states, along with the differences in ^{15}N chemical shifts between the two states could be quantified (28). This methodology was feasible for the Gly-48 mutants because of their unique combination of low stability and fast exchange between folded and unfolded states, leading to measurable peak broadening in ^{1}H-^{15}N heteronuclear sequential quantum correlation spectra. Particularly important was that folding and unfolding rates of individual amide sites in each

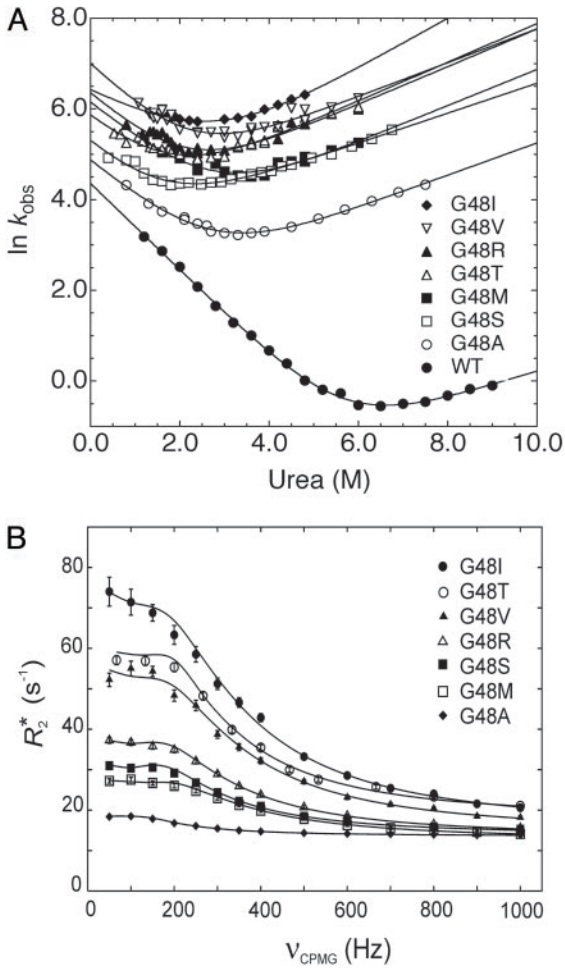

Fig. 2. Kinetic characterization of Fyn SH3 domain Gly-48 mutants. (*A*) Dependence of the rate of folding and unfolding on urea concentration at 25°C of WT and the mutants measured by stopped-flow fluorescence. The data points with an lnk_{obs} of 6 approach the limit of detection because much of the decay or growth of the corresponding kinetic traces is lost in the instrument dead time. The lines shown are the theoretical fits to the data by using the equation shown in *Materials and Methods*. (*B*) Dispersion profiles for Asp-9 of the mutants, recorded at 25°C, 800 MHz (^{1}H frequency). The solid lines are obtained from a global fit of all dispersion data for a given mutant simultaneously, assuming a simple two-state exchange process.

Table 2. The protein folding kinetics of the Gly-48 mutants measured by NMR relaxation dispersion

	$k_f^{H_2O}$, sec^{-1}	$k_u^{H_2O}$, sec^{-1}	$\Delta\Delta G_{\ddagger\to u}$*, kcal·mol^{-1}	$\Delta\Delta G_{f\to u}$†, kcal·mol^{-1}	ϕ_f‡
WT§	78.0 ± 3.6	0.06 ± 0.01	0	0	
G48V	729 ± 8.6	46.9 ± 0.5	−1.32 ± 0.03	2.61 ± 0.13	−0.50 ± 0.03
G48M	516 ± 7.4	15.8 ± 0.1	−1.11 ± 0.03	2.17 ± 0.13	−0.51 ± 0.04
G48I	833 ± 10.7	70.9 ± 1.6	−1.40 ± 0.03	2.78 ± 0.13	−0.50 ± 0.03
G48A	127 ± 14.2	5.0 ± 0.03	−0.29 ± 0.07	2.32 ± 0.15	−0.12 ± 0.01
G48R	439 ± 6.0	25.5 ± 0.1	−1.02 ± 0.03	2.55 ± 0.13	−0.40 ± 0.02
G48S	148 ± 7.3	17.6 ± 0.1	−0.38 ± 0.04	2.97 ± 0.13	−0.13 ± 0.01
G48T	280 ± 8.8	43.0 ± 0.4	−0.75 ± 0.03	3.13 ± 0.13	−0.24 ± 0.01

*$\Delta\Delta G^{\ddagger}_{\to u} = -RT\ln[k_f^{H_2O}(\text{mutant})/k_f^{H_2O}(\text{WT})]$. Note that negative $\Delta\Delta G$ values indicate stabilization compared with WT.

†$\Delta\Delta G_{f\to u} = -RT\ln[k_f^{H_2O}(\text{mutant})/k_f^{H_2O}(\text{WT})]/-RT\ln[k_u^{H_2O}(\text{WT})/k_u^{H_2O}(\text{mutant})]$.

‡$\phi_f = \Delta\Delta G_{\ddagger\to u}/\Delta\Delta G_{f\to u}$

§WT values are from stopped-flow experiments shown in Table 1.

mutant could be measured simultaneously in solution without denaturant. Fig. 2*B* shows dispersion profiles measured for Asp-9 in each of the mutants examined. Dispersion plots from all residues satisfying a set of criteria indicated in *Supporting Methods* were fit to a global two-site exchange process, following the described approach (23), with ≈70% of the dispersion profile well fit collectively to this simple model (Fig. 2*B*). A more detailed analysis of these data will be presented elsewhere.

The $k_f^{H_2O}$ and $k_u^{H_2O}$ values determined by using the relaxation dispersion experiments (Table 2) are determined from a global fit involving many different individual residues in each mutant measured at two different field strengths, and are obtained directly without extrapolation from rates measured in denaturant. Remarkably, the $k_f^{H_2O}$ and $k_u^{H_2O}$ values determined by the NMR experiments generally deviated by <20% from those obtained in the stopped-flow experiments with the exception of the fastest folding mutants, G48I and G48V, which were the most difficult to characterize by stopped-flow techniques. The $\Delta G_{f\to u}$ values determined for each mutant in the two sets of kinetics experiments deviate by <10% in all cases and the values obtained in both experiments correlate well with those determined in the equilibrium denaturation experiments (Fig. 6, which is published as supporting information on the PNAS web site). In further discussion and calculations below, we use the $k_f^{H_2O}$ and $k_u^{H_2O}$ values from the NMR experiments because they are the most accurately determined.

Peptide-Binding Activity of the Gly-48 Mutants. By using an assay based on Trp fluorescence, we measured the binding activity of the Gly-48 mutants to a single high-affinity target peptide previously isolated by phage display screening (20). Despite their low thermodynamic stability, all of the mutants were still able to bind target peptide (Table 1, and Fig. 7, which is published as supporting information on the PNAS web site). One group of mutants (G48A, G48R, G48S, and G48M) displayed K_d values within 10-fold of the WT value, whereas a second group (G48T, G48I, and G48V) show a 15-fold or greater change in K_d values. Interestingly, all β-branched substitutions belong to the second group of mutants, suggesting that binding may be more influenced by the backbone conformational preference of the substituted amino acid rather than its size or charge.

Discussion

The Gly-48 mutants characterized here are remarkable in combining a dramatic acceleration of folding rates with a large degree of destabilization. Typically, unstable mutant proteins display decreased folding rates and/or increased unfolding rates. Although destabilized mutants that fold considerably faster than WT have been previously isolated (7, 29), the Gly-48 mutants present by far the most extreme examples of such behavior, especially considering that they are only single-site substitutions. The use of NMR spin relaxation dispersion experiments to provide accurate measurements of $k_f^{H_2O}$ and $k_u^{H_2O}$ for the Gly-48 mutants in water without denaturant allowed us to eliminate the possibility that the $k_f^{H_2O}$ and $k_u^{H_2O}$ values extrapolated from stopped-flow measurements in urea were aberrantly high due to undetected chevron plot curvature at low denaturant concentrations. The excellent agreement between results of the stopped-flow and NMR folding kinetics experiments strongly validates our findings.

The unusual folding rates of the Gly-48 mutants are not likely to be due to a loss of two-state folding behavior, or to a grossly altered native-state structure. The preservation of two-state folding for these mutants is confirmed by the general agreement between the $\Delta G_{f\to u}$ and $m_{f\to u}$ values determined from both kinetic and equilibrium experiments (Table 1) and by the unfolding behavior of the majority of individual residue positions being consistent with a two-state model in NMR experiments. In addition, there is no sign of curvature in the chevron plots (Fig. 2*A*), which can indicate the formation of a folding intermediate (30). The native structures of the mutants must be close to that of the WT because they still possess peptide-binding activity (Table 1). Furthermore, the ^{1}H-^{15}N-heteronuclear sequential quantum correlation NMR spectra measured for each of the mutants exhibited fully native features (data not shown).

Within the standard folding transition-state framework, in which the protein-folding reaction is modeled as a two-state system with a folded state, an unfolded state, and an energetic barrier (transition state, ‡) in between (2), amino acid substitutions like those at the Gly-48 position that accelerate the folding are presumed to stabilize the transition-state structure with respect to the unfolded state (i.e., $\Delta\Delta G_{\ddagger\to u} < 0$, Fig. 3). The relative effect of a mutation on the stability of the transition-state structure is often expressed as the Φ_f value ($\Delta\Delta G_{\ddagger\to u}/\Delta\Delta G_{f\to u}$) with Φ_f values for individual residues typically ranging from 0 to 1, where a value of 1 indicates that the residue makes the same stabilizing interactions in the transition state as it makes in the folded structure (31, 32). Residues exhibiting negative Φ_f values, as are observed for all of the Gly-48 mutants (Table 2), have been suggested to make nonnative contacts in the transition-state structure (33, 34).

A likely mechanism for the folding transition-state stabilization caused by Gly-48 substitutions can be deduced by considering the structure of the SH3 domain transition state. It has been firmly established that the three-stranded sheet composed of strands *b*, *c*, and *d* and the distal Loop (Fig. 1*A*) comprise the

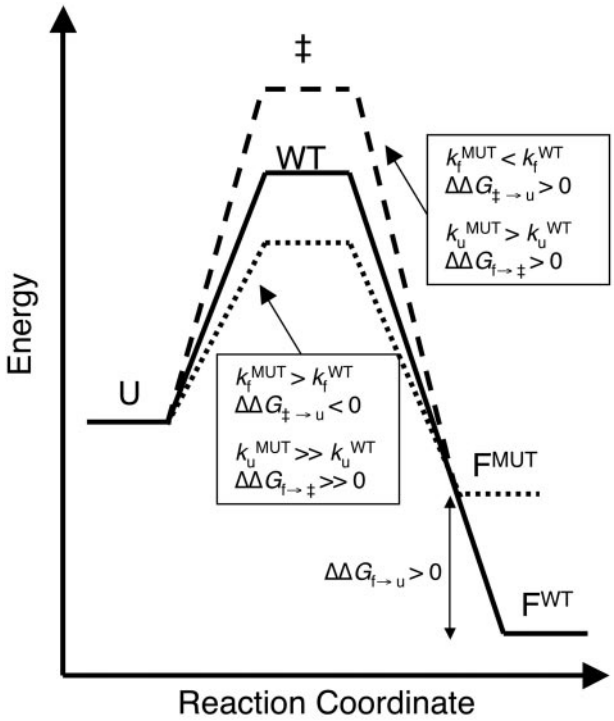

Fig. 3. Energy diagram for mutants affecting protein-folding kinetics. Energy levels of the transition states, unfolded state (U), and the folded states of a WT (F^{WT}) and mutant (F^{MUT}) are shown. The dashed line shows how the energy levels of the transition state and folded state would change in typical destabilizing mutants where the folding rate is slowed and the unfolding rate is accelerated. The dotted line shows how the energy levels of the transition state and folded state are likely altered in the Gly-48 mutants where the folding rate is accelerated over the WT due to stabilization of the transition-state structure relative to that of the WT domain. The folded state is also destabilized, causing the unfolding rate to be extremely fast compared with the WT. It is assumed that mutants do not significantly affect the energetics of the unfolded state.

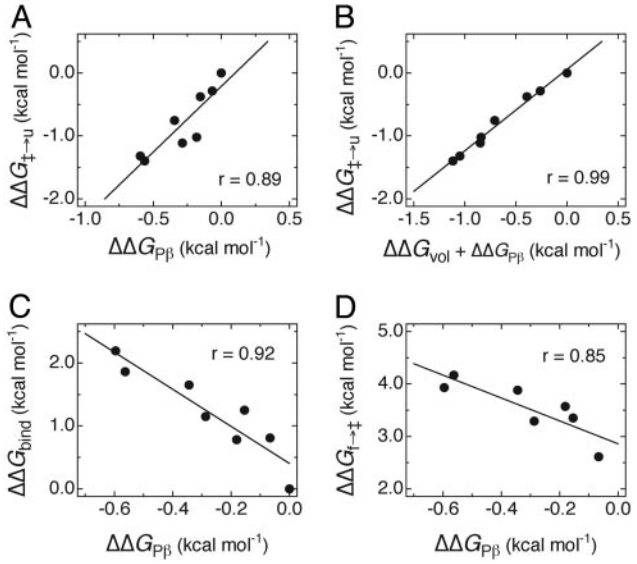

Fig. 4. The correlation between changes in residue β-sheet propensity and volume with changes in folding rates (quantified by $\Delta\Delta G_{‡\to u}$), unfolding rates (quantified by $\Delta\Delta G_{f\to ‡}$), and *in vitro* functional activity (quantified by $\Delta\Delta G_{bind}$) caused by substitutions at Gly-48. (*A*) $\Delta\Delta G_{‡\to u}$ is plotted as a function of the change in β-sheet propensity pseudoenergy ($\Delta\Delta G_{P\beta}$). (*B*) $\Delta\Delta G_{‡\to u}$ is plotted as a function of the sum of the $\Delta\Delta G_{P\beta}$ and the change in residue volume pseudoenergy ($\Delta\Delta G_{vol}$). (*C*) The change in the free energy of binding ($\Delta\Delta G_{bind}$) is plotted as a function of $\Delta\Delta G_{P\beta}$. (*D*) $\Delta\Delta G_{f\to ‡}$ is plotted as a function of $\Delta\Delta G_{P\beta}$. The WT value is not shown in *D* because it is an outlier. Derivation of the pseudoenergy terms and $\Delta\Delta G_{bind}$ is described in *Materials and Methods*. Note that negative $\Delta\Delta G$ values indicate stabilization with respect to the WT.

most highly structured portion of the SH3 domain transition state (7, 14, 16). The location of the Gly-48 within this structured region (strand *d*), explains why substitutions at this position have a large effect on the folding rate. We hypothesize that position 48 is forced into its unusual ϕ/ψ angles by the tight packing of the hydrophobic core, which occurs only in the fully formed native-state structure. In the transition state, where only one β-sheet is well formed and the core is loosely packed (7), the Gly-48 position is likely to be in a standard, relaxed β-strand conformation. Because Gly is relatively unfavorable in β-strand conformation, it follows that substitutions at the Gly-48 position should stabilize the β-strand conformation in the transition-state structure and thus accelerate the folding reaction. Furthermore, residues that are most favorable in β-strand conformation should accelerate folding the most. These hypotheses are confirmed by the strong correlation ($r = 0.89$) between the $\Delta\Delta G_{‡\to u}$ values (values are more negative for faster folders) of each Gly-48 mutant versus a pseudoenergy term based on the β-sheet propensity of the substituted residue ($\Delta\Delta G_{P\beta}$, value is more negative for better β-sheet formers; Fig. 4*A*). Thus, folding of this domain is accelerated by the stabilization of a nonnative backbone conformation in the folding transition state.

Modeling the backbone conformation of the Gly-48 position into a normal β-strand conformation indicates that a side chain at this position would face into the hydrophobic core close to Ala-39 and Ile-50 (Fig. 1*B*), residues that have both been shown to be important in stabilizing the transition-state structure (7). Consequently, increasing the side chain bulk at position 48 might accelerate folding by facilitating hydrophobic collapse. Consistent with this idea, a strong correlation ($r = 0.90$; Fig. 8, which is published as supporting information on the PNAS web site) was also observed between the $\Delta\Delta G_{‡\to u}$ values of each Gly-48 mutant and a pseudoenergy term based on the volume of the substituted residue ($\Delta\Delta G_{vol}$). Furthermore, when the sum of both $\Delta\Delta G_{P\beta}$ and $\Delta\Delta G_{vol}$ pseudoenergy terms is compared with the $\Delta\Delta G_{‡\to u}$ for each mutant (Fig. 4*B*), not only is an astonishingly good fit observed ($r = 0.99$), but the slope of the line is close to 1 with an intercept close to zero. This result implies that the observed stabilization of the transition state caused by the Gly-48 substitutions can be almost fully accounted for by the β-sheet forming propensity and volume of the substituted residue.

Another mechanism by which the Gly-48 substitutions could increase the folding rate is by destabilizing the unfolded state, which would narrow the energetic gap between this state and the transition state (Fig. 3). The increased structure forming potential produced by the Gly-48 substitutions might cause a simultaneous compaction of the unfolded state accompanied by an overall destabilization due to loss of entropy. Previous work has shown that introducing nonnative local structural propensity into a protein can lead to a compaction of the unfolded state (35). In addition, a single Ala-to-Thr substitution in the β-sheet of a variant B1 domain of protein G caused an increase in folding rate and compaction of the unfolded state, as assessed by the $m_{f\to u}$ values derived from folding kinetics experiments (36). A point that argues against a significant alteration in the unfolded-state structures of the Gly-48 mutants is that their $m_{f\to u}$ values do not deviate from the WT value. Because $m_{f\to u}$ values also are proportional to the difference in solvent exposure between the folded and unfolded states (37), a compaction of the unfolded

state would be expected to decrease the m_{kf} and overall $m_{f\rightarrow u}$ value.

Why is Gly-48 with its unusual conformation so highly conserved even though it is clearly not the optimal residue for fast folding of the SH3 domain? The unusual ϕ/ψ angles are not an absolute requirement for the SH3 domain fold because they are not present in other SH3-like proteins unrelated in sequence (38–40). In SH3 domains, the Gly-48 position is close to Pro-51, which is also almost universally conserved and is crucial for peptide-binding activity. Position 49, although less conserved, is also often involved in peptide binding (8). Therefore, the unusual ϕ/ψ angles are likely required for the proper positioning of key side chains on the functional surface. Although Gly-48 mutants are able to function, the backbone is nevertheless strained, as indicated by their extremely fast unfolding rates. Intriguingly, the change in free energy of binding of the Gly-48 mutants ($\Delta\Delta G_{bind}$, a higher value indicates weaker binding) and their $\Delta\Delta G_{f\rightarrow\ddagger}$ values are both inversely correlated with the β-sheet propensity energies for each substituted residue ($r = 0.92$ and 0.85, respectively; Fig. 4 *C* and *D*). The most favorable residues for β-sheet conformation are therefore the least favorable in the native-state structure and also lead to mutants that are the most impaired in their ability to adopt the backbone conformation required for function. Furthermore, substituted residues at the Gly-48 position that are most favorable in the native-state structure in terms of function are the least favorable in the transition-state structure. The Gly-48 position presents a clear case where evolution has selected for an unusual backbone conformation that is optimal for function but is not optimal for fast folding. Other examples where folding "frustration" caused by functional constraints leads to a slowed folding reaction have been recognized (41).

In protein engineering studies, the interpretation of changes in the folding and unfolding rates has relied on the assumption that the transition state is stabilized by the partial or complete formation of side-chain-mediated interactions present in the native folded structure (2). However, the possibility of nonnative interactions playing a role in the folding process has long been recognized because of the fairly frequent occurrence of Φ_f values that are beyond the expected range of 0 to 1. Theoretical results have also suggested that nonnative contacts can accelerate protein folding (42–44). The work presented here provides strong evidence for the potential importance of nonnative interactions in protein folding by elucidating for the first time, to our knowledge, a mechanism by which folding can occur fastest through a nonnative backbone conformation in the folding transition state. At present, the role of nonnative interactions in folding pathways may be underappreciated because the study of single-site substitutions with Ala, which is the predominant substitution investigated in protein engineering studies, may not always reveal the unusual type of behavior seen in this study. In the case of position 48, the Φ_f values for the Fyn (Table 1) and Src (14) SH3 domains calculated from G48A substitutions are very close to 0, and would not be considered notable on their own. To detect the potential roles of nonnative interactions in protein folding, it may be best to substitute positions of interest with a range of amino acids. For example Tyr-to-Phe substitutions have been used to demonstrate that hydrophobic residues outside of the structured portion of the transition state (as determined by Φ value analysis) can stabilize the transition state in a nonnative unspecific manner (45). Further studies into the mechanisms by which nonnative interactions may accelerate protein folding are clearly required to determine whether these interactions play a generally important role in protein folding.

We thank Claudia Vari and William Lester for technical assistance. This work was supported by grants from the Canadian Institutes of Health Research (to A.R.D. and L.E.K.). L.E.K. holds a Canada Research Chair in Biochemistry.

1. Fersht, A. R. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 1525–1529.
2. Fersht, A. R. (1995) *Curr. Opin. Struct. Biol.* **5,** 79–84.
3. Alm, E., Morozov, A. V., Kortemme, T. & Baker, D. (2002) *J. Mol. Biol.* **322,** 463–476.
4. Ladurner, A. G., Itzhaki, L. S., Daggett, V. & Fersht, A. R. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 8473–8478.
5. Northey, J. G. B., Maxwell, K. L. & Davidson, A. R. (2002) *J. Mol. Biol.* **320,** 389–402.
6. Mok, Y. K., Elisseeva, E. L., Davidson, A. R. & Forman-Kay, J. D. (2000) *J. Mol. Biol.* **307,** 913–928.
7. Northey, J. G. B., Di Nardo, A. A. & Davidson, A. R. (2002) *Nat. Struct. Biol.* **9,** 126–130.
8. Larson, S. M. & Davidson, A. R. (2000) *Protein Sci.* **9,** 2170–2180.
9. Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L. & Serrano, L. (1994) *Biochemistry* **33,** 2142–2150.
10. Plaxco, K. W., Guijarro, J. I., Morton, C. J., Pitkeathly, M., Campbell, I. D. & Dobson, C. M. (1998) *Biochemistry* **37,** 2529–2537.
11. Grantcharova, V. P. & Baker, D. (1997) *Biochemistry* **36,** 15685–15692.
12. Grantcharova, V. P., Riddle, D. S., Santiago, J. V. & Baker, D. (1998) *Nat. Struct. Biol.* **5,** 714–720.
13. Grantcharova, V. P., Riddle, D. S. & Baker, D. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7084–7089.
14. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. & Baker, D. (1999) *Nat. Struct. Biol.* **6,** 1016–1024.
15. Martinez, J. C., Pisabarro, M. T. & Serrano, L. (1998) *Nat. Struct. Biol.* **5,** 721–729.
16. Martinez, J. C. & Serrano, L. (1999) *Nat. Struct. Biol.* **6,** 1010–1016.
17. Minor, D., Jr., & Kim, P. S. (1994) *Nature* **367,** 660–663.
18. Smith, C. K., Withka, J. M. & Regan, L. (1994) *Biochemistry* **33,** 5510–5517.
19. Maxwell, K. L. & Davidson, A. R. (1998) *Biochemistry* **37,** 16172–16182.
20. Rickles, R. J., Botfield, M. C., Zhou, X. M., Henry, P. A., Brugge, J. S. & Zoller, M. J. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 10909–10913.
21. Loria, J. P., Rance, M. & Palmer, A. G. (1999) *J. Am. Chem. Soc.* **121,** 2331–2332.
22. Tollinger, M., Skrynnikov, N. R., Mulder, F. A. A., Forman-Kay, J. D. & Kay, L. E. (2001) *J. Am. Chem. Soc.* **123,** 11341–11352.
23. Mulder, F. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. (2001) *Nat. Struct. Biol.* **8,** 932–935.
24. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 211–222.
25. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22,** 2577–2637.
26. Pal, D. & Chakrabarti, P. (2000) *Acta Crystallogr. D* **56,** 589–594.
27. Chothia, C. (1975) *Nature* **254,** 304–308.
28. Palmer, A. G., III, Kroenke, C. D. & Loria, J. P. (2001) *Methods Enzymol.* **339,** 204–238.
29. Ventura, S., Vega, M. C., Lacroix, E., Angrand, I., Spagnolo, L. & Serrano, L. (2002) *Nat. Struct. Biol.* **9,** 485–493.
30. Roder, H. & Colon, W. (1997) *Curr. Opin. Struct. Biol.* **7,** 15–28.
31. Matouschek, A. & Fersht, A. R. (1991) *Methods Enzymol.* **202,** 82–112.
32. Fersht, A. R., Matouschek, A. & Serrano, L. (1992) *J. Mol. Biol.* **224,** 771–782.
33. Jackson, S. E., elMasry, N. & Fersht, A. R. (1993) *Biochemistry* **32,** 11270–11278.
34. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995) *J. Mol. Biol.* **254,** 260–288.
35. Prieto, J., Wilmans, M., Jimenez, M. A., Rico, M. & Serrano, L. (1997) *J. Mol. Biol.* **268,** 760–778.
36. Smith, C. K., Bu, Z., Anderson, K. S., Sturtevant, J. M., Engelman, D. M. & Regan, L. (1996) *Protein Sci.* **5,** 2009–2019.
37. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995) *Protein Sci.* **4,** 2138–2148.
38. Delbruck, H., Ziegelin, G., Lanka, E. & Heinemann, U. (2002) *J. Biol. Chem.* **277,** 4191–4198.
39. Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R. & Hard, K. (1995) *Nat. Struct. Biol.* **2,** 807–810.
40. Narayana, N., Matthews, D. A., Howell, E. E. & Nguyen-huu, X. (1995) *Nat. Struct. Biol.* **2,** 1018–1025.
41. Gruebele, M. (2002) *Nat. Struct. Biol.* **9,** 154–155.
42. Plotkin, S. S. (2001) *Proteins* **45,** 337–345.
43. Treptow, W. L., Barbosa, M. A., Garcia, L. G. & Pereira de Araujo, A. F. (2002) *Proteins* **49,** 167–180.
44. Li, L., Mirny, L. A. & Shakhnovich, E. I. (2000) *Nat. Struct. Biol.* **7,** 336–342.
45. Viguera, A. R., Vega, C. & Serrano, L. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 5349–5354.
46. Noble, M. E., Musacchio, A., Saraste, M., Courtneidge, S. A. & Wierenga, R. K. (1993) *EMBO J.* **12,** 2617–2624.
47. Evans, S. V. (1993) *J. Mol. Graphics* **11,** 134–138.