Supporting Information

Zarrine-Afsar *et al.* 10.1073/pnas.0801874105

SI Methods

Model Simulation of Folding Kinetics. Simulations of model folding kinetics and thermodynamic sampling were performed by using Langevin dynamics. The values of the parameters of the potential function (1) E_0 (which equals the function E defined in ref. 1), the criterion for native contacts, and the Langevin procedure (1) are identical to those in ref. 2. Folding simulations of a given model were performed at approximately the transition midpoint temperature, T_m, of the model (see below). Starting from a random unfolded conformation (with progress variable $Q < 0.25$), we determined folding rate as $k_f = (\mathrm{MFPT})^{-1}$, where (MFPT) is the mean first passage time, measured in simulation time steps, for the folding chain to arrive at the model native state ($Q > 0.75$). As in previous studies (2, 3), kinetic relaxation at simulation temperatures around T_m was found to be essentially single-exponential. In this study, each folding rate and each free energy profile was computed by using 3×10^8 simulation time steps.

Folding simulations were first performed for the wild-type Fyn SH3 domain with the κ_i value for all hydrophobic residues along its sequence set to 1.0. These simulations showed that folding rate of this model first increases then decreases with increasing overall hydrophobic strength K_{HP} (Fig. S4). This trend is consistent with previous analytical (4) and computational (5, 6) analyses of general effects of nonnative interactions. Because the decrease in folding rate at higher K_{HP} values signals significant kinetic trapping effects that are uncharacteristic of the experimental cooperative folding processes we aimed to model, we adopted a lower intermediate K_{HP} value of 0.8 for all other simulations in this study.

Within this theoretical construct, effects of point mutations on folding kinetics may be modeled readily. A polar amino acid at any given sequence position i can be substituted for a hydrophobic amino acid, and vice versa, by choosing a hydrophobic ($\kappa_i > 0$) or hydrophilic ($\kappa_i = 0$) value for κ_i in the model. Such a sequence change does not affect the native-centric E_0 term, but it changes the general nonnative hydrophobic interactions in the E_{HP} term. By the same token, the substitution of one hydrophobic residue by another with a different hydrophobicity corresponds to a change in the positive κ_i value for that position.

We chose to conduct folding simulations at model temperatures coinciding with or very near the transition temperature, T_m, of the model because simulated folding kinetics is two-state-like under model T_m conditions and thus serves well in this regard as a mimicry for cooperative folding of real, single-domain proteins. Our simulations for single mutants of the Fyn SH3 domain were conducted at $T_m = 1.051$ (in model units), because their T_m values were practically identical. The T_m values for the simulated double mutants varied slightly, from 1.041 to 1.052, but were all within ≈0.5% of their mean value. In other words, therefore, all model folding rates reported in this work were obtained under essentially the same folding conditions. In our analysis, these folding rates simulated at $\approx T_m$ in the model were applied to rationalize the differences in experimental folding rates of Fyn SH3 mutants in general, including rates measured under strongly folding conditions at low denaturant concentrations. This approach to matching theory with experiment is tenable because the folding arms of the experimental chevron plots for the Fyn SH3 mutants are essentially parallel (see Fig. S1); hence, the rank order of experimental folding rate is independent of folding conditions. In light of this observation, we stipulate that our model can provide useful predictions for the kinetics of folding, even though its lack of nonadditive many-body interactions suggests that the model is likely insufficient for predicting mutational effects on unfolding rate and native stability (7), as noted in the main text.

Experimental Measurement of Folding Kinetics. The kinetics of folding and unfolding were measured at 25°C through monitoring the changes upon GuHCl unfolding of the *Trp* fluorescence of the domain, using a Bio-Logic SFM-4 stopped flow device (BioLogic Instruments). Samples were excited at 295 nm, and the total fluorescence emission above 309 nm was collected by using a custom-made cut-on optical filter (Omega Filters) mounted on the PMT detector. The final protein concentration inside the cuvette was ≈5 μM. At each final concentration of GuHCl, kinetic traces from at least five separate kinetic runs were averaged. The averaged folding and unfolding traces for each mutant were fit to appropriate single-exponential functions in BioKine (BioLogic Instruments) using a Padé–Laplace method (8) and the Simplex algorithm (9) to obtain the rate constants of folding and unfolding as well as the uncertainty associated with the exponential fit (as standard deviation of the trace from best-fit exponential curve).

Kinetic chevron plots of each mutant were fit based on the following chevron equation by nonlinear least-squares minimization in Matlab 7.1.0.183 (The Mathworks), as well as in Kaleidagraph (Synergy Software):

$$\ln k_{obs} = \ln\{\exp[\ln k_f + m_{kf}(0.5 - [\mathrm{GuHCl}])] + \exp[\ln k_u + m_{ku}([\mathrm{GuHCl}] - 5.0)]\} \quad [1]$$

where k_{obs} is the observed rate constant at a given GuHCl concentration, k_f and k_u are the folding and unfolding rates at 0.5 M and 5 M GuHCl, respectively, and m_{kf} and m_{ku} are the slopes of the dependence of ln k_f and ln k_u, respectively, on the concentration of GuHCl. To minimize the errors associated with linear extrapolation of the observed rates to zero concentration of GuHCl, we have reported the folding and the unfolding rates of the mutants at 0.5 and 5 M concentrations of GuHCl, respectively, and used these data for our analyses.

Values for changes in free energy between the unfolded state (u), transition state (‡), and native state (f) were calculated as follows (in obvious notation):

$$\Delta\Delta G_{‡\to u} = -RT\ \ln(k_f^{mut}/k_f^{wt}) \quad [2]$$

$$\Delta\Delta G_{f\to ‡} = -RT\ \ln(k_u^{wt}/k_u^{mut}) \quad [3]$$

$$\Delta\Delta G_{f\to u} = \Delta\Delta G_{‡\to u} + \Delta\Delta G_{f\to ‡} \quad [4]$$

where RT is gas constant times absolute temperature. Errors on folding rates were evaluated using a Monte Carlo (10) sampling in Matlab 7.1.0.183 (Mathworks). It is immediately obvious that the error on the individual rates k_{obs} ([GuHCl]) associated with deviation from the best-fit single-exponential *Trp* fluorescence recovery curves is typically two orders of magnitude smaller than the actual scatter of the data points around the best-fit chevron plot. Accordingly, propagating these error estimates via Monte Carlo simulations yields unrealistically small error estimates for the fitted parameters k_f, m_{kf}, k_u, and m_{ku} (data not shown). The scatter, therefore, primarily reflects other, more significant sources of experimental error, such as slight variations in the stopped-flow mixing, final denaturant concentration, performance of the UV lamp, etc. Visual inspection of the chevron plot suggested that the

scatter is relatively uniform for all data points. Thus, making use of the error propagation law, the errors on data points could be estimated from the rms scatter of ln k_{obs}([GuHCl]) around the best-fit chevron plots ln k_{fit}([GuHCl]). These errors (typically ≈5–10% and assumed to be normally distributed) were propagated by Monte Carlo simulation to obtain error margins for the fitted parameters k_f, m_{kf}, k_u, and m_{ku} reported in Table S1. As argued above, we are confident that these error margins accurately reflect the experimental uncertainties associated with our folding kinetic data. Two synthetic Monte Carlo simulations assuming relative errors of 5% and 10% for all individual rates k_{obs}([GuHCl]) showed that error propagation is linear in this regime, suggesting that estimation of errors for the fitting parameters k_f, m_{kf}, k_u, and m_{ku} via the covariance matrix is also a valid approach (10); accordingly, fitting the chevron equation by Levenberg–Marquardt nonlinear least-squares fitting with covariance matrix error analysis in Kaleidagraph (Synergy) yields folding parameters with errors that are very similar to those reported in Table S1. The Monte Carlo errors (dx_i) were propagated by using the equation below to obtain the errors (du) of $\Delta\Delta G_{\ddagger\rightarrow u}$:

$$du = \sqrt{\sum_i [(\partial u/\partial x_i)^2\ (dx_i)^2]} \qquad [5]$$

where $u = f(x_i)$ is a function of x_i, and dx_i is the residual error of x_i.

Temperature-Induced Unfolding Experiments. Changes in ellipticity at 220 nm were measured at equilibrium as a function of temperature using an Aviv Circular Dichroism spectrometer model 62A DS (Aviv Instruments). The T_m values were obtained as previously described (11).

NMR Spectroscopy and Sequence Assignment. [^{1}H,^{15}N] HSQC spectra were recorded on 1.0 mM [U-^{15}N] WT and N53I Fyn SH3 with 0.2 mM EDTA, 0.05% NaN_3, 50 mM sodium phosphate (pH 7.0) in H_2O/D_2O (9:1) at 25°C on Varian Inova spectrometers operating at ^{1}H frequencies of 800 MHz and 500 MHz, respectively. Sequence-specific ^{1}H and ^{15}N resonance assignments for the WT Fyn SH3 were obtained from [^{1}H,^{15}N]-TOCSY-HSQC and [^{1}H,^{15}N]-NOESY-HSQC spectra at 25°C as described previously (12). NMR spectra were processed and analyzed as described previously (12).

X-Ray Crystallography. Fyn SH3 (N53I-V55L) was crystallized at 20°C by using hanging drops containing equal volumes of well solutions (0.1 M sodium cacodylate, pH 6.5/1.4 M sodium acetate) and protein solution (20 mg/ml protein/20 mM Tris, pH 8.0). Crystals grew over 2 weeks. Data were collected in house by using an R-Axis IV^{++} Rigaku RUH3R rotating anode generator with Osmic optics and processed by using the program XDS (13). The data collection and processing statistics are summarized in Table S2.

Fyn SH3 (N53I-V55L) was crystallized in the space group $P3_121$, with one molecule per asymmetric unit. The structure was solved by molecular replacement at 4 Å with COMO (14) using the crystal structure of the SH3 domain of Fyn (15) as search model (PDB entry 1SHF) after removing the water molecules.

Structure refinement of Fyn SH3 (N53I-V55L) was alternated between REFMAC5 (16) and manual rebuilding using Coot (17). The progress of the refinement was monitored by the improvement of R_{free} with 5% cross-validated data. Water molecules were added automatically by using ARP/wARP (18) and individually inspected. The model was refined to an R/R_{free} of 0.207/0.24 and contains no Ramachandran violations. Ninety-six percent of the dihedral angles are in the most-favored, four percent in the additionally allowed, and none in the generally allowed regions (19). The crystallographic refinement statistics are summarized in Table S2.

1. 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.
2. 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.
3. 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.
4. Plotkin SS (2001) Speeding protein folding beyond the Gō model: How a little frustration sometimes helps. *Proteins* 45:337–345.
5. Clementi C, Plotkin SS (2004) The effects of nonnative interactions on protein folding rates: Theory and simulation. *Protein Sci* 13:1750–1766.
6. Fan K, Wang J, Wang W (2002) Folding of lattice protein chains with modified Gō potential. *Eur Phys J B* 30:381–391.
7. Kaya H, Chan HS (2003) Simple two-state protein folding kinetics requires near-Levinthal thermodynamic cooperativity. *Proteins* 52:510–523.
8. Yeramian E, Schaeffer F, Caudron B, Claverie P, Buc H (1989) A rigorous mathematical treatment for the excluded volume effect in Monte Carlo simulations of polymeric chains. *Biopolymers* 28:2059–2070.
9. Enderlein J, Erdmann R (1997) Fast fitting of multi-exponential decay curves. *Opt Commun* 134:371–378.
10. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Numerical Recipes* (Cambridge Univ Press, New York), 3rd Ed.
11. 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.
12. 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.
13. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
14. Jogl G, Tao X, Xu Y, Tong L (2001) COMO: A program for combined molecular replacement. *Acta Crystallogr D* 57:1127–1134.
15. Noble MEM, Musacchio A, Saraste M, Courtneidge SA, Wierenga RK (1993) Crystal structure of the SH3 domain in human Fyn—comparison of the 3-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *EMBO J* 12:2617–2624.
16. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
17. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
18. Perrakis A, Sixma TK, Wilson KS, Lamzin VS (1997) wARP: Improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr D* 53:448–455.
19. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK—A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
20. Cruickshank DW (1999) Remarks about protein structure precision. *Acta Crystallogr D* 55:583–601.
21. 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.

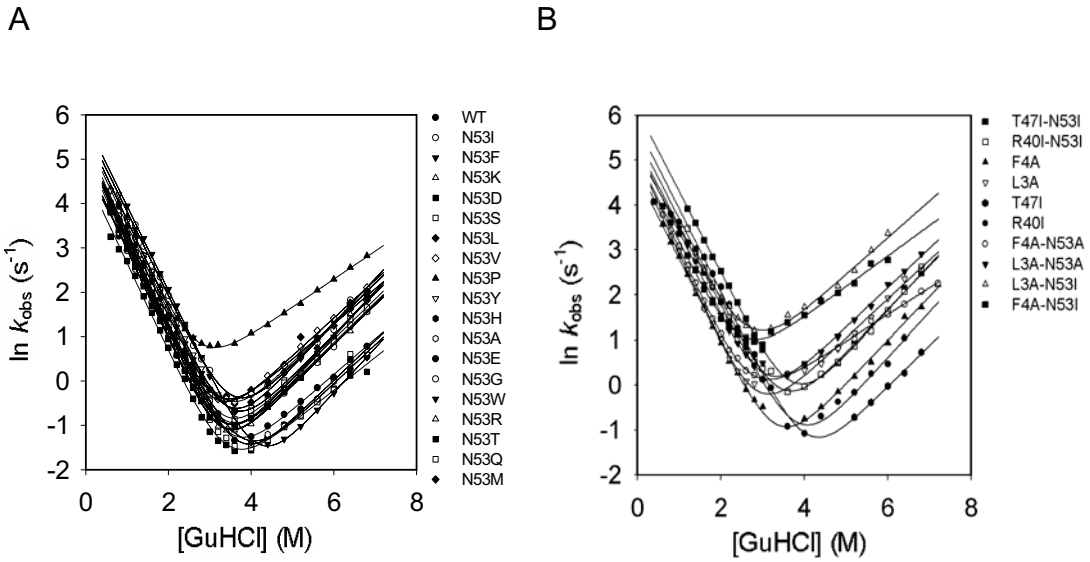

Fig. S1. The kinetic chevron plots of the mutants used in this study. Shown are chevron plots of N53 mutants (*A*) and those of other mutants (*B*) used in the analysis of interaction energies presented in Fig. 4. Plots of the logarithms of the observed rate constants (ln k_{obs}) at various concentrations of GuHCl for the WT Fyn SH3 domain and all of the mutants except for the Cys substitution exhibited straight folding and unfolding arms indicating that two-state folding has been maintained. Note that the N53W mutant folds faster than the N53I mutant (see also Table S1). We did not choose N53W for further investigation using double mutant cycles because the presence of two Trps in this mutant would likely cause complications in the interpretation of the Trp fluorescence signal that was used as a probe to monitor the folding/unfolding transition.

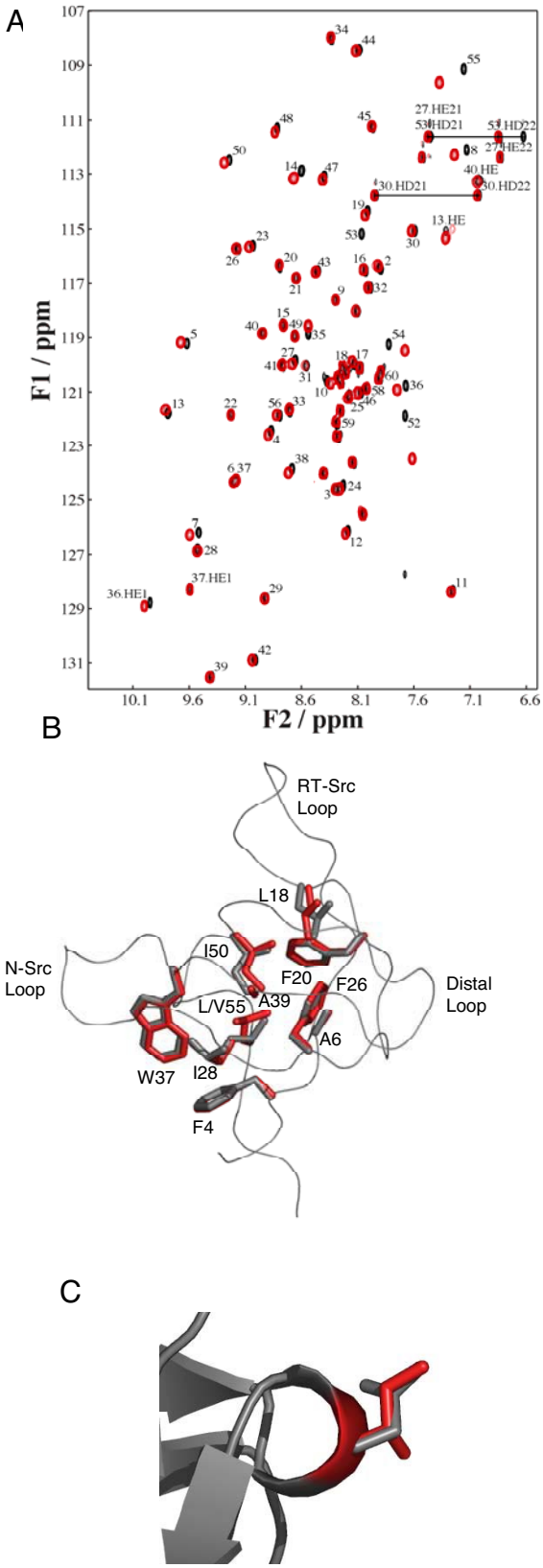

Fig. S2. Structural studies of the N53I substitution in the Fyn SH3 domain. (*A*) An overlay of the [^{1}H,^{15}N] HSQC spectrum of N53I (red) on top of the WT spectrum (black). (*B*) The overlay of the crystal structure of Fyn SH3 N53I-V55L double mutant (3CQT) to that of WT (1SHF) suggests that neither mutation causes a dramatic alteration in the hydrophobic core of the protein. (*C*) An expanded view of the overlay suggesting that the Ile side chain in N53I (stick representation) retains its WT-like solvent exposure.

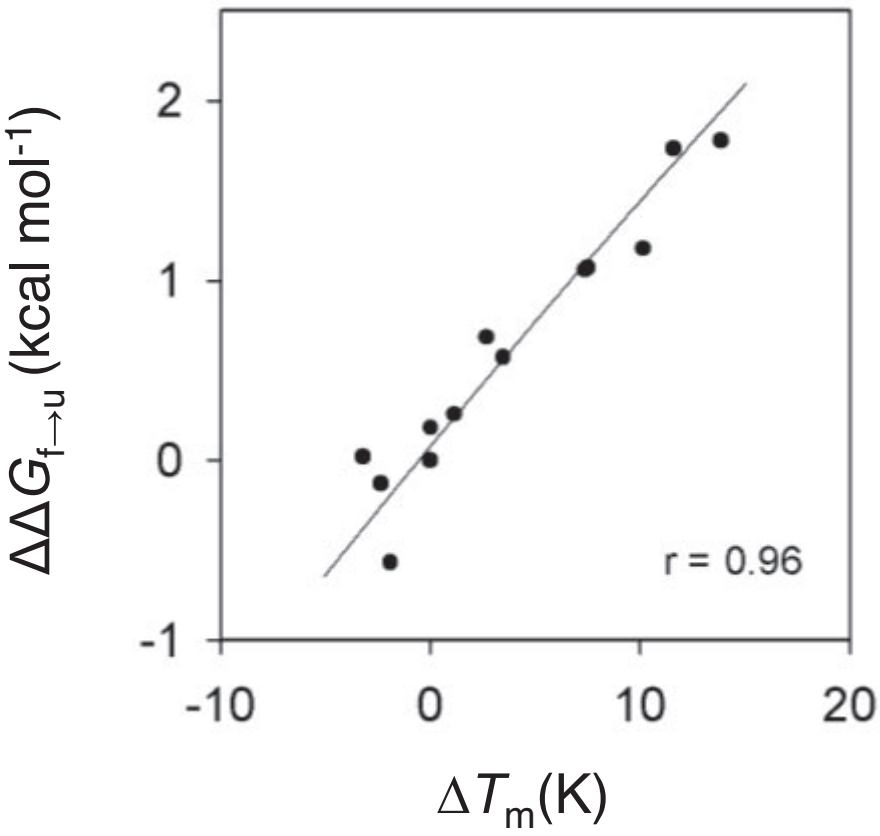

Fig. S3. The correlation between ΔT_m (Table S1, determined by measuring change in ellipticity at 220 nm upon heat unfolding) and $\Delta\Delta G_{f\rightarrow u}$ values (calculated from folding and unfolding rates in GuHCl) of all of the mutants used in thermodynamic cycles presented in Fig. 4. The data were fit to an expanded form of the Gibbs–Helmholtz equation as described in detail in ref. 21.

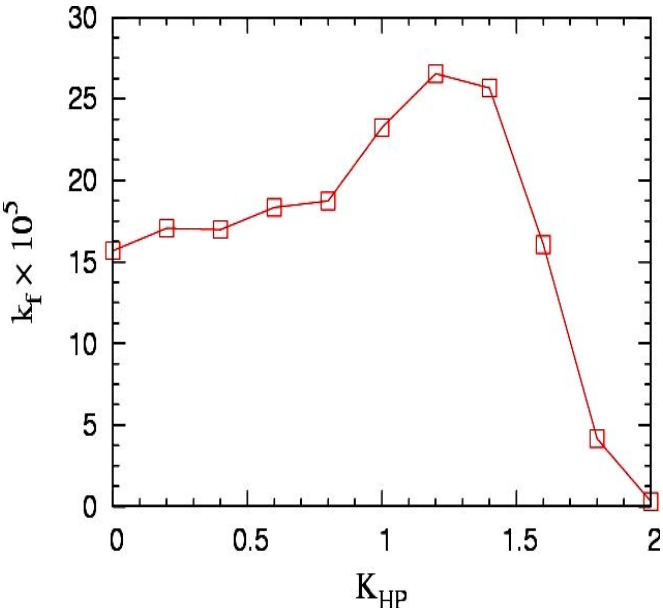

Fig. S4. The effect of increasing the overall strength of hydrophobic interactions, K_{HP}, on simulated folding rate (see Eq. **2** in the main text). Hydrophobic residues were assigned according to the amino acid sequence of WT Fyn SH3 domain. This figure shows that the folding rate initially increased as hydrophobic interactions were made stronger; folding rate then decreased sharply for K_{HP} value above ≈1.3. This trend is consistent with previous theoretical and computational studies of nonnative interactions as ''noise'' added to Gō-type potentials (4–6). Each k_f value was computed at the transition midpoint temperature T_m of the model at the given K_{HP}.

Table S1. Folding kinetics and thermal stability properties of the Fyn SH3 domain mutants used in this study

Mutants	k_f,* s^{-1}	m_{kf}†	k_u,* s^{-1}	m_{ku}†	ΔT_m,‡ °C
WT	50.28 ± 3.41	1.78 ± 0.05	0.39 ± 0.02	0.88 ± 0.04	0.00
N53A	65.52 ± 3.85	1.95 ± 0.04	0.41 ± 0.01	0.91 ± 0.03	−2.36 ± 0.29
N53I	118.67 ± 8.62	1.91 ± 0.05	1.42 ± 0.04	0.93 ± 0.03	1.16 ± 0.44
N53F	92.32 ± 4.84	1.99 ± 0.04	1.50 ± 0.05	0.89 ± 0.03	3.20 ± 0.40
N53K	65.90 ± 3.10	2.13 ± 0.03	0.90 ± 0.02	0.92 ± 0.02	0.20 ± 0.30
N53D	38.54 ± 2.05	1.98 ± 0.02	0.38 ± 0.01	0.73 ± 0.04	2.00 ± 0.30
N53S	65.60 ± 2.00	2.04 ± 0.02	0.38 ± 0.01	0.84 ± 0.02	0.10 ± 0.20
N53L	69.84 ± 4.60	1.99 ± 0.05	1.63 ± 0.06	0.79 ± 0.04	0.90 ± 0.30
N53V	75.53 ± 3.30	2.10 ± 0.03	1.75 ± 0.03	0.86 ± 0.02	5.90 ± 0.20
N53P	101.17 ± 2.83	2.01 ± 0.03	5.28 ± 0.07	0.63 ± 0.01	13.00 ± 0.30
N53Y	101.77 ± 5.53	1.98 ± 0.04	1.11 ± 0.04	0.93 ± 0.04	2.44 ± 0.20
N53H	78.82 ± 3.10	2.09 ± 0.03	0.95 ± 0.02	0.90 ± 0.02	2.70 ± 0.30
N53E	59.96 ± 1.29	1.99 ± 0.01	0.51 ± 0.01	0.80 ± 0.01	4.70 ± 0.33
N53T	53.50 ± 1.56	2.00 ± 0.02	0.97 ± 0.02	1.01 ± 0.02	−0.63 ± 0.30
N53R	67.50 ± 2.10	2.05 ± 0.02	1.04 ± 0.02	0.97 ± 0.01	−1.79 ± 0.30
N53Q	69.96 ± 2.76	2.01 ± 0.03	1.05 ± 0.03	0.90 ± 0.02	−0.46 ± 0.25
N53M	74.11 ± 3.24	1.89 ± 0.03	1.36 ± 0.04	0.99 ± 0.03	−0.27 ± 0.25
N53W	147 ± 6.11	1.95 ± 0.02	0.28 ± 0.01	0.99 ± 0.02	−4.96 ± 0.40
N53G	52.73 ± 1.51	2.00 ± 0.02	0.93 ± 0.02	0.87 ± 0.01	0.64 ± 0.20
L3A	50.21 ± 2.68	2.01 ± 0.05	2.85 ± 0.07	0.87 ± 0.03	10.50 ± 0.19
F4A	41.73 ± 1.70	1.86 ± 0.03	1.03 ± 0.03	0.97 ± 0.02	2.69 ± 0.19
R40I	83.14 ± 6.55	1.77 ± 0.05	0.67 ± 0.04	1.02 ± 0.07	−3.20 ± 0.50
T47I	127.43 ± 11.65	1.86 ± 0.05	0.38 ± 0.04	0.92 ± 0.07	−1.90 ± 0.20
L3A-N53A	76.02 ± 4.38	1.87 ± 0.04	3.61 ± 0.09	0.88 ± 0.02	7.50 ± 0.71
L3A-N53I	74.41 ± 12.09	1.68 ± 0.12	11.65 ± 0.34	0.82 ± 0.04	13.88 ± 0.33
F4A-N53A	56.94 ± 1.59	2.04 ± 0.03	2.64 ± 0.04	0.59 ± 0.01	7.37 ± 0.31
F4A-N53I	61.40 ± 3.84	1.90 ± 0.07	8.92 ± 0.37	0.68 ± 0.04	11.6 ± 0.28
R40I-N53I	99.59 ± 8.81	1.82 ± 0.05	2.04 ± 0.08	0.97 ± 0.04	3.48 ± 0.21
T47I-N53I	182.97 ± 7.76	1.81 ± 0.03	1.93 ± 0.05	1.00 ± 0.02	0.02 ± 0.20

The errors reported in this table are the uncertainties obtained through a Monte Carlo sampling of fitting errors, as described in *Materials and Methods* and *SI Methods*. The errors reported for ΔT_m values are fitting errors.

*Folding (k_f) and unfolding (k_u) rates are reported at 0.5 M and 5 M concentrations of GuHCl, respectively.

†m_{kf} and m_{ku} are the slopes of dependence of $\ln(k_f)$ and $\ln(k_u)$, respectively, on the concentration of GuHCl.

‡ΔT_m is the melting temperature of the WT minus that of the mutant, i.e., $\Delta T_m = T_m^{WT} - T_m^{mut}$.

Table S2. Crystallographic data and refinement statistics for the Fyn SH3 domain N53I-V55L mutant (PDB ID code 3CQT)

Data collection statistics	
Space group	$P3_121$
Cell dimensions	
a, Å	52.032
b, Å	52.032
c, Å	44.817
α , °	90
β , °	90
γ , °	120
Wavelength, Å	1.54
Resolution, Å	45.06–1.60 (1.70–1.60)*
R_{int}†	0.0103 (0.0241)
R_σ†	0.0311 (0.0181)
$\langle I \rangle / \sigma$ (*I*)	46.71 (28.89)
Redundancy	1.74 (1.84)
Completeness, %	93.7 (97.8)
B values, Wilson plot, Å^2	15.3

Refinement statistics	
Resolution, Å	45.06–1.6
R/R_{free}‡	0.207/0.24
No. of protein atoms	501
No. of water molecules	92
rmsd from ideal geometry	
Bonds, Å	0.008
Angles, °	1.1
Dihedral angles, °	16.8
Improper angles, °	0.66
Estimated coordinate error,§ Å	0.1
Average *B* factor (Å^2)	21
Ramachandran,¶ (%)	94/6/0/0

*Values in parentheses correspond to the values in the highest-resolution shell.

†$R_{int} = \Sigma\Sigma|F_o^2 - \langle F_o^2 \rangle| / \Sigma\Sigma < F_o^2 >$; $R_\sigma = \Sigma[\sigma (F_o^2)] / \Sigma[F_o^2]$.

‡$R = \Sigma_h||F_o| - |F_c||/\Sigma_h|F_o|$; R_{free} is *R* for the 5% cross-validated test data.

§Cruickshank's diffraction component precision index (DPI) as an estimate of coordinate error (20).

¶Fractions of residues in most favored/allowed/generously allowed/disallowed regions of the Ramachandran plot according to PROCHECK (19).