the rate constant; and (iii) above 15 cP, where the solvent friction dominates. Extrapolation of these results to low temperatures indicates that solvent viscosity could have very large effects on the rate of conformational change. At −95°C in 79% by weight glycerol-water, the pre-exponential factor in Eq. 3 for this conformational change is predicted from the viscosity (*16*) to be reduced by a factor of about 10^{11} compared to the value at 20°C, whereas the exponential term is predicted to be reduced by a factor of less than 200. This extrapolation suggests that the marked decrease in the rate of interconversion of conformational substates at temperatures near the glass transition (*10*, *17*, *18*) results more from the enormous viscosity of the solvent than from potential energy barriers that are large compared to the average thermal energy. That is, at low temperatures conformational substates may not be "frozen" so much as "stuck." This hypothesis predicts that the low-temperature kinetics of Mb, which has a broad distribution of ligand rebinding rates resulting from a distribution of noninterconverting conformational substates (*10*, *17*), would be mimicked at room temperature in very high viscosity solvents. Distributed kinetics at room temperature have, in fact, been observed in solid polyvinyl alcohol (*17*).

What is the influence of solvent viscosity on the rate of protein conformational changes under physiological conditions? The viscosity of cytoplasm has been estimated to be about 2 to 3 cP (*19*). Although we do not yet know anything about the values of σ for other proteins, the present results suggest that intracellular viscosities could slow protein conformational changes significantly and therefore influence the kinetics of protein function (*20*).

REFERENCES AND NOTES

1. M. Smoluchowski, *Z. Phys. Chem.* **92**, 129 (1917); S. A. Rice, *Diffusion-Limited Reactions*, vol. 25 of *Comprehensive Chemical Kinetics* (Elsevier, Amsterdam, 1985).
2. H. A. Kramers, *Physica (Utrecht)* **7**, 284 (1940); P. Hänggi, P. Talkner, M. Borkovec, *Rev. Mod. Phys.* **62**, 251 (1990).
3. D. Beece *et al.*, *Biochemistry* **19**, 5147 (1980).
4. B. A. Chrunyk and C. R. Matthews, *ibid.* **29**, 2149 (1990).
5. E. R. Henry, J. H. Sommer, J. Hofrichter, W. A. Eaton, *J. Mol. Biol.* **166**, 443 (1983); A. Ansari *et al.*, *Biochemistry* **25**, 3139 (1986).
6. D. G. Lambright, S. Balasubramanian, S. G. Boxer, *Chem. Phys.* **158**, 249 (1991).
7. L. P. Murray, E. R. Henry, J. Hofrichter, W. A. Eaton, *Biophys. Chem.* **29**, 63 (1988).
8. J. Kuriyan, S. Wilz, M. Karplus, G. A. Petsko, *J. Mol. Biol.* **192**, 133 (1986).
9. N. Agmon and J. J. Hopfield, *J. Chem. Phys.* **79**, 2042 (1983); E. R. Henry, J. Hofrichter, J. H. Sommer, W. A. Eaton, in *Brussels Hemoglobin Symposium*, A. G. Schnek and C. Paul, Eds. (Edition de l'Université de Bruxelles, Brussels, 1983), pp. 193–203; J. Hofrichter *et al.*, *Biochemistry* **24**, 2667 (1985).
10. P. J. Steinbach *et al.*, *Biochemistry* **30**, 3988 (1991).
11. X. Xie and J. D. Simon, *ibid.*, p. 3682.
12. L. Genberg, L. Richard, G. McLendon, R. J. D. Miller, *Science* **251**, 1051 (1991).
13. One of the smaller amplitude relaxations may correspond to the 700-ns process observed in water at 20°C by photoacoustic calorimetry [J. A. Westrick, J. L. Goodman, K. S. Peters, *Biochemistry* **26**, 8313 (1987)].
14. H. Frauenfelder *et al.*, *J. Phys. Chem.* **94**, 1024 (1990); H. Frauenfelder, S. G. Sligar, P. G. Wolynes, *Science* **254**, 1598 (1991).
15. The rate constants shown were corrected for the activation energy to the reference temperature of 20°C. The viscosities were calculated from the empirical function of Hasinoff (*23*). The rate constants were obtained in a three-step procedure designed to remove the contribution of the slower, smaller amplitude processes. First, we obtained a set of rate constants, parameterized by Eq. 3, by simultaneously fitting both the ligand rebinding curves (V_1) and the unnormalized progress curves (V_2) for the deoxyheme spectral changes to a sum of four relaxations (*22*). In this fit a stretched exponential was used to describe the first relaxation, exponentials to describe the next two relaxations, and the integrated form of the second-order rate equation to describe the final bimolecular relaxation. Second, the contribution of the bimolecular relaxation was subtracted from the progress curves for the deoxyheme spectral changes, and these curves were normalized to 100% deoxyhemes with the use of the ligand rebinding curves (Fig. 2A). Finally, we refit the normalized data for the deoxyheme spectral changes (Fig. 2, B and C) by using Eq. 1 for the first process with a single value of β and by allowing the rate constant to be optimized independently for each progress curve. In this fit the amplitude of the first relaxation was constrained to be independent of both solvent and temperature, the rates for the two slower processes were constrained to the values obtained in the simultaneous fit with the ligand rebinding curves, and the amplitude of the second relaxation was constrained to be independent of the solvent. The major results of the fit do not depend significantly on the specific assumptions used in fitting the slower processes. We obtained the uncertainties in the data by optimizing k in Eq. 1 for various fixed values of β. The rates were found to decrease monotonically with increasing values of β. Increasing β to 1.0 ($A_0 = 0.041$) or decreasing β to 0.4 ($A_0 = 0.088$) increased the χ^2 by 2.7, which represents the 90% confidence limit, and the corresponding values of k were taken as the uncertainties that are plotted in Fig. 3. The values obtained at these limits of β were used to obtain the errors in the parameters C, σ, and E_0.
16. A. A. Newman, *Glycerol* (CRC Press, Cleveland, 1968); D. H. Rasmussen and A. P. MacKenzie, *J. Phys. Chem.* **75**, 967 (1971); C. A. Angell and W. Sichina, *Ann. N.Y. Acad. Sci.* **279**, 53 (1976).
17. R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, I. C. Gunsalus, *Biochemistry* **14**, 5355 (1975).
18. A. Ansari *et al.*, *Biophys. Chem.* **26**, 337 (1987); I. E. T. Iben *et al.*, *Phys. Rev. Lett.* **62**, 1916 (1989); V. Šrajer, L. Reinisch, P. M. Champion, *Biochemistry* **30**, 4886 (1991).
19. A. M. Mastro and A. D. Keith, *J. Cell Biol.* **99** (part 2), 180s (1984).
20. B. Gavish and M. M. Werber, *Biochemistry* **18**, 1269 (1979); A. P. Demchenko, O. I. Rusyn, E. A. Saburova, *Biochim. Biophys. Acta* **998**, 196 (1989).
21. J. Hofrichter *et al.*, *Biochemistry* **30**, 6583 (1991).
22. J. Hofrichter, J. H. Sommer, E. R. Henry, W. A. Eaton, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2235 (1983); E. R. Henry and J. Hofrichter, *Methods Enzymol.* **210**, 129 (1992).
23. B. B. Hasinoff, *Arch. Biochem. Biophys.* **183**, 176 (1977).

24. We thank B. Bagchi, M. Edidin, F. Stillinger, A. Szabo, and R. Zwanzig for helpful discussions.

27 February 1992; accepted 14 April 1992

Probing Protein Stability with Unnatural Amino Acids

David Mendel, Jonathan A. Ellman, Zhiyuh Chang, David L. Veenstra, Peter A. Kollman, Peter G. Schultz*

Unnatural amino acid mutagenesis, in combination with molecular modeling and simulation techniques, was used to probe the effect of side chain structure on protein stability. Specific replacements at position 133 in T4 lysozyme included (i) leucine (wt), norvaline, ethylglycine, and alanine to measure the cost of stepwise removal of methyl groups from the hydrophobic core, (ii) norvaline and *O*-methyl serine to evaluate the effects of side chain solvation, and (iii) leucine, *S,S*-2-amino-4-methylhexanoic acid, and *S*-2-amino-3-cyclopentylpropanoic acid to measure the influence of packing density and side chain conformational entropy on protein stability. All of these factors (hydrophobicity, packing, conformational entropy, and cavity formation) significantly influence protein stability and must be considered when analyzing any structural change to proteins.

Mutational studies of the amino acids that form the hydrophobic core of proteins are beginning to define how these residues influence protein structure and stability (*1*–*10*). However, it is difficult to make mutations with the natural 20 amino acids that perturb one interaction without simultaneously affecting several others. For example, mutation of $Leu^{133} \rightarrow$ Phe or $Ala^{129} \rightarrow$ Val in T4 lysozyme (T4L) in an attempt to increase packing density, and as a consequence, thermal stability, resulted in a less

D. Mendel, J. A. Ellman, Z. Chang, P. G. Schultz, Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, and Center for Advanced Materials, Lawrence Berkeley Laboratory, Berkeley, CA 94720.

D. L. Veenstra and P. A. Kollman, Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94143.

*To whom correspondence should be addressed.

stable protein because of a local increase in strain energy (*11*). The ability to site-specifically incorporate unnatural amino acids into proteins makes possible more precise changes in the steric or electronic properties of an amino acid and expands the scope of structural perturbations that can be made (*12*, *13*). We have used molecular modeling and simulation techniques to design unnatural amino acids to probe the effects of hydrophobicity, packing, cavity formation, and side chain conformational entropy on protein stability. The effects of these substitutions on the thermal stability of T4L were evaluated from the melting properties of the purified mutant proteins determined by circular dichroism (CD) spectroscopy.

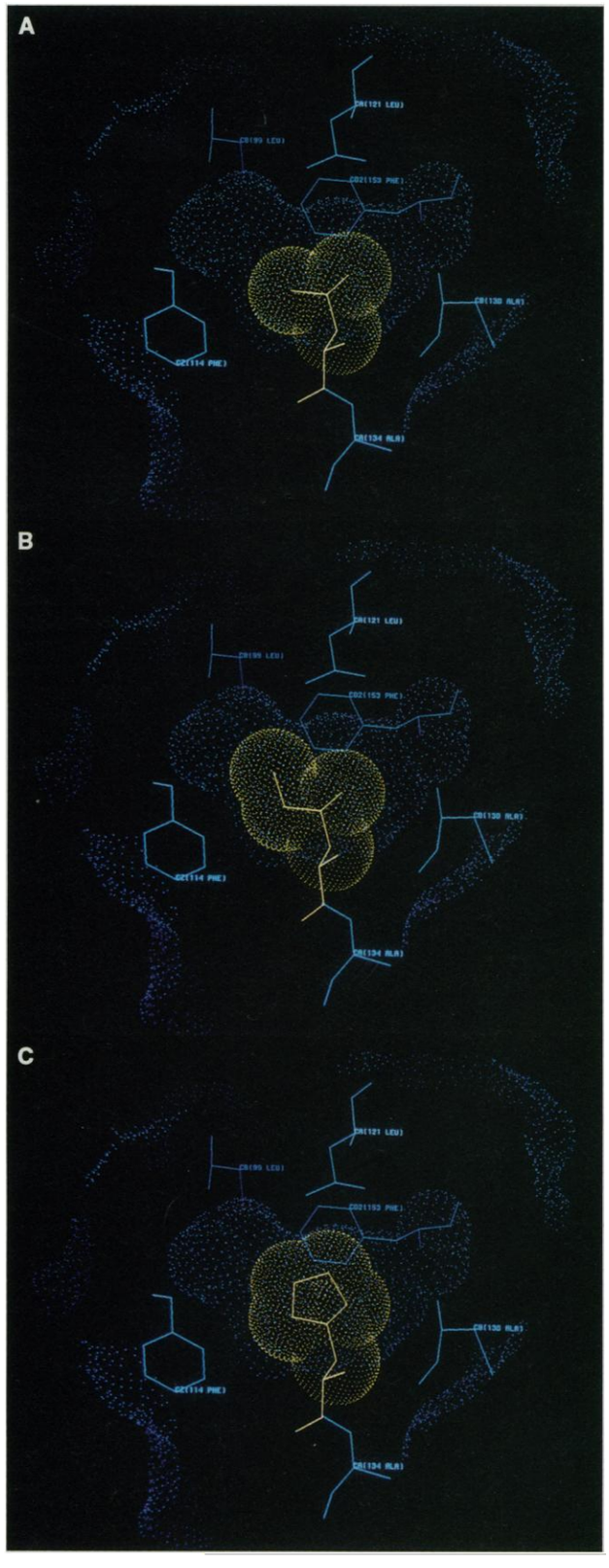
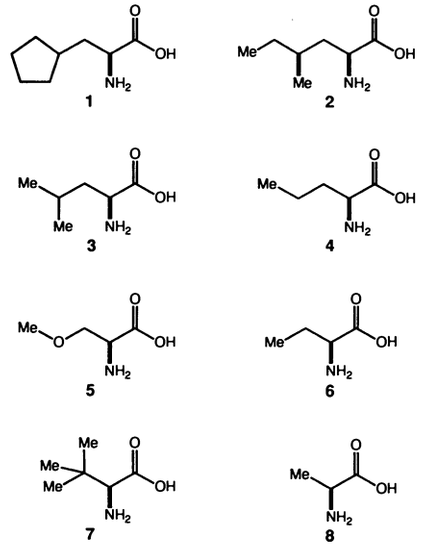


Fig. 1. Graphic representation of the side chain van der Waals surfaces (yellow) at position 133 in T4L of: (**A**) Leu, (**B**) *S,S*-2-amino-4-methylhexanoic acid **2**, and (**C**) *S*-2-amino-3-cyclopentylpropanoic acid **1**. Amino acids were designed with the model building program LEaP starting from the coordinates of Leu^{133} and by using standard geometries (*32*). The residues were then superimposed on the crystal structure. Visualization of the cavity (blue) was made possible by the generation of a molecular surface in the absence of a side chain at position 133 with MIDAS PLUS and represents the contact area between the protein and a water probe of 1.4 Å radius (*33*). Unnatural amino acid **1** was synthesized using the method of Evans (*34*) starting from 3-cyclopentylpropionyl chloride. Amino acid **2** was constructed according to the method of Seebach (*35*) with *S*-2-methyl-1-butanol.

Amino acids were substituted for Leu^{133} in T4L, a residue that is buried in the hydrophobic core of T4L and that defines one face of a large cavity (Fig. 1A). Mutations Leu^{133} → Ala and Leu^{133} → Phe have been shown not to significantly perturb overall protein structure, simplifying the analysis of mutations at this site (*3*, *11*). Substitutions for Leu^{133} were designed to: (i) bury more hydrophobic surface area, (ii) incrementally enlarge the cavity, (iii) alter the hydrophobicity of isosteric amino acid side chains (*14*, *15*), and (iv) alter side chain configurational entropy. Molecular dynamics (MD) calculations were carried out to evaluate the structural effects of these mutations on the surrounding protein (Table 1). Simulations of wild-type protein (Leu^{133}) gave low root-mean-square (rms) values for cavity residues, as expected. Simulation of the Leu^{133} → Phe mutant, as a test of the ability of MD to detect unfavorable interactions, yielded an unfavorable χ_1 dihedral of −91°, close to the Phe^{133} crystal structure value of −87° (*11*), and a higher rms for Leu^{121}, which appears to have bad van der Waals contacts in the Phe^{133} crystal structure (*16*).

Novel amino acid replacements for Leu were then designed and incorporated at position 133, and their influences on T4L stability were evaluated (scheme 1). *S,S*-2-Amino-4-methylhexanoic acid **2** and *S*-2-



Scheme 1. Amino acids substitutions for Leu^{133}; Me, methyl.

amino-3-cyclopentylpropanoic acid **1** should extend the side chain of Leu by one $-CH_3$ and two $-CH_2-$ groups, respectively, and fill the cavity more completely than Leu (**3**) without creating unfavorable packing contacts (Fig. 1). For the cyclic side chain there should be less loss of conformational entropy on folding than with a straight chain alkyl group. The rms deviations of cavity residues surrounding $Leu^{133} \rightarrow$ **1** and $Leu^{133} \rightarrow$ **2** from MD simulations agree quite well with those of Leu^{133}. *Tert*-leucine **7** buries approximately the same hydrophobic surface area as Leu but is expected to introduce significantly greater strain because of a clash between the γ carbons and the carbonyl oxygen of Ala^{129} in the preceeding turn of the α helix. Rotation of the side chain of **7** about χ_1 cannot alleviate this strain because it is symmetrically substituted, and so, to accommodate the steric bulk, Φ increases in the simulation (−53° to −59° in the protein and −56° to −66° in the isolated helix). Norvaline **4**, ethylglycine **6**, and alanine **8** were included to assay the effects of systematic removal of $-CH_3$ ($-CH_2-$) groups from Leu^{133}. Finally, O-methylserine **5**, which is isosteric with norvaline **4** but has a $CH_2 \rightarrow O$ substitution, was included to assess the effects of side chain solvation (polar and nonpolar) on protein stability. The simulations suggest that the norvaline **4**, methylserine **5**, and ethylglycine **6** side chains have similar conformations to Leu^{133}, except for a ~7° change in ψ, which is probably related to the smaller size of these residues and reflects slight repacking of the cavity (*1*, *3*). In addition, **5** has a smaller χ_1 dihedral angle (−61°) than Leu (−78°), which may be due to decreased bulk at the γ position and perhaps some electrostatic effects. Although the changes in dihedral angles and rms deviations are small, they do permit qualitative estimates of structural disruptions.

Incorporation of the above amino acids at position 133 of T4L was accomplished by in vitro suppression of a $Leu^{133} \rightarrow$ TAG stop mutation (encoded on plasmid pT4LL133am) (*17*) with a chemically aminoacylated suppressor tRNA derived from yeast $tRNA^{Phe}$ (*12*, *13*, *18*). We have used this method to site-selectively incorporate Phe analogs into β-lactamase (*12*) and a photoactivatable β-2-nitrobenzyl-Asp (*19*) as well as a wide variety of amino acids with novel backbone structures into T4L (*20*). When the in vitro–coupled transcription-translation system was programmed with pT4LL133am and supplemented with aminoacyl $tRNA_{CUA}$'s, full-length T4L was produced (lanes 4 to 11, Fig. 2). At 4 mM added magnesium acetate, amino acids **1** to **8** were incorporated with suppression efficiencies ranging from 27% (**8**) to 62% (**2**). In contrast, when $tRNA_{CUA}$ was omitted (lane 2, Fig. 2) or did not carry an amino acid (lane 3, Fig. 2), less than 1% full-length T4L was produced, compared to expression of wild-type T4L by pHSe54,97.TA (lane 1, Fig. 2) (*20*, *21*). Thus the in vitro system does not contain endogenous suppressor tRNAs capable of reading through the amber stop codon, nor do the aminoacyl tRNA synthetases present in the *Escherichia coli* S-30 extract aminoacylate the suppressor $tRNA_{CUA}$ with any of the 20 natural amino acids.

Mutant T4L's were purified to homogeneity (as judged by silver-stained denaturing polyacrylamide gel electrophoresis) from 5.0-ml in vitro protein synthesis reactions (*22*). The mutants have specific activities indistinguishable from that of wild-type T4L with the exception of $Leu^{133} \rightarrow$ **7**, whose activity is reduced fivefold (*23*). This decrease likely reflects the unfavorable

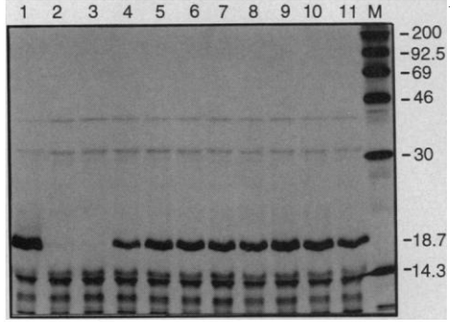


Fig. 2. Autoradiogram of in vitro suppression reactions labeled with L-[^{35}S]Met and containing the following plasmids and tRNAs: lane 1, pHSe54,97.TA (wild type); lane 2, pT4LL133am without $tRNA_{CUA}$; lane 3, pT4LL133am and 8 μg of full-length, unacylated $tRNA_{CUA}$; lane 4, pT4LL133am and 8 μg of alanyl-$tRNA_{CUA}$; lane 5, pT4LL133am and 8 μg of ethylglycyl-$tRNA_{CUA}$; lane 6, pT4LL133am and 8 μg of norvalyl-$tRNA_{CUA}$; lane 7, pT4LL133am and 8 μg of *O*-methylseryl-$tRNA_{CUA}$; lane 8, pT4LL133am and 8 μg of leucyl-$tRNA_{CUA}$; lane 9, pT4LL133am and 8 μg of *S,S*-2-amino-4-methylhexanoyl-$tRNA_{CUA}$; lane 10, pT4LL133am and 8 μg of *S*-2-amino-3-cyclopentylpropanoyl-$tRNA_{CUA}$; lane 11, pT4LL133am and 8 μg of *tert*-leucyl-$tRNA_{CUA}$. Lane M contains ^{14}C-methylated molecular weight standards. Cleared supernatants (20 μl) from terminated 30-μl in vitro reactions were incubated with 4 μl of 2.5 mg ml^{-1} ribonuclease (RNase) A for 15 min at 37°C and analyzed by 15% SDS–polyacrylamide gel electrophoresis. Each 30-μl reaction contained 7 μCi of L-[^{35}S]Met.

Table 1. Root-mean-square deviations and dihedral angles averaged over 20 ps of molecular dynamics. A 17 Å cap of water molecules was centered on the γ carbon of 133. Only residues defining the cavity (98, 99, 102, 106, 111, 114, 117, 118, 120, 121, 129 to 134, 138, 139, 146, 149, 150, and 153) and waters within the cap were allowed to move (*11*) (an assumption supported by the x-ray crystal structures of the Phe^{133} and Ala^{133} mutants). Data were collected after minimization and 10-ps constrained molecular dynamics. Simulations were carried out with AMBER 3A with the Weiner *et al.* force field (*36*). All of the bond lengths were held constant. A time step of 2.0 fs and a constant dielectric were used. The rms deviations from the native crystal structure for heavy atoms only were calculated from coordinates collected every 0.5 ps and averaged.

Residue 133	Rms deviations of cavity residues near 133						Dihedral angles (degrees)		
	102	114	117	121	129	153	Φ	ψ	χ_1
3 (wt)	0.16	0.16	0.18	0.10	0.09	0.13	−53	−40	−78
Phe	0.18	0.14	0.21	0.16	0.10	0.13	−50	−42	−91
2	0.15	0.11	0.23	0.11	0.08	0.12	−53	−40	−79
4	0.15	0.11	0.19	0.10	0.09	0.13	−55	−32	−76
1	0.17	0.13	0.19	0.12	0.08	0.14	−54	−40	−78
7	0.16	0.14	0.20	0.11	0.13	0.11	−59	−39	−65
6	0.17	0.12	0.18	0.12	0.09	0.12	−54	−35	−73
5	0.16	0.13	0.20	0.11	0.09	0.13	−56	−34	−61

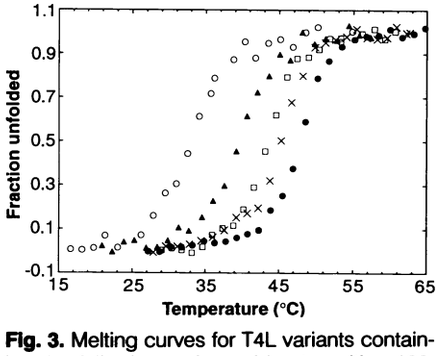


Fig. 3. Melting curves for T4L variants containing the following amino acids at position 133: (○), *O*-methyl serine **5**; (▲), norvaline **4**; (□), leucine **3**; (×), *S,S*-2-amino-4-methyl hexanoic acid **2**; and (●), *S*-2-amino-3-cylcopentylpropanoic acid **1**. Curves represent the average of three runs and were displayed with the relation $F_u = (CD_f - CD_x)/(CD_f - CD_u)$ where F_u is the fraction of unfolded protein, CD_f is the average CD signal (223 nm) at the earliest portion of the curve, CD_u is the average CD signal at the latest part of the curve, and CD_x is the average CD signal observed at given points between these two extremes.

Table 2. Amino acids incorporated at position 133 of T4L and the experimental and estimated thermodynamic stabilities of the resulting enzymes. Proteins containing these replacements were purified to homogeneity from 5.0-ml scale in vitro suppression reactions (*22*). The CD melts were performed in triplicate (except for Leu^{133} → **6**, one determination) according to the procedure of Becktel and Baase (*25*) with a Jasco 600 polarimeter for CD readings (*37, 38*). Protein concentrations used in the CD measurements were between 5 and 10 μg ml^{-1}.

Amino acid	T_m*	ΔH_{expt}† (kcal mol^{-1})	$\Delta\Delta G_{expt}$‡§ (kcal mol^{-1})	ΔSurface area‖ (Å^2)	$\Delta\Delta G_{calc}$‡ (kcal mol^{-1})				
					HE¶	Cavity#	Pack**	ΔS_{conf}††	Total
1	47.8 ± 0.3	98 ± 2	1.24 ± 0.09	+22.0	+0.5/+1.0	+0.5	+0.4	0.0	+1.4/+1.9
2	45.4 ± 0.2	103 ± 16	0.60 ± 0.13	+22.9	+0.6/+1.1	+0.5	+0.5	−0.5	+1.1/+1.6
3	43.5 ± 0.2	96 ± 4	0.00	0.0	0.0/0.0	0.0	0.0	0.0	0.0/0.0
4	39.5 ± 0.3	79 ± 4	−1.07 ± 0.10	−21.0	−0.5/−1.0	−0.4	−0.4	+0.4	−0.9/−1.4
5	32.9 ± 0.5	63 ± 13	−2.63 ± 0.56	−50.7	−2.3	−0.4	−0.4	+0.4	−2.7§§
6	~31		~−3.3‡‡	−49.3	−1.2/−2.3	−1.0	−1.0	+0.5	−2.7/−3.8
7				−4.6	−0.1/−0.2	−0.1	−0.1	+0.9	+0.6/+0.5

*The melting temperatures (T_m) of the enzymes correspond to the average of three determinations. The error estimates correspond to the 95% confidence limit of these measurements. †The ΔH is the enthalpy of unfolding at the T_m as determined by van't Hoff analysis of the CD melts. The error estimates correspond to the 95% confidence limit from three determinations. ‡Relative to the unfolded state so that a mutant with a positive $\Delta\Delta G$ is more stable than the wild-type Leu. §Isothermal $\Delta\Delta G$ values were calculated at 43.5°C with the use of a thermodynamic model in which a constant change in heat capacity ΔC_p was used, estimated (*3*) to be 2.5 kcal mol^{-1} K^{-1}. $\Delta\Delta G$ values were also calculated at the T_m of the mutants by using ΔH (96 kcal mol^{-1}) and ΔC_p (1.80 kcal mol^{-1}) of the wild-type enzyme according to the method of Dao-pin *et al.* (*39*) and were within the error estimates obtained for the isothermal $\Delta\Delta G$ determinations. ‖Nonpolar solvent-accessible surface area only. The difference in nonpolar solvent-accessible surface area compared to Leu was calculated with Richmond's algorithm (*40*), using a 1.4 Å radius water probe and Richards' van der Waals values (*41*) for the tetrapeptide *N*-acetyl-Asn-X-Ala-*N*-methyl amide in the extended conformation. ¶For the pairs of values, the hydrophobic effect (HE) is calculated with a value of 24 cal mol^{-1} Å^{-2} of buried nonpolar surface area (*28*) for the first entry and with a value of 47 cal mol^{-1} Å^{-2} (*26*) for the second. #Estimated from ΔG of sublimation of hydrocarbons, 20.7 cal mol^{-1} Å^{-2} (*27*). This term accounts for changes in dispersion interactions as well as changes in the mobility of surrounding residues. **Estimated from ΔH of melting of hydrocarbons, 20.3 cal mol^{-1} Å^{-2} (*27*), which accounts for more favorable packing interactions in the solid-like native state versus the liquid-like denatured state. ††Estimated from $\Delta G = -RT\ln(q_2/q_1)$ where q is the partition function for side-chain rotamers and R is the gas constant; q was calculated based on a *gauche* versus *trans* difference of 0.9 kcal mol^{-1} and was assumed to be 1 for the native state, except for **4**, which can assume two conformations. ‡‡$\Delta\Delta G$ was calculated at the mutant T_m (31°C) with ΔH (96 kcal mol^{-1}) and ΔC_p (1.80 kcal mol^{-1}) of the wild-type enzyme according to (*39*). §§Cavity, dispersion, and conformational entropy estimates are assumed to be the same as for norvaline (**4**).

steric interactions that result from introduction of a tertiary β center within an α helix, leading to perturbation of the three-dimensional protein structure (*24*) (Table 1). The thermal stabilities of the mutants were evaluated with purified protein (extensively dialyzed against 20 mM potassium phosphate, 25 mM KCl, pH 2.5) by determining the midpoint of the reversible thermal denaturation curve as monitored by CD with the use of a two-state denaturation model (*25*) (Fig. 3 and Table 2).

Semiquantitative estimates of stability differences based on hydrophobicity, packing effects, and side chain conformational entropy were also made for a comparison with the experimental values (Table 2). Transfer free energies ΔG's of hydrocarbons between hydrocarbon liquids and water were used to estimate the maximum ΔΔG of desolvating nonpolar amino acid side chains upon folding (*26, 27*). Two values were used: 24 cal mol^{-1} Å^{-2} of solvent-accessible surface area removed from water (*28*) and the more recent value of 47 cal mol^{-1} Å^{-2}, which reflects an additional entropy term due to size differences between solvent and solute (*26*). The difference in packing interactions between the liquid-like denatured and solid-like folded states was estimated with the value of the heats of fusion for alkanes derived by Nicholls *et al.* of 20.3 cal mol^{-1} Å^{-2} (multiplying this value by the difference in surface area compared to Leu gives an upper estimate of the stabilization or destabilization due to changes in packing) (*27*). Rather than use the entropy of melting to estimate the contribution of freezing side chain motion, side-chain configurational entropy ΔS_{conf} was calculated from the partition functions of the side chain dihedrals, giving entropy terms that are a function of the accessible conformations of the side chains rather than a function of their surface area. In general, the packing and ΔS_{conf} terms nearly cancel one another (*27*). Cavity formation or filling must be accounted for when comparing mutations of hydrophobic residues to transfer experiments (*1, 3*). The cost of cavity formation has been estimated from sublimation ΔG's to be ~20.7 cal mol^{-1} Å^{-2} (*27*). Although this number gives a maximum value because it assumes that the entire change in surface area contributes either to filling a cavity or creating one, the value is very close to that determined experimentally (20 cal mol^{-1} Å^{-2}) by Eriksson *et al.* (*3*), where repacking is taken into consideration.

In vitro synthesized T4L containing Leu at position 133 (generated with leucyl $tRNA_{CUA}$) melted at 43.51° ± 0.19°C, in good agreement with our previous determinations under identical conditions (*20*). Stepwise removal of $-CH_3$ ($-CH_2-$) groups led to a stepwise decrease in protein stability (*4, 5, 7*), but the change was nonlinear; ΔΔG became increasingly large with decreasing side chain surface area (ΔΔG = 1.1 kcal mol^{-1} for **3** versus **4** and ΔΔG = 2.2 kcal mol^{-1} for **4** versus **6**). This trend is accurately reflected in the semiquantitative estimates of the ΔΔG's and results from the net effect of changes in the hydrophobic, entropic, packing, and cavity terms (*29*).

Unnatural amino acid mutagenesis allows us to directly correlate the transfer of amino acid side chains from aqueous solution to the protein hydrophobic core with transfer from water to octanol in the absence of other variables (*30*). The difference in stability between the mutant enzyme containing **4** and the mutant enzyme containing the isosteric residue **5** ($-CH_2-$ replaced by $-O-$), 1.7 kcal mol^{-1}, compares favorably with the difference in the octanol-water partitioning ratios of the respective *N*-acetyl amide derivatives (ΔΔG = 1.8 kcal mol^{-1}) (*31*), suggesting in this case that octanol-water partitioning ratios provide an accurate measure of solvation effects without additional entropy terms (*26*).

The extended amino acids **2** and **1**, which were designed to fill the cavity with minimal strain, stabilize T4L by 1.9° (0.60 kcal mol^{-1}) and 4.3°C (1.24 kcal mol^{-1}), respectively, demonstrating that amino acids that increase the bulk of buried hydrophobic residues without concomitant strain can significantly increase protein stability (*4, 7, 9*). The experimental difference between the stabilities of the mutant proteins containing these two amino acids at position 133 (0.54 kcal mol^{-1}) in part reflects the effects of increasing ΔS_{conf} while keeping surface area relatively constant (Δsurface area = 0.9 Å^2). Again, this value is in reasonable agreement with the calculated ΔS_{conf} and emphasizes the importance of this term in interpreting mutagenesis results. The Leu^{133} → **1** and Leu^{133} → **2** mutants are likely not as stable as the

calculated estimates because of nonideal packing interactions or incomplete filling of the cavity.

Hydrophobicity, packing effects, cavity formation, and side chain conformational entropy all play important roles in determining protein stability and must be considered in any mutational study. Additional mutations coupled with intensive free energy calculations should provide greater insight into the various contributions of these factors to protein folding and stability. This approach can be applied to other forces that determine protein folding, including hydrogen bonding, electrostatic interactions, and main chain effects.

REFERENCES AND NOTES

1. K. A. Dill, *Biochemistry* **29**, 7133 (1990).
2. M. J. Behe, E. E. Lattman, G. D. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4195 (1991).
3. A. E. Eriksson *et al.*, *Science* **255**, 178 (1992).
4. M. Matsumura, W. J. Becktel, B. W. Matthews, *Nature* **334**, 406 (1988).
5. J. T. Kellis, Jr., K. Nyberg, D. Sali, A. R. Fersht, *ibid.* **333**, 784 (1988).
6. W. S. Sandberg and T. C. Terwilliger, *Science* **245**, 54 (1989).
7. K. Yutani, K. Ogasahara, T. Tsujita, Y. Sugino, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4441 (1987).
8. D. Shortle, W. E. Stites, A. K. Meeker, *Biochemistry* **29**, 8033 (1990).
9. W. A. Lim and R. T. Sauer, *J. Mol. Biol.* **219**, 359 (1991).
10. _____, *Nature* **339**, 31 (1989).
11. M. Karpusas, W. A. Baase, M. Matsumura, B. W. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8237 (1989).
12. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **244**, 182 (1989).
13. J. A. Ellman, D. Mendel, C. J. Noren, S. Anthony-Cahill, P. G. Schultz, *Methods Enzymol.* **202**, 301 (1991).
14. T. L. Doering *et al.*, *Science* **252**, 1851 (1991).
15. R. O. Heuckroth, L. Glaser, J. I. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8795 (1988).
16. Phe adopted a low-energy dihedral angle (χ_1 = −56°) in a minimized isolated helix (Trp^{129} to Ala^{134}), suggesting that the unfavorable angle in the native protein is caused by unfavorable steric interactions with another residue within the core of the protein. That residue is Ser^{117}, which forms a hydrogen bond with the carbonyl O of the Asn^{132} side chain and is thus unlikely to move to accommodate Phe^{133}.
17. Mutagenesis was performed according to the Eckstein method with 5′ TCT ACT TTT AGC CTA GTT AAC TGC 5′ (mismatches underlined) as the mutagenic oligonucleotide; J. R. Sayers, W. Schmidt, F. Eckstein, *Nucleic Acids Res.* **16**, 791 (1988).
18. C. J. Noren *et al.*, *ibid.* **18**, 83 (1990).
19. D. Mendel, J. A. Ellman, P. G. Schultz, *J. Am. Chem. Soc.* **113**, 2758 (1991).
20. J. A. Ellman, D. Mendel, P. G. Schultz, *Science* **255**, 197 (1992).
21. Plasmid pHSe54,97.TA encodes a cysteine-free T4L behind a twin *tac* promoter.
22. Proteins were purified by polyethyleneimine·HCl precipitation of the crude in vitro suppression reaction followed by passing the supernatant over DEAE- and CM-cellulose cartridges in tandem. T4L's were eluted from the CM-cellulose cartridge with a salt gradient (pH 7.5). A second round of cation exchange chromatography at pH 4.9 afforded homogeneous protein. See also (*20*).
23. Suppression efficiencies, determined by scintillation counting of SDS-polyacrylamide gel slices, were used to estimate the amount of T4L produced. These values were then compared to the rates of lysis of *E. coli* strain NAPIV cells by the mutant proteins to estimate the specific activity relative to wild-type enzyme. See also (*19*) and (*20*).
24. P. C. Lyu, J. C. Sherman, A. Chen, N. R. Kallenbach, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5317 (1991).
25. W. J. Becktel and W. A. Baase, *Biopolymers* **26**, 619 (1987).
26. K. A. Sharp, A. Nicholls, R. Friedman, B. Honig, *Biochemistry* **30**, 9686 (1991); K. A. Sharp, A. Nicholls, R. F. Fine, B. Honig, *Science* **252**, 106 (1991).
27. A. Nicholls, K. A. Sharp, B. Honig, *Proteins* **11**, 281 (1991).
28. C. Chothia, *J. Mol. Biol.* **105**, 1 (1976).
29. Cavity terms can also be estimated from cavity surface areas (*3*). Changes in cavity surface areas using a 1.4 Å² probe are: Leu^{133} → **1**, −7.7 Å²; Leu^{133} → **2**, −14.6 Å²; Leu^{133} → **4**, 16.4 Å²; Leu^{133} → **6**, 66.1 Å². Estimates using these values give cavity terms similar to those reported in Table 2.
30. The only isosteric replacement provided by the natural amino acids is Val to Thr; however, this would result in the introduction of an unsatisfied hydrogen bond.
31. Octanol-water partition values were determined according to the protocol of J. Fauchere and V. Pliska [*Eur. J. Med. Chem.* **18**, 369 (1983)] by vapor phase chromatography. The $\Delta\Delta G$ value of 1.8 kcal mol^{-1} for **4** versus **5** represents the average of five determinations.
32. C. Schafmeister, LEaP, University of California, San Francisco (1991).
33. C. Huang *et al.*, MIDAS PLUS, Computer Graphics Lab, University of California, San Francisco (1991).
34. D. A. Evans, T. C. Britton, J. A. Ellman, R. L. Dorow, *J. Am. Chem. Soc.* **112**, 4011 (1990).
35. R. Naef and D. Seebach, *Helv. Chim. Acta.* **68**, 135 (1985).
36. G. Seibel, H. C. Singh, P. K. Weiner, J. Caldwell, P. A. Kollman, AMBER 3.0, Rev. A, University of California, San Francisco (1991); S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, *J. Comput. Chem.* **7**, 230 (1986).
37. Temperature was measured as described (*20*).
38. W. J. Becktel and J. A. Schellman, *Biopolymers* **26**, 1859 (1987).
39. S. Dao-pin, D. E. Anderson, W. A. Baase, F. W. Dahlquist, B. W. Matthews, *Biochemistry* **30**, 11521 (1991).
40. T. J. Richmond, *J. Mol. Biol.* **178**, 63 (1984).
41. F. M. Richards, *Annu. Rev. Biophys. Bioeng.* **6**, 151 (1977).

42. We are grateful for support by the Director, Office of Energy Research, Office of Basic Energy Sciences, Division of Material Sciences, and also by the Division of Energy Biosciences of the U.S. Department of Energy (DE-AC03-76SF00098). D.M. was supported by American Cancer Society postdoctoral fellowship PF-4014A and J.A.E. by NSF postdoctoral fellowship CHE-8907488. P.G.S. is an NSF Waterman Awardee. D.L.V. was supported by NIH through a Biotechnology Training Grant (GM-08388-02), and P.A.K. thanks NIH (GM-29072), for research support. We also thank the NIH for support of the UCSF Computer Graphics Laboratory, R. Langridge, P.I. (NIH-RR-1081).

26 February 1992; accepted 23 April 1992

Nuclear Localization of *Agrobacterium* VirE2 Protein in Plant Cells

Vitaly Citovsky, John Zupan, Debra Warnick, Patricia Zambryski*

The *Agrobacterium* single-stranded DNA (ssDNA) intermediate T-strand is likely transferred to the plant cell nucleus as a complex with a single VirD2 molecule at its 5′ end and multiple VirE2 molecules along its length. VirD2 contains a nuclear localization signal (NLS); however, because the T-strand is principally coated with VirE2 molecules, VirE2 also might assist in nuclear uptake. Indeed, VirE2 fused to a reporter protein localizes to plant cell nuclei, a process mediated by two amino acid sequences with homology to the bipartite NLS of *Xenopus* nucleoplasmin. Moreover, tumorigenicity of an avirulent *virE2* mutant is restored when inoculated on transgenic plants expressing VirE2, supporting in planta function of VirE2.

The interaction of *Agrobacterium* with plant cells results in crown gall tumors. Most functions for *Agrobacterium*–plant cell DNA transfer are carried on the Ti (tumor-inducing) bacterial plasmid. One portion, the T-DNA, is copied and transferred to the plant cell, but the products that mediate its movement are encoded by a separate virulence (*vir*) region. After induction of *vir* gene expression by small phenolic molecules excreted from wounded plant cells, a single-stranded copy of the T-DNA (T-strand) is generated and transferred [reviewed in (*1*)].

The T-strand associates with two protein products of the *vir* region, VirD2 and VirE2. The VirD2 protein is bound to the 5′ end of the T-strand (*2*), and an ssDNA binding protein (SSB), VirE2, coats the T-strand along its entire length (*3*, *4*). The T-strand associated with VirD2 and VirE2 is designated the T-complex (*5*). The T-complex travels from *Agrobacterium* into the plant cell where the T-DNA is integrated into the plant nuclear genome.

Recently, Howard *et al.* (*6*) identified a bipartite NLS at the COOH-terminus of VirD2; deletion of this sequence reduced *Agrobacterium* tumorigenicity (*7*). Thus, it was proposed that VirD2 mediates nuclear

V. Citovsky, J. Zupan, P. Zambryski, Department of Plant Biology, Genetics and Plant Biology Building, University of California, Berkeley, CA 94720.
D. Warnick, Sandoz Crop Protection Corporation, Palo Alto, CA 94304.

*To whom correspondence should be addressed.