Methionine and Alanine Substitutions Show That the Formation of Wild-Type-like Structure in the Carboxy-Terminal Domain of T4 Lysozyme Is a Rate-Limiting Step in Folding†

Nadine C. Gassner,‡ Walter A. Baase,§ Joel D. Lindstrom,§ Jirong Lu,‡,|| Frederick W. Dahlquist,‡ and Brian W. Matthews*,§,⊥

Institute of Molecular Biology, Howard Hughes Medical Institute, and Departments of Chemistry and Physics, 1229 University of Oregon, Eugene, Oregon 97403-1229

Received July 6, 1999; Revised Manuscript Received August 25, 1999

ABSTRACT: In an attempt to identify a systematic relation between the structure of a protein and its folding kinetics, the rate of folding was determined for 20 mutants of T4 lysozyme in which a bulky, buried, nonpolar wild-type residue (Leu, Ile, Phe, Val, or Met) was substituted with alanine. Methionine, which approximated the size of the original side chain but which is of different shape and flexibility, was also substituted at most of the same sites. Mutations that substantially destabilize the protein and are located in the carboxy-terminal domain generally slow the rate of folding. Destabilizing mutations in the amino-terminal domain, however, have little effect on the rate of folding. Mutations that have little effect on stability tend to have little effect on the rate, no matter where they are located. These results suggest that, at the rate-limiting step, elements of structure in the C-terminal domain are formed and have a structure similar to that of the fully folded protein. Consistent with this, two variants that somewhat increase the rate of folding (Phe104 → Met and Val149 → Met) are located within the carboxy-terminal domain and maintain or improve packing with very little perturbation of the wild-type structure.

The analysis of protein folding by the combination of protein engineering and kinetics has yielded a number of important insights regarding possible pathways and mechanisms (*1*−*8*).

The effect of mutations on the kinetics of folding can be more pronounced and is more complex than the effect on equilibrium thermodynamics (*2*). It is possible that a single key residue, or a small subset of residues, might play a critical role in the rate-limiting step for folding. Recently it has been suggested that folding rates can be slowed because of the stabilization of intermediates along the folding pathway (*9*). A similar view can arise when the folding pathway is considered from a folding funnel perspective (*10*). Thus it is also possible that mutational substitutions in T4 lysozyme might slow folding by differential stabilization of the early intermediate observed in the T4 lysozyme folding pathway (*4*).

To try to clarify concerns of this sort, and also in an attempt to obtain insights into the kinetics of folding of T4 lysozyme, we have used an approach in which alanine and methionine are substituted in parallel at a number of sites within T4 lysozyme (Figure 1). Alanine substitutions within the core of the protein typically have a much larger effect on structure and stability than surface substitutions (*11*, *12*). It is to be expected that internal sites will be more important for the kinetics of folding (*4*) and the present analysis was therefore restricted to such locations. At 16 sites for which the residue in wild-type lysozyme is leucine, isoleucine, phenylalanine, or valine, both alanine and methionine were substituted. At two additional sites where the residue in the native protein was already methionine, alanines were substituted. Finally, in one case in which the wild-type protein had alanine, a methionine was incorporated. Methionine was chosen because it is roughly the same size as leucine, isoleucine, phenylalanine, and valine and is flexible enough to more-or-less occupy the same volume.

MATERIALS AND METHODS

Chemicals and Solutions. Chemicals were "Baker analyzed" except for urea (Amresco). Solutions were made with Nanopure water (Barnstead). Urea stock solutions (6 M) were made and deionized with AG 501-X8(D) mixed-bed resin (Bio-Rad) until immediately prior to use. The PHM84 meter, a GK2401C electrode, and the S series of standards (Radiometer) were used to determine solution pH values. (1*S*)-(+)-10-Camphorsulfonic acid (Aldrich) in water was used for intensity calibration of circular dichroism (CD) measurements ($\Delta\epsilon = +2.37$ M^{-1} cm^{-1} at 290.5 nm).

† This work was supported in part by NIH grants GM57766 to F.W.D. and GM21967 to B.W.M.

* To whom correspondence should be addressed: Institute of Molecular Biology, 1229 University of Oregon, Eugene, OR 97403; Phone (541) 346-2572; FAX (541) 346-5870; E-mail brian@uoxray.uoregon.edu.

‡ Institute of Molecular Biology and Department of Chemistry.

§ Institute of Molecular Biology and Howard Hughes Medical Institute.

|| Present address: Eli Lilly and Company, Lilly Corporate Center, DC1505, Indianapolis, IN 46285.

⊥ Department of Physics.

10.1021/bi9915519 CCC: $18.00
Published on Web 10/15/1999

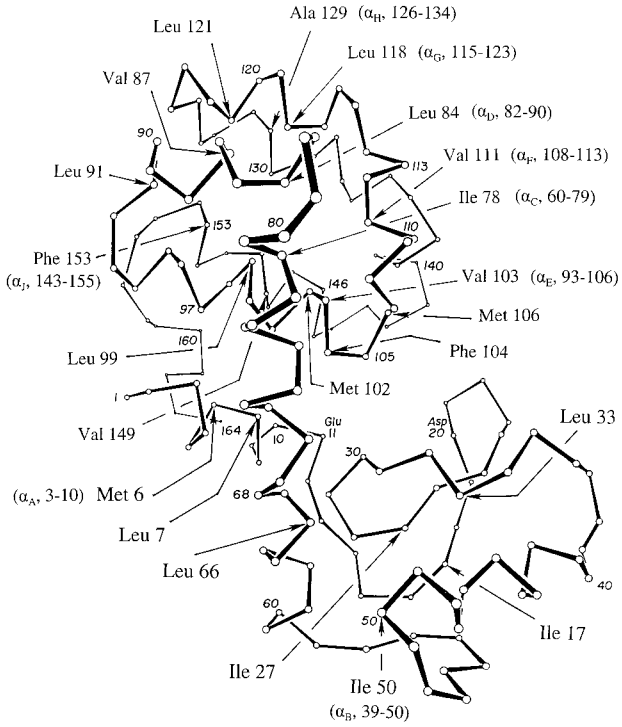

FIGURE 1: Sites within T4 lysozyme that were substituted with alanine and/or methionine. For representative amino acids that are located within α-helices, the name of the α-helix and the residues in the helix are indicated in parentheses.

Protein Purification. Mutations were introduced into the pseudo-wild-type (WT*)[1] T4 lysozyme gene (*13*) by the method of Kunkel et al. (*14*). Proteins were plasmid-amplified, purified, and crystallized by procedures standard for soluble T4 variants (*15*−*18*). Protein concentrations were determined by means of absorbance at 280 nm using a Varian 2290 spectrophotometer and an extinction coefficient of 24 170 M^{-1} cm^{-1} (*19*).

Measurement of the Kinetics of Folding. Solutions were outgassed for 10 min at $^1/_3$ atm by use of an outgassing station with controlled stirring from Microcal. Proteins were unfolded at 1.00 (± 0.03) mg/mL in 2.0 M urea, 20 mM H_3PO_4, and 41.5 mM HCl, pH 2.0 at 20 °C. The degree of unfolding was monitored by both circular dichroism (CD) and fluorescence. Refolding was initiated by rapid mixing with 66.7 mM $K_{3/2}H_{3/2}PO_4$ (also outgassed) at a ratio of 19:1 to give a final solution that was 0.050 mg/mL in protein and 64.4 mM KPO_4, 0.1 M urea, and 2.1 mM KCl, pH 6.7, also at 20 °C.

Stopped-flow mixing experiments were done in a Bio-Logic TC100/15 Z cell and the Bio-Logic SFM/3 (Molecular Kinetics, Pullman, WA; control software V1.08) mixing head mounted on a Jasco J-720 circular dichroism (CD) spectropolarimeter (control software V1.32.00) with a mounting bracket equipped with UV-grade focusing optics (Jasco). Cell and injection syringe temperatures were controlled at 20.0 °C by a Neslab RTE−111 circulating water bath.

The J-720 was operated at 223 nm with a bandwidth of 2 nm and was modified to record photomultiplier tube current (rather than voltage) in channel 2. Data were collected at a response time of 1 ms with the current control feedback circuit disabled (manual mode) except for the three most slowly refolding mutants, which were collected at 2 ms. In this configuration the J-720 operated at a constant photomultiplier tube voltage of between 287 and 310 depending on the age of the xenon arc lamp.

The time course of the circular dichroism was taken as the AC component of the photomultiplier tube signal (channel 1) divided by the DC component of the photomultiplier tube signal (channel 2; *20*). Autocorrelation (AUTOCOR; *21*) of the noise signal of channel 1 demonstrated that the time constant of that channel was 1.5 ms.

Fluorescence emission intensity was collected simultaneously with CD at right angles to the transmitted beam by using an Edmund Scientific fiber optical cable ($^1/_8$ in. diameter) and a spherical mirror that reflected emitted light back through the center of the cell. The optical cable transmitted light above 260 nm to a blocking filter (cutoff below 338 nm) and an EMI 6256S phototube operating at 920 V. After conversion of this signal from current to voltage by a PAR 181 preamplifier, the signal components below 6000 Hz were isolated by an 8-pole butterfly filter (Maxim Corp.) and digitized via a DT2805 (Data Translation) analogue-to-digital interface board. Autocorrelation of the noise in this signal confirmed that the time constant of the fluorescence signal was clearly less than 1 ms.

Timing relationships among channel 1, channel 2, and the fluorescence intensity were established by variation of the concentration of lysozyme (to compare channel 1 and channel 2 at 223 nm) and of *N*-acetyl-L-tryptophanamide (to compare channel 2 and fluorescence) by very slow mixing. This demonstrated that channel 1 was 1.2 ms behind channel 2 and that channel 2 was 1 ms behind the fluorescence.

The SFM/3 was operated at a nominal dead time of 6.5 ms. However, transients were seen in the raw data for up to an additional 5 ms, increasing the instrumental deadtime to 11.5 ms. Data taken within this overall instrumental dead time were discarded.

Each refolding experiment consisted of at least 50 individual pushes in the SFM/3. Pushes were culled if variation of the transmitted light intensity (channel 2) was more than 2.5% during the observation phase of a push. The remaining pushes were averaged. Fluorescence and CD data were fit with Origin (version 4.10; Microcal Software, Inc.) to the sum of two exponentials to obtain amplitudes and rate constants except for the L99A, F153A, and M102A/M106A data, which were fit to a single exponential. For those sites for which more than one set of pushes was done, the simple averages of the rate parameters determined by Origin are shown.

Thermal Stability. Thermal unfolding and van't Hoff analysis was done as described previously (*18*) in 0.10 M sodium chloride, 1.4 mM acetic acid, and 8.6 mM sodium acetate, pH 5.4, with protein concentrations between 0.5 and 1.5 μM. Reversibility averaged 80% for data presented here. A ΔC_p value of 2500 cal/mol·deg at an isotherm of 59 °C was used to compute ΔG.

X-ray Crystallography. The structures of WT* (Brookhaven Protein Data Bank access code 1L63) and all alanine mutants except I78A and L91A were taken from previous work (*12*, *22*−*25*). All variants crystallized (*18*) in space group $P3_2$-21, isomorphous with WT*. X-ray diffraction data were collected at room temperature using a San Diego Multiwire

[1] Abbreviations: WT*, cysteine-free pseudo-wild-type lysozyme; CD, circular dichroism; SFM, stopped-flow mixing.

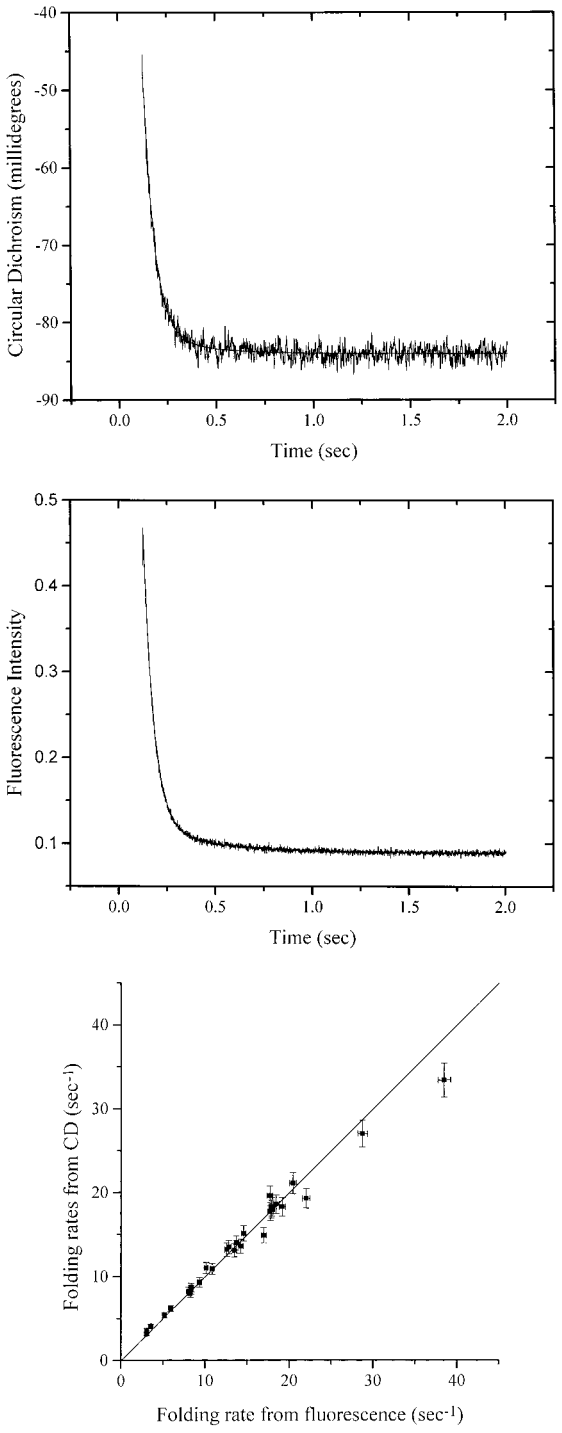

FIGURE 2: (a, top). Folding of WT* lysozyme at 0.05 mg/mL as monitored by circular dichroism. The intensity for the unfolded protein in 2 M urea, 0.042 N HCl, and 0.020 M H_3PO_4, pH 2.0 at 20 °C, is −14 mdeg. For the folded protein in 0.1 M urea, 0.002 M KCl, and 0.64 M $K_{3/2}H_{3/2}PO_4$, pH 6.7 at 20 °C, the intensity is −88 mdeg. The trace begins after mixing (112 ms) and instrument dead time (11.5 ms). (b, middle) Folding of WT* lysozyme monitored by fluorescence, under the same conditions described for panel a. The fluorescence for the unfolded protein is 0.3 unit on the arbitrary scale used in the figure, increases rapidly during the burst phase, and then decreases as shown. The value for the fully unfolded protein is 0.09 unit. (c, bottom) Comparison of the principal component of the rate of folding for different lysozyme mutants as monitored by circular dichroism and by fluorescence.

System area detector (*26*) for all crystals except I78A, I78M, and L91A. These three data sets were collected on a Molecular Structure Corp. R-Axis IV, integrated with DENZO and scaled with SCALEPACK (*27*). Refinement of atomic coordinates was carried out according to standard protocols with TNT (*28*, *29*). The final models do not include solvent molecules with *B*-factors higher than 80 Å^2. L118A, which has been previously described (*12*, *22*), was subjected to several additional cycles of positional and *B*-factor refinement and has been submitted to the Brookhaven Protein Data Bank along with all other structures solved here (access codes 1CTW, 1CU0, 1CU2, 1CU3, 1CU5, 1CU6, 1CUP, 1CUQ, 1CV0, 1CV1, 1CV3, 1CV4, 1CV5, 1CV6, 1CVK, and 1QSQ).

Table 1: Rates of Refolding of Alanine- and Methionine-Substituted Lysozymes

alanine variants		methionine variants		
protein	rate, k_{Ala} (s^{-1})	protein	rate, k_{Met} (s^{-1})	relative rate k_{Met}/k_{Ala}
M6A	19.3	WT* (M6)	18.6	1.0
L7A	28.6			
I17A	23.7	I17M	21.0	0.9
		I27M	18.4	
		I27M/L33M	17.8	
L33A	16.9	L33M	19.1	1.1
I50A	16.1	I50M	17.3	1.1
L66A	15.5	L66M	17.9	1.2
I78A	14.9	I78M	11.0	0.7
L84A	4.1	L84M	8.7	2.1
V87A	6.2	V87M	13.5	2.2
L91A	8.4	L91M	15.1	1.8
L99A	3.6	L99M	17.7	4.9
		I100M	10.3	
M102A	8.2	WT* (M102)	18.6	2.3
M102A/M106A	5.4	WT*	18.6	3.4
V103A	13.6	V103M	21.1	1.6
F104A	19.6	F104M	33.4	1.7
M106A	10.9	WT* (M106)	18.6	1.7
V111A	18.3	V111M	13.1	0.7
L118A	8.0	L118M	14.0	1.8
L121A	9.3	L121M	18.3	2.0
WT* (A129)	18.6	A129M	19.3	1.0
V149A	18.0	V149M	27.0	1.5
F153A	3.2	F153M	13.2	4.1

RESULTS

The sites in T4 lysozyme at which alanine and/or methionine substitutions were made are shown in Figure 1.

The refolding of WT*, as monitored by circular dichroism (CD) and fluorescence emission, is shown in Figure 2. Forty to fifty percent of the CD signal (Figure 2a) develops within the overall dead time of the instrument, which was 11.5 ms. After the dead time, the signal is well modeled as the sum of two exponentials, a principal relaxation component (18.6 s^{-1}) followed by a low-amplitude, slow relaxation component (2.7 s^{-1}). At longer times, the CD signal tends to the CD of folded protein in a monotonic manner. The fluorescence intensity (Figure 2b) also shows a principal relaxation component followed by a low amplitude, slow relaxation component. Refolding rates as determined from the principal decay components from both fluorescence and CD are in reasonable agreement for the WT* and mutant proteins (Figure 2c).

The refolding rate for WT* is 18.6 s^{-1}. For the alanine mutants the rates tend to be slower, ranging from 3 to 20 s^{-1} (Table 1; Figure 3). For the methionine variants the range is 9−33 sec^{-1}. The relative reproducibility in the rate

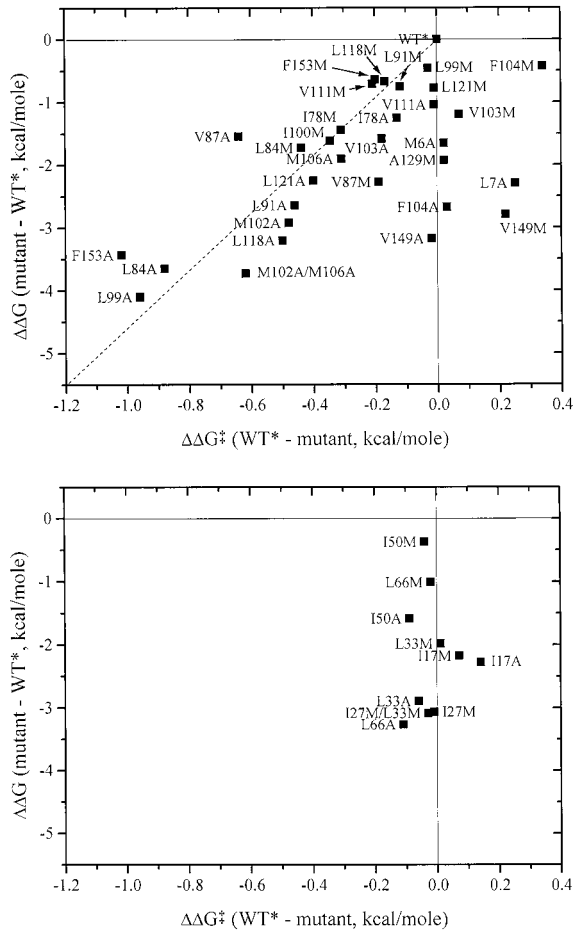

FIGURE 3: (a, top) Comparison of the effect of mutations within the carboxy-terminal domain on the stability of T4 lysozyme and on the kinetics of folding. The "carboxy-terminal" domain of T4 lysozyme is conventionally defined to include residues 80−162, as well as the A-helix (residues 1−10) (*33*). $\Delta\Delta G$ is the change in the Gibbs free energy of unfolding of the mutant relative to WT*. $\Delta\Delta G^{\ddagger}$ is derived from rate constants for folding (Table 1) by the equation $\Delta\Delta G^{\ddagger} = -RT \ln (k_{WT*}/k_{mut})$ where k_{mut} and k_{WT*} are the observed refolding rates for the mutant and WT*, respectively. The dashed line is drawn with a slope of $\Delta\Delta G^{\ddagger}/\Delta\Delta G = 0.22$ (see text). (b, bottom). Plot, similar to that in panel a, for mutations located within the amino-terminal domain of T4 lysozyme (residues 15−59).

constants was about ±6% as monitored by circular dichroism and ±2% as monitored by fluorescence. The thermal stabilities of all the variants are summarized in Table 2.

The following sections highlight the characteristics of those mutant crystal structures not described previously (Table 3). Figures are included only for a subset of the mutants, in particular those showing the largest changes in structure or kinetics relative to WT*.

Isoleucine 78 to Alanine and Methionine. Ile78 is located toward the carboxy terminus of the C-helix (residues 60−79; Figure 1). On truncation to alanine there are relatively large structural changes, compared to other such mutants (see below). The backbone in the immediate vicinity of the substitution moves 0.8−1.0 Å toward the site vacated by the isoleucine side chain (Figure 4a,b). Concomitant with this structural change, the C-helix becomes somewhat bent and the D-helix undergoes a bodily rotation of 1°. Atoms within the side chains of Leu84, Leu99, Ile100, and Val103 also move 0.5−0.7 Å inwards to partly fill the space that was occupied by Ile78.

In many respects the structural changes associated with the Ile78 → Met replacement are the opposite of those observed for Ile78 → Ala, although less in magnitude (Figure 4c,d). The backbone at site 78 moves outwards by 0.7 Å relative to WT*, and the D-helix rotates by 3° in a direction opposite to that seen for Ile78 → Ala. Overall, site 78 is characterized by considerable plasticity, relaxing inward when the smaller alanine side chain is substituted and moving outward to better accommodate the shape of the methionine side chain.

Leucine 84 to Alanine and Methionine. Truncation of Leu84 to alanine results in substantial distortion in the F-helix and formation of a channel to the surface (*25*). In contrast, replacement with methionine leaves the structure quite close to that of WT*. Met84 is well-ordered and adopts an unstrained rotamer conformation similar to that of WT*.

Valine 87 to Alanine and Methionine. When Val87 is replaced with alanine there are relatively modest changes in structure, including an overall shift in helix D, which includes residue 87, of about 0.35 Å toward the putative cavity (*25*). When methionine is substituted, the increase in volume of the side chain causes somewhat larger backbone shifts (up to 0.7 Å) in the opposite direction (not shown).

Leucine 91 to Alanine and Methionine. Located in a loop between the D- and E-helices, the main chain of Leu91 is 16% solvent-exposed, although the side chain is completely buried. As is apparent from both the electron density maps and the refined structures (not shown), there are minimal changes in structure associated with either the alanine or methionine substitution.

Valine 103 to Alanine and Methionine. In common with many of the alanine replacements, the truncation of the larger WT* side chain is associated with modest shifts of surrounding atoms toward the vacated space. In this case the largest such shift is 0.6 Å for the Val111 side chain (*25*). In contrast to most of the methionine replacements, however, the introduction of Met103 causes shifts both toward and away from the mutated side chain. One of the methyl groups of Val111, for example, moves 0.7 Å inward (somewhat like its behavior in the Val111 → Ala mutant). Atoms within the side chain of Leu84 and main chain of Glu108, however, move outwards by 0.5−0.8 Å.

Phenylalanine 104 to Alanine and Methionine. The substitution Phe104 → Ala opens a channel to the surface of the protein (*24*). For the methionine replacement, however (Figure 5), there are only minor adjustments (up to 0.3−0.4 Å) in atoms surrounding the substituted side chains.

Valine 111 to Alanine and Methionine. Helix F, which includes Val111, is short and relatively mobile. When the valine is truncated to alanine, main-chain atoms within the helix move up to 0.5 Å toward the putative cavity (*25*). Conversely, when the larger methionine side chain is substituted at site 111, the F-helix is substantially distorted with some main-chain atoms moving up to 2 Å (Figure 6).

Leucine 118 to Alanine and Methionine. Both the alanine and methionine substitutions at site 118 leave the structure largely unaffected (*12*, *22*; this work). The methionine structure, especially, is extremely similar to WT*, made possible by the fact that the C^{ϵ} methyl group of the methionine occupies a site where the C^{δ} methyl of the leucine makes contact with solvent (not shown).

Table 2: Thermostabilities of Alanine and Methionine Mutant Lysozymes[a]

alanine variants				methionine variants			
protein	T_m (°C)	ΔH (kcal/mol)	$\Delta\Delta G$ (kcal/mol)	protein	T_m (°C)	ΔH (kcal/mol)	$\Delta\Delta G$ (kcal/mol)
M6A	60.8	121	−1.6	WT* (M6)	65.3[b]	130[b]	
L7A	59.0	112	−2.3				
I17A	58.9	93	−2.3	I17M	59.4	102	−2.2
				I27M	55.2	62	−3.1
				I27M/L33M	55.0	61	−3.1
L33A	56.5	83	−2.9	L33M	60.0	108	−2.0
I50A	61.1	116	−1.6	I50M	64.7	120	−0.4
L66A	55.2	84	−3.3	L66M	62.6	122	−1.0
I78A	61.8	127	−1.2	I78M	61.6[b]	117[b]	−1.5[a]
L84A	54.8	101	−3.7	L84M	60.4[b]	110[b]	−1.9[a]
V87A	61.0	127	−1.5	V87M	59.0	113	−2.3
L91A	57.9	113	−2.6	L91M	63.3[b]	125[b]	−0.8[b]
L99A	53.6	104	−4.1	L99M	64.0[c]	134[c]	−0.4[c]
M102A	57.1	108	−2.9	WT* (M102)	65.3[b]	130[b]	
I100A	58.2	118	−2.5	I100M	60.8	125	−1.6
M102A/M106A	54.5	101	−3.7	WT* (M102/M106)	65.3	130	
V103A	60.9	122	−1.6	V103M	62.2[b]	117[b]	−1.2[b]
F104A	57.8	109	−2.7	F104M	64.5	121	−0.4
M106A	60.1	114	−1.9	WT* (M106)	65.3	130	
V111A	62.4	127	−1.0	V111M	63.3	127	−0.7
L118A	56.3	109	−3.2	L118M	63.5[b]	130[b]	−0.7[b]
L121A	59.1	108	−2.2	L121M	63.2[b]	129[b]	−0.8[b]
WT* (A129)	65.3[b]	130[b]	- -	A129M	60.1	109	−1.9
V149A	56.3	106	−3.2	V149M	57.5	111	−2.8
F153A	55.6	108	−3.4	F153M	63.7[c]	128[c]	−0.6[c]

[a] T_m is the melting temperature at pH 5.4 (see text) and has an uncertainty of about ±0.14 °C. ΔH is the enthalpy of unfolding at the T_m of the mutant. $\Delta\Delta G$ gives the difference between the free energy of unfolding of the mutant and that of WT*. Its uncertainty is about ±0.15 kcal/mol for the more stable variants and increases to about ±0.3 kcal/mol for the least stable ones. [b] From Gassner et al. (*34*). [c] From Eriksson et al. (*18*).

Table 3: Crystallographic Data Collection and Refinement Statistics

		cell dimensions						
mutant	resolution (Å)	*a*, *b* (Å)	*c* (Å)	completeness of data (%)	R_{merge} (%)	*R* (%)	Δ_{bond} (Å)	Δ_{angle} (deg)
I78A	2.10	61.1	97.7	83	8.8	16.8	0.015	2.1
I78M	2.20	60.5	97.7	90	7.2	15.7	0.021	2.6
L84M	1.85	61.1	96.9	92	4.1	15.5	0.016	2.4
V87M	2.12	61.0	97.2	85	5.9	15.2	0.016	2.6
L91A	2.10	61.3	96.5	94	4.0	16.5	0.021	2.7
L91M	2.05	61.0	97.1	89	6.7	15.8	0.018	2.5
I100M	1.89	60.9	97.1	93	4.2	16.0	0.017	2.6
V103M	2.05	61.0	97.3	93	5.9	15.3	0.016	2.4
F104M	2.12	60.9	97.2	93	4.9	16.1	0.020	2.6
M106A[a]	1.85	60.9	96.5	64	4.6	15.3	0.018	2.3
V111M	1.90	61.0	97.4	92	5.5	15.3	0.017	2.4
L118M[b]	1.80	61.1	97.1	81	4.1	15.1	0.015	2.3
L121M	1.80	61.0	96.9	93	3.6	16.6	0.020	2.2
V149M	1.90	60.9	97.3	92	6.1	15.7	0.017	2.5

[a] From Xu et al. (*25*). [b] Following Eriksson et al. (*22*) (see text).

Leucine 121 to Alanine and Methionine. The alanine and methionine replacements at site 121 have somewhat different consequences. The former is associated with fairly substantial rearrangements that diminish the size of the putative cavity (*23*). In contrast, the methionine replacement causes very little change in the structure (not shown).

Valine 149 to Alanine and Methionine. In wild-type lysozyme there is a small declivity with polar atoms lining the walls near Val149, which, in the Val149 → Ala mutant, is occupied by a water molecule (*25*). In the methionine mutant the C^{ϵ} methyl group extends toward this same space, allowing the large side chain to be accommodated with very modest structural changes (Figure 7).

DISCUSSION

The sites that are the subject of the present study are shown in Figure 1. Some are well within the protein; others are closer to the surface. When a side chain such as valine, leucine, isoleucine, methionine, or phenylalanine is replaced with alanine, the structural consequences differ, depending on the location within the protein and the rigidity with which the surrounding side chains are held (*22*, *25*). In the case of Ile78 → Ala, for example, surrounding atoms move inward by up to 1 Å (Figure 4b). In contrast, for Leu91 → Ala there are only very minor structural adjustments (not shown). Similarly, methionine replacements at different sites also

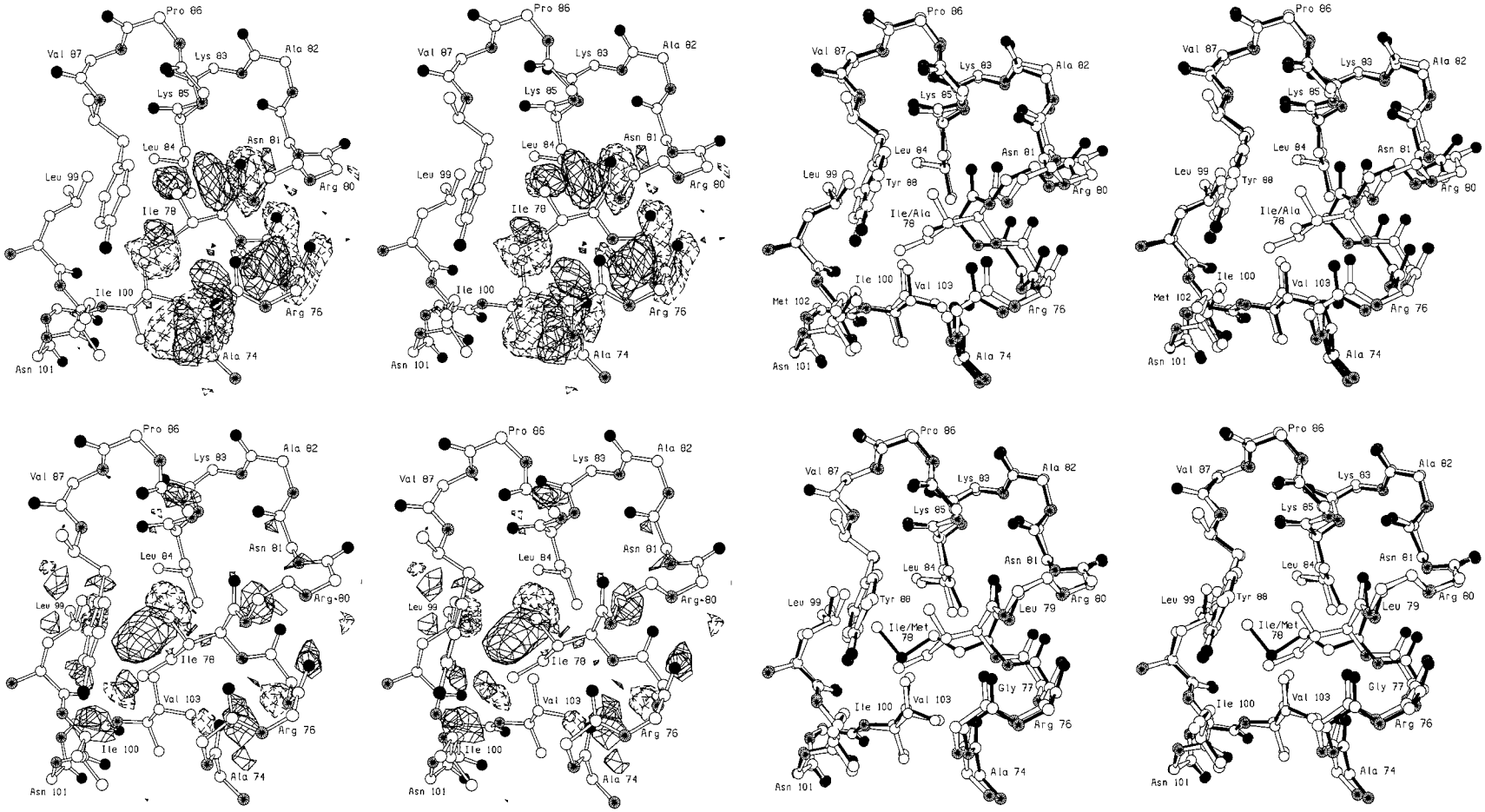

FIGURE 4: (a, upper left) Difference in electron density between mutant I78A and WT* lysozyme superimposed on the structure of WT*. Coefficients are ($F_{Mut} - F_{WT*}$), where F_{Mut} and F_{WT*} are the observed scattering amplitudes for the mutant and WT* crystals. Phases are from the refined structure of WT*. Electron density, positive solid and negative broken, is contoured at $\pm 3\sigma$. Resolution is as in Table 3. Carbon atoms are drawn as open circles, oxygen and sulfur solid, and nitrogen with spokes. (b, upper right). Superposition of mutant I78A (solid bonds) on WT* (open bonds). (c, lower left). Difference density map for mutant I78M; all conventions are as in Figure 2a. (d, lower right). Superposition of mutant I78M (solid bonds) on WT* (open bonds).

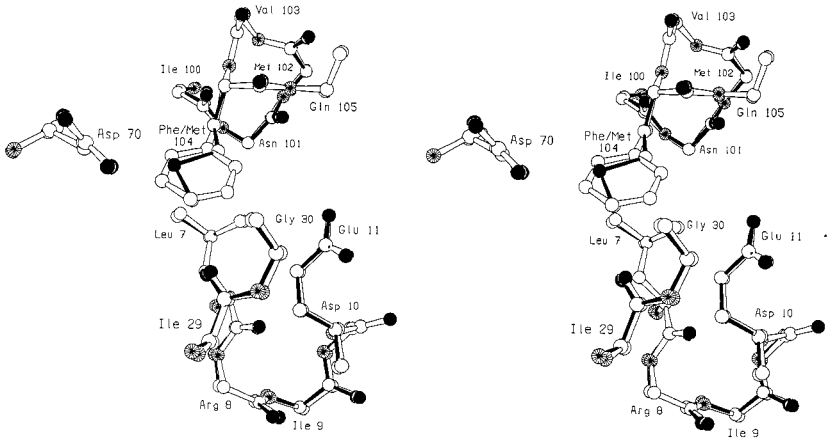

FIGURE 5: Superposition of mutant F104M (solid bonds) on WT* (open bonds).

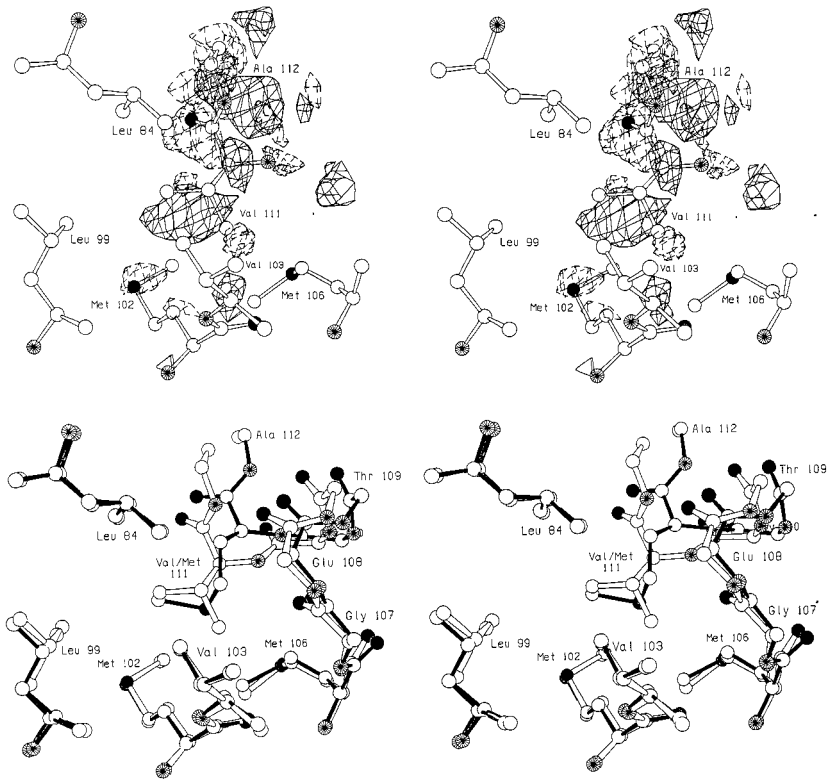

FIGURE 6: (a, top) Difference density map for mutant V111M; all conventions are as in Figure 2a. (b, bottom) Superposition of mutant V111M (solid bonds) on WT* (open bonds).

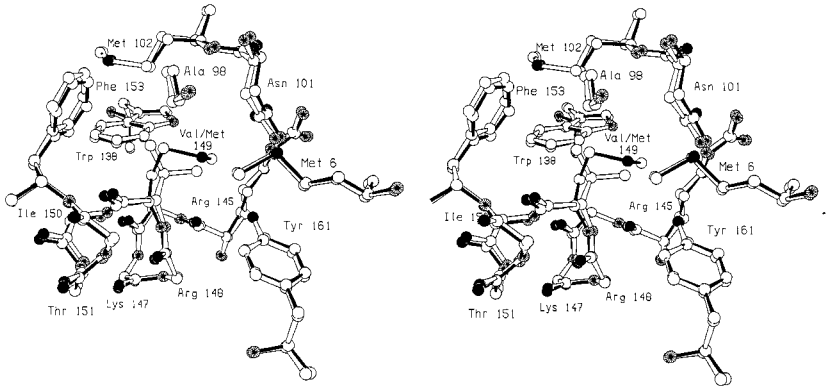

FIGURE 7: Superposition of mutant V149M (solid bonds) on WT* (open bonds).

elicit different responses. At site 78, there are outward shifts of up to 0.7 Å (Figure 4d), while at site 91 there are hardly any changes at all (not shown). Some of the sites are known to be protected from hydrogen exchange in the burst-phase intermediate (*4*). Mutations at such sites may stabilize the burst-phase intermediate through an increase in helical

propensity, better solvation of polar residues upon removal of a bulky residue, or enhanced freedom of rotation of side chains in the local environment.

Generally speaking, the substitution of alanine at a given site is more destabilizing than methionine (Table 2). This is not surprising in view of the reduced solvent transfer free energy of alanine relative to the larger hydrophobic side chains, coupled with the energetic cost of cavity creation (*22*, *25*). For sites that have a valine residue in the wild-type protein, the alanine and methionine replacements are much closer in stability (Table 2), consistent with valine being intermediate in size.

Within the time window that kinetic data were accessible (Figure 2) the refolding of all of the mutant proteins can be well modeled as the sum of two exponentials. (For mutants L99A, F153A, and M102A/M106A a single-exponential sufficed, the putative slower exponential not falling within the observational window.) About half of the circular dichroism signal (Figure 2a) develops within the dead time of the stopped-flow instrument (about 12 ms). This has been seen before for T4 lysozyme by pulsed hydrogen exchange methods (*4*) and is similar to the burst phase of protein folding seen in other systems. Thus, three well-separated kinetic regimes are apparent. First, there are processes that take place within the dead time of the system. Second, there is the main or principal exponential decay. Third, there is a low-amplitude, slow exponential decay at about $2-3\ s^{-1}$ due to proline isomerization (*4*).

We interpret the observed kinetics with the following simple model. The unfolded state, U, rapidly forms an intermediate, I, that accounts for the burst phase observed in the circular dichroism signal. We observe the first-order conversion of this intermediate into the folded state, F. Thus for wild type:

$$\mathrm{U} \xrightarrow[\text{fast}]{} \mathrm{I} \xrightarrow[19\ s^{-1}]{k_2} \mathrm{F}$$

In the discussion that follows, the much slower process, identified as proline isomerization, is neglected. Since all three proline residues in the native state have the trans conformation, the amplitude of the very slow step is low, corresponding to the population of cis proline isomers in the unfolded state (*4*).

Table 1 compares refolding rates for alanine and methionine substitutions at the same sites. For sites 153 and 99, alanine decreases the rate of refolding to give a ratio of 4 and 5 in rates, respectively. Refolding rates at sites 78, 111, and 129 vary little from that of WT* when substituted with alanine but slow markedly when substituted with methionine to give ratios of between 0.7 and 1.0. For most sites in this study the ratio of refolding rates when methionine is present to that when alanine is present is close to 1.8.

Inspection of the data suggests that the behavior of the different mutations can be most easily rationalized not by considering methionine substitutions versus alanine substitutions but by segregating the mutations according to the domain of the protein within which they are located.

In Figure 3a we plot, for all mutations that are located within the carboxy-terminal domain, the rate of folding relative to wild type as a function of the change in protein stability associated with the substitution. A similar plot for those mutations that are located within the amino-terminal domain is shown in Figure 3b. Most mutations within the carboxy-terminal domain that substantially destabilize the protein slow the rate of folding. A very different result is obtained in the amino-terminal domain (Figure 3b). Here, no mutation has a comparably large effect on the rate of folding, whether it destabilizes the protein or not. The different consequences for folding, according to whether the alanine and methionine substitutions occur in the N- or C-terminal domain, are illustrated in Figure 8.

This different behavior for mutations within the two domains suggests that these domains contribute very differently to the transition state. Using native-state hydrogen exchange, Llinás et al. (*30*) have shown that the N-terminal domain is less stable than the C-terminal domain. The correlation between the rate of folding and the overall stability of the protein, seen in Figure 3a for a large number of mutations throughout the C-terminal domain, suggests that at the rate-limiting step the structure of this domain is being formed and its structure must be similar to that of the wild-type protein.

For a mutation that introduces a cavity there is a correlation between the size of the cavity and the loss of stability (*22*, *25*). Therefore, for mutations within the C-terminal domain one would expect a correlation between cavity size and rate. That this is the case is shown in Figure 9. This may provide one way to rationalize the behavior of these mutants. A substitution that creates a cavity can be viewed to involve the removal of potential packing interactions at critical times in protein folding, thus leading to the decrease in folding rates. In this view the protein structure at the rate-limiting step in folding must have a structure that is quite close to that of the fully folded protein.

In an alternative view, the truncation of the wild-type side chain or the slightly increased helix propensity of an alanine residue could differentially stabilize the burst-phase intermediate. Of the three single mutations that slow folding to the greatest degree [i.e., L84A, L99A, and F153A (Figure 3a)], only one (L99A) is at a site that was identified by Lu and Dahlquist (*4*) as being part of the burst-phase intermediate. Thus changes to this early intermediate, per se, do not appear to explain the behavior of all of the mutations studied here. Evidently structure forms at early times near residue 99 and the rate-limiting step represents elaboration of that structure to form the essence of the C-terminal domain core. Residues 84 and 153 are close to residue 99 in the final folded structure. The amides of residues corresponding to helix E, which includes residue 99 and is largely buried within the C-terminal domain (Figure 1), are protected from hydrogen exchange in the burst phase (*4*). Thus it is very likely that this helix has formed in the intermediate.

There is an approximate linear free energy relationship between the observed folding rate and the equilibrium constant for folding for the residues involved in forming the hydrophobic core of the C-terminal domain. This is shown in Figure 3a as a straight line whose slope ($\Delta\Delta G^{\ddagger}/\Delta\Delta G$) is approximately 0.22. This slope has a similar interpretation to the ϕ value analysis suggested by Fersht for the analysis of strictly two-state folding kinetics (*31*, *32*). However, in our case we observe the folding rate constant from the intermediate rather than the unfolded state, so the interpretation of the slope is somewhat more complicated. The slope

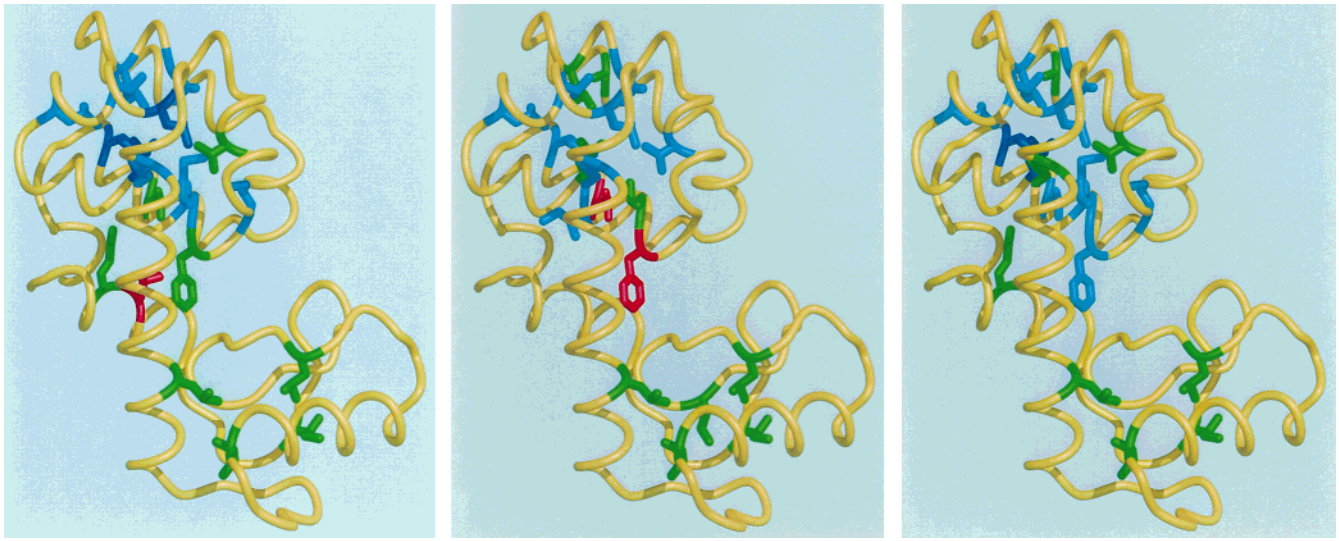

FIGURE 8: (a, left) Schematic illustration showing the effect of alanine substitutions on the rate of folding of T4 lysozyme. Amino acids whose substitution with alanine have little if any effect on rate ($-0.1 \leq \Delta\Delta G^{\ddagger} \leq 0.1$ in Figure 3a) are colored green; those that result in faster folding ($\Delta\Delta G^{\ddagger} > 0.1$) are colored red; those that result in somewhat slower folding ($-0.7 \leq \Delta\Delta G^{\ddagger} < -0.1$) are colored light blue; and the slowest folders ($\Delta\Delta G^{\ddagger} < -0.7$) are colored dark blue. The lysozyme backbone is drawn in the same alignment as in Figure 1. All substitutions that significantly accelerate or decelerate the rate of folding are in the upper, C-terminal, domain. (b, center). Schematic illustration showing the effect of methionine substitutions on the rate of folding. The coloring follows the same conventions as in panel a, i.e., blue for methionine substitutions with significantly slow folding, green for little if any effect, and red for faster folding. (c, right). Schematic illustration comparing the effects on the rate of refolding of T4 lysozyme of alanine substitutions at different sites relative to methionine substitutions at the same sites. Green indicates sites with k_{Met}/k_{Ala} less than unity (Table 1), light blue indicates sites with the ratio between 1.0 and 2.0, and dark blue indicates sites with the ratio greater than 2.0 (i.e., sites for which the alanine substitutions are slowest relative to their methionine counterparts).

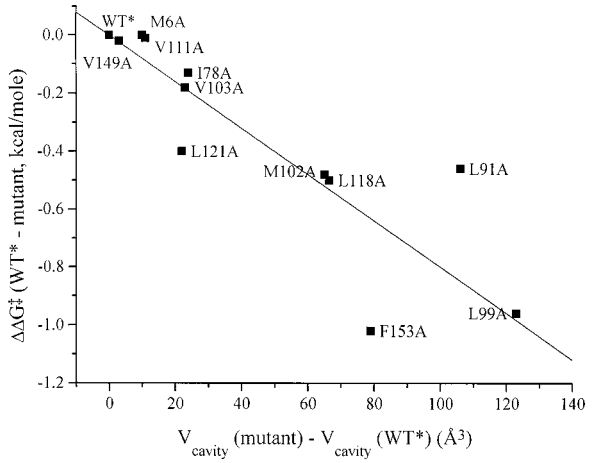

FIGURE 9: Relation between the rate of folding of mutant lysozymes and the increase in cavity volume created by the mutation. All mutations shown are within the C-terminal domain. Folding rates are relative to that of wild type and are plotted as in Figure 7a. Cavity volumes are from Eriksson et al. (*22*) and Xu et al. (*25*) except for alanine mutants I78A and L91A, which have volumes of 24 Å^3 and 106 Å^3.

of 0.2 shows that on average only about 20% of the equilibrium destabilization energy associated with mutational substitution is expressed in the folding kinetics. Under the conditions used in this study, the folding kinetics reflects the energy difference between the intermediate and the transition state. For mutations of residues in helix E, a detailed interpretation is difficult because that region has formed structure in the intermediate and the kinetics will reflect the difference in energy and thus structure between the intermediate and the transition state. Pulsed hydrogen exchange measurements show that most of the remaining residues in the C-terminal domain have not formed structure in the intermediate (*4*). For these residues the interpretation is more straightforward. Consider the mutations Leu84 → Ala and Phe153 → Ala (Figure 3a). In the context of the fully formed structure, each of these substitutions lowers the stability of the protein by about 3.5 kcal/mol as a result of a combination of reduced hydrophobic interactions and an increase in packing defects in the folded state (*22*, *25*). In the kinetic measurement, the same mutations perturb the difference between the transition state and the intermediate where the residue is unstructured by about 1 kcal/mol. This is about half the energetic change expected on the basis of hydrophobic considerations alone. This suggests that at the transition state, the hydrophobic core of the C-terminal domain is forming, but it is still incomplete and is probably partially solvent-accessible. In a structural sense, the core is developing from the residues that will be in contact in the final folded state but their interactions are still incomplete.

There are three mutants that fold somewhat faster than WT*, namely, L7A, F104M, and V149M (Figure 3a). Each of these lies within the carboxy-terminal domain (Figure 8). In the case of F104M and V149M the structures have been determined (Table 3). At site 104 the methionine occupies the space vacated by the phenylalanine with very little change in the surrounding atoms (Figure 5). In the case of V149M the methionine takes advantage of space nearby to replace the smaller valine side chain, but again with very little perturbation of the surrounding atoms (Figure 7). These structural correlations are consistent with the beneficial effect of both F104M and V149M on the rate of folding being due to stabilization of a transition state that is similar in structure to the folded protein. The observation that the mutation L7A increases the rate of folding is consistent with helix A (residues 3−10) being considered as part of the C-terminal domain (*33*). It also suggests that helix A, although remote

in sequence from the C-terminal domain, nevertheless contributes to the structure of the transition state.

ACKNOWLEDGMENT

We thank Yash Paul Myer for advice on how to collect rapid kinetic data by means of circular dichroism, James E. Swirczynski for advice on analogue electronic circuits, especially multipole filters, Jack Landis and Julie L. Sohl for advice on J-720 and Bio-Logic software, David Senkovitch for advice on reduced-load mounting of the SFM/3 stopped-flow head, Larry Weaver for advice on X-ray data collection and analysis, Leslie Gay and Hong Xiao for protein preparation and crystal growth, Enoch P. Baldwin and Jian Xu for the structure of M106A, and Hong Qian, Michael Schimerlik, Eric Anderson, and John A. Schellman for helpful discussions, especially in the areas of protein folding and modulation spectroscopy.

REFERENCES

1. Matouschek, A., Kellis J. T., Jr., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature 346*, 440−445.
2. Chen, B.-L., Baase, W. A., Nicholson, H., and Schellman, J. A. (1992) *Biochemistry 31*, 1464−1476.
3. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) *J. Mol. Biol. 224*, 771−782.
4. Lu, J., and Dahlquist, F. W. (1992) *Biochemistry 31*, 4749−4756.
5. Serrano, L. (1994) *Curr. Opin. Struct. Biol. 4*, 107−111.
6. Kalnin, N. N., and Kuwajima, K. (1995) *Proteins: Struct., Funct., Genet. 23*, 163−176.
7. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997) *Nat. Struct. Biol. 4*, 305−310.
8. Munson, M., Anderson, K. S., and Regan, L. (1997) *Folding & Des. 2*, 77−87.
9. Xu, Y., Mayne, L., and Englander, S. W. (1998) *Nat. Struct. Biol. 9*, 774−778.
10. Dill, K. A., and Chan, H. S. (1997) *Nat. Struct. Biol. 1*, 9−10.
11. Matthews, B. W. (1993) *Annu. Rev. Biochem. 62*, 139−160.
12. Blaber, M., Baase, W. A., Gassner, N., and Matthews, B. W. (1995) *J. Mol. Biol. 246*, 317−330.
13. Matsumura, M., and Matthews, B. W. (1989) *Science 243*, 792−794.
14. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.
15. Alber, T., and Matthews, B. W. (1987) *Methods Enzymol. 154*, 511−533.
16. Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., and Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 44−73.
17. Poteete, A. R., Dao-pin, S., Nicholson, H., and Matthews, B. W. (1991) *Biochemistry 30*, 1425−1432.
18. Eriksson, A. E., Baase, W. A., and Matthews, B. W. (1993) *J. Mol. Biol. 229*, 747−769.
19. Elwell, M., and Schellman, J. (1975) *Biochim. Biophys. Acta 386*, 309−323.
20. Jensen, H. P., Schellman, J. A., and Troxell, T. (1978) *Appl. Spectrosc. 32*, 192−200.
21. Qian, H., and Elson, E. l. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5479−5483.
22. Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) *Science 255*, 178−183.
23. Baldwin, E., Xu, J., Hajiseyedjavadi, O., Baase, W. A., and Matthews, B. W. (1996) *J. Mol. Biol. 259*, 542−559.
24. Baldwin, E., Baase, W. A., Zhang, X.-J., Feher, V., and Matthews, B. W. (1998) *J. Mol. Biol. 277*, 467−485.
25. Xu, J., Baase, W. A., Baldwin, E., and Matthews, B. W. (1998) *Protein Sci. 7*, 158−177.
26. Hamlin, R. (1985) *Methods Enzymol. 114*, 416−452.
27. Otwinowski, Z. (1993) in *Proceedings of the CCP4 Study Weekend* (Sawyer, L., Isaac, N., Borley, S., Eds.) pp 56−62, SERC Daresbury, Laboratory, Warrington, U.K.
28. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) An efficient general-purpose least-squares refinement program for macromolecular structures, *Acta Crystallogr. A43*, 489−503.
29. Tronrud, D. E. (1996) *J. Appl. Crystallogr. 29*, 100−104.
30. Llinás, M., Gillespie, B., Dahlquist, F. W., and Marqusee, S. (1999) *Nat. Struct. Biol.* (in press).
31. Serrano, L., Kellis, J. T., Jr., Cann, P., Matouschek, A., and Fersht, A. R. (1992) *J. Mol. Biol. 224*, 783−804.
32. Fersht, A. R., Itzhaki, L. S., ElMasry, N. F., Matthews, J. M., and Otzen, D. E. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10426−10429.
33. Zhang, X.-J., Wozniak, J. A., and Matthews, B. W. (1995) *J. Mol. Biol. 250*, 527−559.
34. Gassner, N. C., Baase, W. A., and Matthews, B. W. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 12155−12158.

BI9915519