ELSEVIER

Biophysical Chemistry 101–102 (2002) 43–56

Biophysical Chemistry

www.elsevier.com/locate/bpc

A helix initiation signal in T4 lysozyme identified by polyalanine mutagenesis

Xue-jun Zhang[1], Walter A. Baase, Brian W. Matthews*

Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, OR 97403-1229, USA

Abstract

To better understand the relation between sequence and structure, and in an attempt to simplify the protein folding problem, a series of alanine substitutions was introduced into bacteriophage T4 lysozyme. In contrast to previous studies in this system, which were restricted to single α-helices, the present analysis included a helix–turn–helix region, a loop–helix region, and two α-helices that were well separated in the three-dimensional structure. It was shown previously that T4 lysozyme is very tolerant of alanine substitutions within α-helices, especially at solvent-exposed sites. The present study shows that the protein is also tolerant of such substitutions in turn and loop regions, although less than in helices. The results confirm that the structural information in the amino acid sequence is highly redundant. For example, the protein with the sequence ^{127}AAAAAALAAAAWAAA^{141} folds normally, has melting temperature only 0.8 °C lower than wildtype, and has a crystal structure that is also very similar to wildtype. Polyalanine substitutions within turns or loops can, however, lead to differences in structure and in folding. In one example the triple substitution K35A/S36A/P37A caused this region of the molecule to change to a more helical conformation. In a second case the mutant with the sequence ^{34}AAAAALAAAKAALAAA^{49}, which spans a loop–helix region, had a dramatically altered thermal unfolding transition, suggesting that this region may tend to form a single, uninterrupted, helix. Substitution of Ala38 in the above construct with aspartic acid caused the unfolding to be more like wildtype, suggesting that residue 38, which is at a helix-capping position in the wildtype structure, provides an initiation signal that is essential in the polyalanine mutant for the correct formation of α-helix 39–50. In a typical protein, the information that codes for the 3D structure is presumably distributed over many amino acids. The present results suggest that in simplified sequences the key folding information may be restricted to a subset of critical residues, and so be more readily accessible to experimental analysis.

Keywords: Phage T4; Alanine scanning mutagenesis; Lysozyme; Protein stability; α-Helix

*Corresponding author. Tel.: +1 541 346 2572; fax: +1 541 346 5870.
E-mail address: brian@uoxray.uoregon.edu (B.W. Matthews).
[1] Present address: Crystallography Program, Oklahoma Medical Research Foundation, 825 NE 13th Street, Oklahoma City, OK 73104, USA.

0301-4622/02/$ - see front matter
PII: S0301-4622(02)00193-X

1. Introduction

1.1. Dedication

It is a special pleasure to dedicate this article to our good friend, colleague and mentor, John Schellman, and, at the same time, to acknowledge the contributions and friendship of his wife Charlotte. John and Charlotte have both been involved in, and have made critical contributions to, what has become known as 'the lysozyme project'.

Prior to our determination of the crystal structure of bacteriophage T4 lysozyme [1], we invited colleagues to predict the locations of the elements of secondary structure. Among the 11 helix predictions that were received, Charlotte Schellman's was the most accurate (essentially co-equal with that of Chou and Fasman [2]) [3]. More recently we used T4 lysozyme to investigate the role of the 'Schellman motif', named in Charlotte's honor [4,5].

Our studies of the structures of mutant T4 lysozymes have been intimately coupled with analyses of effects of the mutations on stability. The experimental procedures to permit the stability measurements, together with the theoretical underpinning, all originate directly from John Schellman's laboratory [6–13]. Without these contributions the 'lysozyme project' would not have been possible.

1.2. The problem of protein folding

The question as to how the amino acid sequence of a protein determines its three-dimensional structure has been attacked by a variety of experimental and theoretical methods [14–18]. One aspect of the problem is the information that is encoded in the sequence relative to the information that is necessary to specify the structure. If we make the admittedly gross oversimplification that the backbone conformation of each residue can be specified by a binary choice, i.e. 'helical' or 'extended', then the number of choices required to specify the fold of a polypeptide chain of n residues is on the order of 2^n. This corresponds to an information content of $-n\sum_{i=1}^{2} p_i \ln p_i \simeq n\ln 2$ nats [19], where p_i is the probability of the ith conformation. The fact that there are more than two possible conformations per residue will increase this estimate. The effects of excluded volume, however, will reduce it. Bearing in mind that any one of the 20 amino acids can occur at any position in a polypeptide chain, the chain can be estimated as having on the order of 20^n bits of information, or an information content of $-n\sum_{i=1}^{20} p_i \ln p_i \simeq n\ln 20$ nats. The information in the sequence therefore greatly exceeds that necessary to specify the structure. Indeed, since $\ln 2/\ln 20 = 0.23$, it can be estimated that, in principle, only about a quarter of the amino acids in a given sequence are required to specify the 3D structure. In practice, it would seem unlikely that nature would construct such a protein since it would be very vulnerable to mutation. Rather, the structural information is likely to be shared by different amino acids to a different extent. Having redundant information would permit some mutation while maintaining the 3D structure. Experimentally, however, one may be able to create such proteins, thereby providing a simplified model that could provide new insights into the problem of protein folding.

One way to experimentally identify the residues of high information content is by substitution with alanine [20–25]. Because of the redundancy of the structural information encoded in the amino acid sequence, the site-by-site substitution technique may not be sufficient to identify the minimum set of critical residues (if such a set exists). Therefore, a strategy of combining multiple alanine substitutions has been proposed to systematically eliminate less-important interactions and to thereby simplify the protein folding problem [22].

Previously, alanine substitution in T4 lysozyme has been restricted to helical regions [22,26–28]. It could be argued that such substitutions are especially favorable since alanine is a residue of high helical propensity (e.g. see [29] and references therein). Indeed, alanine substitution at appropriate sites within helices can increase the stability of the protein [22,28,29]. In the present report we describe a series of polyalanine substitutions in both helical and non-helical regions of

T4 lysozyme. These studies identify a helix initiation (or termination) signal that, in a certain context, seems to be critical for correct folding.

2. Materials and methods

2.1. Protein purification

Mutants were generated via the method of Kunkel et al. [30] either in wildtype lysozyme (WT) or in the cysteine-free pseudo-wildtype protein (WT*) which includes the mutations C54T and C97A [31].

Proteins were purified as described by Poteete et al. [32]. The yield of some of the less stable proteins was substantially reduced relative to wildtype.

2.2. Crystallography

Crystallization was first attempted using ~2 M phosphate, pH ~6.7–7.0, as for wildtype [33]. As necessary, a 'fast screen' of alternative conditions, based on Jancarik and Kim [34] was also used. X-ray data were collected using a multiwire detector [35,36] or oscillation photography [37,38]. Structures of the mutants that crystallized isomorphously with wildtype were determined and refined [39] by standard procedures (e.g. [40]). Mutants that crystallized in novel space groups [41] (Table 3) were determined by molecular replacement [42,43]. For crystal forms with multiple copies of the molecule in the asymmetric unit, but limited resolution, averaging of the electron density maps was used to improve their quality.

2.3. Thermal unfolding

Circular dichroism monitored thermally promoted unfolding and data analysis were as described [33]. Protein concentrations were between 5 and 30 μg/ml and reversibility upon cooling was in excess of 90%. Where appropriate (Table 2), free energy changes were calculated using a two-state model at an isotherm of 50 °C using a constant ΔC_p of 1.8 kcal/mol deg [33].

3. Results

The regions of secondary structure of T4 lysozyme that are the subject of the present analysis, and the mutants that have been made, are shown in Table 1 and Fig. 1. The locations within the molecule as a whole are shown in Fig. 2.

3.1. Analysis of a helix–turn–helix

Previously it was shown that the combination of the 4 substitutions D127A, E128A, V131A and N132A within the α-helix that includes residues 126–134 increased the melting temperature of the protein by 4.0 °C at pH 2.02 [26]. The sequence of the resultant mutant within the α-helical region is ^{126}WAAAAAALA^{134}. It can also be identified as A127–134(L^{133}) (Fig. 1a), indicating that there are alanines from residue 127 to 134 except at site 133 where leucine is retained as in the wildtype protein.

One part of the present study was to extend the alanine substitutions beyond α-helix 126–134 up to residue 141. Overall, the segment investigated includes the initial helix, 126–134, a short turn (residues 135–136) and the following short helix, 137–141 (Figs. 1 and 2). Two triple-alanine substitutions K135A/S136A/R137A and Y139A/N140A/Q141A were added in two steps. The two-step procedure is convenient for the oligonucleotide-directed mutagenesis and also allows the two segments to be evaluated independently. These additional alanine substitutions were chosen to be at the C-terminus rather than the N-terminus of helix 126–134 because more of the sites were solvent-exposed and the replacements therefore less likely to perturb the protein core. Trp138, which has 93% of its sidechain buried, was left unchanged. It has previously been shown that the substitutions Trp138→Ala and Trp138→Tyr destabilize T4 lysozyme by 3.2 and 3.0 kcal/mol, respectively (Xu et al., unpublished; [10]).

With the exception of Trp138, as well as the buried residue Leu133, alanine substitutions within the helix–turn–helix region 127–141 do not, in general, destabilize the protein significantly. For example, the melting temperatures of the mutants

Table 1
Identification of mutant lysozymes

Name of mutant	Lab code	PDB code	Substitutions included in mutant
A128–132	37AAA	1L36	E128A/V131A/N132A
A127–132	4006A	1L73	D127A/E128A/V131A/N132A
A128–134	4007A	1L74	E128A/V131A/N132A/L133A
A127–134	5003A	1L75	D127A/E128A/V131A/N132A/L133A
A128–137(L^{133})	6004A	168L	E128A/V131A/N132A/K135A/S136A/R137A
A127–137(L^{133})	7002A	–	D127A/E128A/V131A/N132A/K135A/S136A
A127–141($L^{133}W^{138}$)	10A01	–	D127A/E128A/V131A/N132A/K135A/S136A/R137A/Y139A/N140A/Q141A
A128–141($L^{133}W^{138}$)	9001A	169L	E128A/V131A/N132A/K135A/S136A/R137A/Y139A/N140A/Q141A
A40–49	A40–49	1L64	N40A/K43A/S44A/E45A/L46A/D47A/K48A/WT*
A40–49($K^{43}L^{46}$)	–	–	N40A/S44A/E45A/D47A/K48A/WT*
A40–49($K^{43}L^{46}$)/A127–132	I001A	192L	N40A/S44A/E45A/D47A/K48A/D127A/E128A/V131A/N132A/WT*
A34–37	4008A	151L	T34A/K35A/S36A/P37A
A34–49($L^{39}K^{43}L^{46}$)	J001A	–	T34A/K35A/S36A/P37A/S38A/N40A/S44A/E45A/D47A/K48A/WT*
S38A	S38A	–	S38A
S38D	S38D	1L19	S38D
A34–49($D^{38}L^{39}K^{43}L^{46}$)	J002A	174L	T34A/K35A/S36A/P37A/S38D/N40A/S44A/E45A/D47A/K48A/WT*
A34–49($D^{35}L^{39}K^{43}L^{46}$)	J003A	–	T34A/K35D/S36A/P37A/S38A/N40A/S44A/E45A/D47A/K48A/WT*
A34–49($D^{36}L^{39}K^{43}L^{46}$)	J004A	–	T34A/K35A/S36D/P38A/S38A/N40A/S44A/E45A/D47A/K48A/WT*
A34–49($D^{37}L^{39}K^{43}L^{46}$)	J005A	–	T34A/K35A/S36A/P37D/S38A/N40A/S44A/E45A/D47A/K48A/WT*
A34–49($L^{39}D^{40}K^{43}L^{46}$)	J006A	–	T34A/K35A/S36A/P37A/S38A/N40D/S44A/E45A/D47A/K48A/WT*
A53–57($C^{54}G^{56}$)	33AAA	190L	N53A/N55A/V57A
A53–58	6005A	–	N53A/C54A/N55A/G56A/V57A/I58A
A52–58	7004A	–	R52A/N53A/C54A/N55A/G56A/V57A/I58A
A53–57($C^{54}G^{56}$)/A128–132	6003A	191L	N53A/N55A/V57A/E128A/V131A/N132A
A53–62($T^{59}K^{60}$)	7005A	–	N53A/C54A/N55A/G56A/V57A/I58A/E62A
A34–37/A128–132	7003A	–	T34A/K35A/S36A/P37A/E128A/V131A/N132A

Variants that are designated WT* include the substitutions C54T and C97A.

A127–141($L^{133}W^{138}$) and A128–141($L^{133}W^{138}$) (^{127}AAAAAALAAAAWAAA^{141} and ^{128}AAAAALAAAAWAAA^{141}) which have 10 and 9 alanine substitutions, respectively, are only 0.9 °C lower than that of wildtype lysozyme at pH 3.0 (Table 2).

The triple substitution E128A/V131A/N132A within helix 126–134 (mutant A128–132) increases the melting temperature by 2.2 °C at pH 3.0. Addition of the 3 substitutions K135A/S136A/R137A, the first two of which are in the turn region (Fig. 1a, Fig. 2), to give mutant A128–137(L^{133}), i.e. ^{128}AAAAALAAAA^{137}, reduces stability by 3.2 °C at pH 3.0 (i.e. by approximately 1.4 kcal/mol). This suggests that alanine substitutions in turn regions are tolerated, although perhaps less well than in α-helices.

The structures of two of the mutants with multiple alanine substitutions in the 127–141 region were determined, namely A128–137(L^{133}) and A128–141($L^{133}W^{138}$).

The first of these, with sequence ^{128}AAAAALAAAA^{137}, has 6 substitutions in the helix–turn region of the helix–turn–helix. It crystallized in space group $P2_12_12$ with five lysozyme molecules per asymmetric unit in a pentameric arrangement [41]. If the backbone atoms of the carboxy-terminal domain of the protein (residues 80–160) are superimposed on wildtype the root-mean-square discrepancy is 0.45 Å. For the region of the substitution (residues 126–137) the discrepancy is 0.65 Å. Thus, the overall backbone structure is quite well conserved in the mutant.

The second mutant whose structure was determined has 9 alanine substitutions extending throughout the entire helix–turn–helix region, to give the sequence ^{128}AAAAALAAAAWAAA^{141}. It crystallized in space group $P2_1$, but again with

(a) Protein Sequence

	126			130					135						140		
WT	W	D	E	A	A	V	N	L	A	K	S	R	W	Y	N	Q	T
Solv. access.	3	1	0	8	0	1	8	3	0	4	8	1	8	0	4	9	7	2
Location	α	α	α	α	α	α	α	α	α	s	s	α	α	α	α	α	α	s
Mutant																		
A128-132			**A**	A	A	**A**	**A**	L	A									
A127-132		**A**	**A**	A	A	**A**	**A**	L	A									
A128-134			**A**	A	A	**A**	**A**	**A**	A									
A127-134		**A**	**A**	A	A	**A**	**A**	**A**	A									
A128-137(L^{133})			**A**	A	A	**A**	**A**	L	A	**A**	**A**	**A**						
A127-137(L^{133})		**A**	**A**	A	A	**A**	**A**	L	A	**A**	**A**	**A**						
A127-141($L^{133}W^{138}$)		**A**	**A**	A	A	**A**	**A**	L	A	**A**	**A**	**A**	W	**A**	**A**	**A**		
A128-141($L^{133}W^{138}$)			**A**	A	A	**A**	**A**	L	A	**A**	**A**	**A**	W	**A**	**A**	**A**		

(b) Protein Sequence

	34					40					45					50					55		58
WT	T	K	S	P	S	L	N	A	A	K	S	E	L	D	K	A	I	G	R	N	C	N	G	V	I
Solv. access.	2	8	6	1	0	6	5	8	3	0	2	6	4	0	2	6	2	1	9	5	1	0	0	6	0
Location	β	s	s	s	s	α	α	α	α	α	α	α	α	α	α	α	α	s	s	s	s	s	s	s	s
Mutant																									
A40-49								**A**	A	A	**A**	**A**	**A**	**A**	**A**	**A**	A								
A40-49($K^{43}L^{46}$)								**A**	A	A	K	**A**	**A**	L	**A**	**A**	A								
A40-49($K^{43}L^{46}$)/A127-132																									
A34-37	**A**	**A**	**A**	**A**																					
A34-49($L^{39}K^{43}L^{46}$)	**A**	**A**	**A**	**A**	**A**	L	**A**	A	A	K	**A**	**A**	L	**A**	**A**	A									
S38D					**D**																				
A34-49($D^{38}L^{39}K^{43}L^{46}$)	**A**	**A**	**A**	**A**	**D**	L	**A**	A	A	K	**A**	**A**	L	**A**	**A**	A									
A34-49($D^{35}L^{39}K^{43}L^{46}$)	**A**	**D**	**A**	**A**	**A**	L	**A**	A	A	K	**A**	**A**	L	**A**	**A**	A									
A34-49($D^{36}L^{39}K^{43}L^{46}$)	**A**	**A**	**D**	**A**	**A**	L	**A**	A	A	K	**A**	**A**	L	**A**	**A**	A									
A34-49($D^{37}L^{39}K^{43}L^{46}$)	**A**	**A**	**A**	**D**	**A**	L	**A**	A	A	K	**A**	**A**	L	**A**	**A**	A									
A34-49($D^{40}L^{39}K^{43}L^{46}$)	**A**	**A**	**A**	**A**	**A**	L	**D**	A	A	K	**A**	**A**	L	**A**	**A**	A									
A53-57($C^{54}G^{56}$)																				**A**	C	**A**	G	**A**	
A53-58																				**A**	**A**	**A**	**A**	**A**	**A**
A52-58																			**A**	**A**	**A**	**A**	**A**	**A**	**A**
A53-57($C^{54}G^{56}$)/A128-132																									
A34-37/A128-132																									

Fig. 1. Identifications of polyalanine mutants. The name A127–141($L^{133}W^{138}$), for example, means that residues 127–141 are all alanine except that residue 133 is leucine and residue 138 is tryptophan. See Table 1 for additional information. Amino acids that are altered relative to wildtype are indicated in boldface. The 'solvent accessibility' is the fraction of the sidechain accessible to solvent where 10 corresponds to 100% solvent-accessible and 0 is fully buried. The location of each residue is indicated as either α ('α-helix'), s ('strand'), or β ('β-sheet'). The reference protein in which each mutant was constructed is given in Table 2. (a) Mutations in the vicinity of residues 126–147. (b) Mutations in the vicinity of residues 34–58.

five crystallographically independent molecules in a pentameric arrangement [41]. In both of the pentameric arrangements none of the alanine substitution sites are involved in the protein–protein contacts within the ring, suggesting that the polyalanine substitution does not, of itself, favor the formation of the pentamer. In the wildtype crystal structure, however, residues Lys135–Ser136–Arg137 participate in a crystal contact with Lys48–Ala49–Ile50–Gly51 of a neighboring molecule [44]. Presumably, it is the elimination of these contact residues that leads to alternative crystal forms, although it is not obvious why the pentameric association occurs in these two cases.

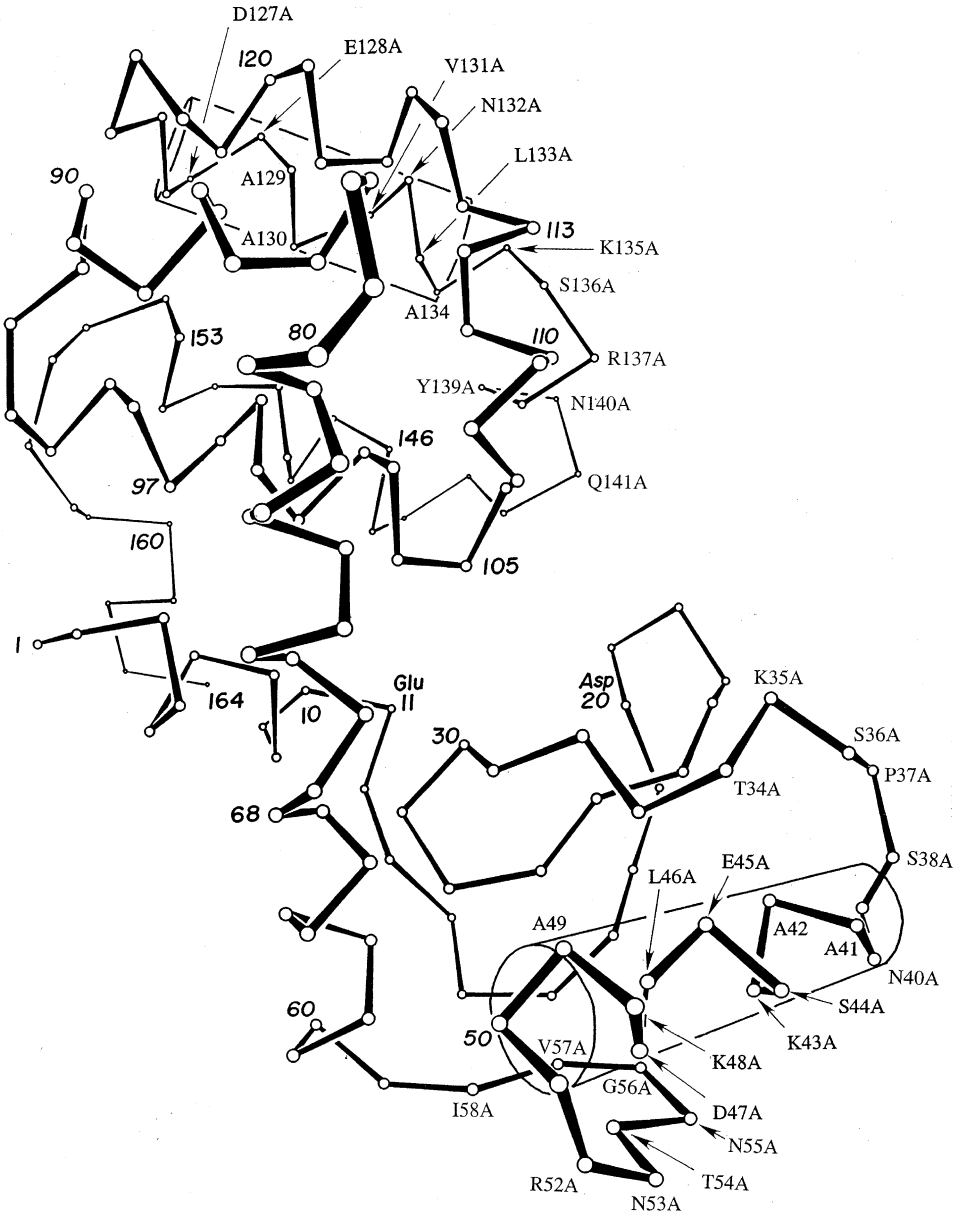

Fig. 2. Backbone of T4 lysozyme showing the residues that were replaced with alanine.

Based on superposition of the carboxy-terminal domain of the refined structure, the backbone atoms in the vicinity of the substitutions (residues 126–141) differ from wildtype by 0.73 Å (root-mean-square). Thus, even the substitution of nine residues with alanine does not change the local structure to a very great degree. There is a rather large change in the hinge-bending angle between the amino- and carboxy-terminal domains but this is observed in a number of mutant lysozymes [41] and is not attributed to the polyalanine substitution, per se.

Table 2
Stabilities of mutant lysozymes

Mutant lysozyme	Reference protein	ΔT_m (°C)	ΔH (kcal/mol)	$\Delta\Delta G$ (kcal/mol)
A128–132	WT	2.2	137	0.9
A127–132	WT	2.4	129	0.85
A128–134	WT	−7	102	−2.5
A127–134	WT	−7	103	−2.5
A128–137(L^{133})	WT	−1.0	107	−0.5
A127–137(L^{133})	WT	−0.9	102	−0.5
A127–141($L^{133}W^{138}$)	WT	−0.8	103	−0.5
A128–141($L^{133}W^{138}$)	WT	−0.9	102	−0.5
A40–49	WT*	−8.5[a]	70[a]	−2.1[a]
A40–49($K^{43}L^{46}$)	WT*	3.1[a]	117[a]	1.1[a]
A40–49($K^{43}L^{46}$)/A127–132	WT*	5.2	117	1.75
A34–37	WT	−1.6	109	−0.7
A34–49($L^{39}K^{43}L^{46}$)	WT*	−10	–	–
S38A	WT	−1.95	118	−0.77
S38D	WT	0.5[b]	123[b]	0.2[b]
A34–49($D^{38}L^{39}K^{43}L^{46}$)	WT*	−3.8	–	–
A34–49($D^{35}L^{39}K^{43}L^{46}$)	WT*	−9.2	–	–
A34–49($D^{36}L^{39}K^{43}L^{46}$)	WT*	−8.5	–	–
A34–49($D^{37}L^{39}K^{43}L^{46}$)	WT*	−7.4	–	–
A34–49($L^{39}D^{40}K^{43}L^{46}$)	WT*	−9.7	–	–
A53–57($C^{54}G^{56}$)	WT	−4.0	104	−1.5
A53–58	WT	−15.8	43	~ −4
A52–58	WT	−16.4	40	~ −4
A53–57($C^{54}G^{56}$)/A128–132	WT	−2.0	105	−0.8
A53–62($T^{59}K^{60}$)	WT	−16.4	42	~ −4
A34–37/A128–132	WT	0.5	97	−0.2

ΔT_m is the melting temperature change relative to the reference protein. ΔH is the enthalpy of unfolding measured at T_m. $\Delta\Delta G$ is the difference between the free energies of unfolding of the mutant and reference proteins. In the solvent used here, which was 25.0 mM KCl, 20.0 mM KPO_4, pH 3.0, the T_m of WT is 53.4 °C with a ΔH of 127 kcal/mol and the T_m of WT* is 51.7 °C with a ΔH of 114 kcal/mol. Errors in $\Delta\Delta G$ for these data are expected to not exceed $\pm$0.2 kcal/mol. Those proteins for which ΔH and $\Delta\Delta G$ are not given did not unfold in a two-state manner and there was clear persistence of dichroism beyond the initial transition. For these data a 'T_m' was estimated from the midpoint of the first, relatively high enthalpy transition by means of overlaying the thermal unfolding curves. No attempt was made to fit this single wavelength unfolding data to more complex thermodynamic models. Errors in 'T_m' determined in this fashion are on the order of $\pm$0.3 °C.

[a] From Heinz et al. [27].

[b] From Nicholson et al. [50].

3.2. Analysis of a loop–helix–loop

In a second phase of the present study, the prior analysis of α-helix 39–50 [27] was extended to include both the preceding and the following loops (residues 34–38 and 51–57) (Figs. 1 and 2).

First, the substitutions K35A, S36A and P37A were introduced in the loop that precedes helix 39–50. At the same time, Thr34, the last residue of the preceding β-sheet was also changed to alanine to give the mutant A34–37 (Fig. 1). This variant had limited effect on stability and on the nature of the unfolding transition (Table 2). (Substitutions that include the fourth turn residue, Ser38, behave in a very different manner and will be discussed separately.)

Mutant A34–37 was crystallized under conditions similar to wildtype but in space group R3 rather than $P3_221$ [41] (Table 3). The protein molecules associate to form a tightly packed trimer

Table 3
X-ray data collection and refinement statistics

Mutant	I001A	33AAA	6003A	6004A	9001A	J002A	4008A
Space group	$P3_221$	$P3_221$	$P3_221$	$P2_12_12$	$P2_1$	$P6_3$	R3
Cell dimensions							
a (Å)	61.1	60.6	61.0	157.2	40.4	89.7	100.5
b (Å)	61.1	60.6	61.0	177.9	112.3	89.3	100.5
c (Å)	95.1	95.8	97.2	40.5	135.2	87.1	40.8
β (°)					91.7		
Data collection							
Resolution (Å)	1.9	2.0	1.95	2.9	3.0	2.3	2.2
Molecules per asymmetric unit	1	1	1	5	5	2	1
Data completeness (%)	66	88	86	82	56	98	86
R_{merge} (%)	7.6	6.4	5.0	9.9	10.6	3.0	
Refinement							
R (%)	16.9	17.1	17.1	19.9	16.1	19.8	18.7
$\Delta_{bond\ length}$ (Å)	0.016	0.017	0.016	0.017	0.015	0.015	0.016
$\Delta_{bond\ angle}$ (°)	2.2	2.3	2.3	2.7	2.4	2.7	2.7
PDB code	192L	190L	191L	168L	169L	174L	151L

Mutants are listed by their laboratory code numbers. The full identifications are given in Table 1. Data for I001A were collected by oscillation photography and the remainder by use of an area detector.

with the substituted alanines at positions 34, 35 and 37 participating in the trimer interface. During refinement, difference electron density maps clearly indicated a substantial change in the mutated region (Fig. 3a). If the rigid parts of the amino-terminal domain plus the amino-terminal α-helix (i.e. residues 3–11, 15–18, 25–30 and 40–60) are superimposed on the wildtype structure, the backbone within the mutated region (residues 33–38) differ by 3.3 Å (Fig. 3b). In wildtype lysozyme residues 33–38 are in a somewhat extended (i.e. non-compact) loop with no two consecutive residues being in the 'α' region of the Ramachandran diagram (Table 4). In the mutant, however, Ala34 and Ala35 adopt the α-conformation, and Ala37 and Ser38 are roughly in this conformation as well (Table 4). Thus, residues 33–36 and 36–39 form two, consecutive, 'αα-type' β-turns [45].

It could be asked whether this change to a more 'α-like' conformation is due to the substitution of the 4 alanines or might, rather, be due to the replacement of Pro37. This would seem not to be the case in that the structure of the single mutant Pro37→Ala [46] has been determined and found to be virtually identical with wildtype (root-mean-square discrepancy of 0.09 Å for the backbone atoms of residues 15–60). Since it is just a matter of deleting the appropriate sidechains, it is trivial to construct a model with the mutant sequence and the wildtype conformation. The inverse model-building experiment, however, in which the wildtype sidechains are built into the crystal structure of the mutant, results in steric conflicts which suggest that the wildtype sequence cannot adopt the conformation seen in the mutant. In the mutant structure, Ala36 is completely buried, suggesting that the substitution S36A may be required for the conformational change. Also the substitution T34A would tend to create a cavity, presumably destabilizing the native structure.

Substitution of some of the residues in the loop that follows helix 39–50 (e.g. N53A/N55A/V57A to give mutant A53–57($C^{54}G^{56}$)), have limited effect on stability. If, however, 7 consecutive alanines are substituted in this region the resultant mutant A52–58 is substantially destabilized (Table 2). This is presumably because Arg52 and Glu62 form a salt bridge that is estimated to contribute approximately 2 kcal/mol to the stability of the folded protein at neutral pH [47]. Also

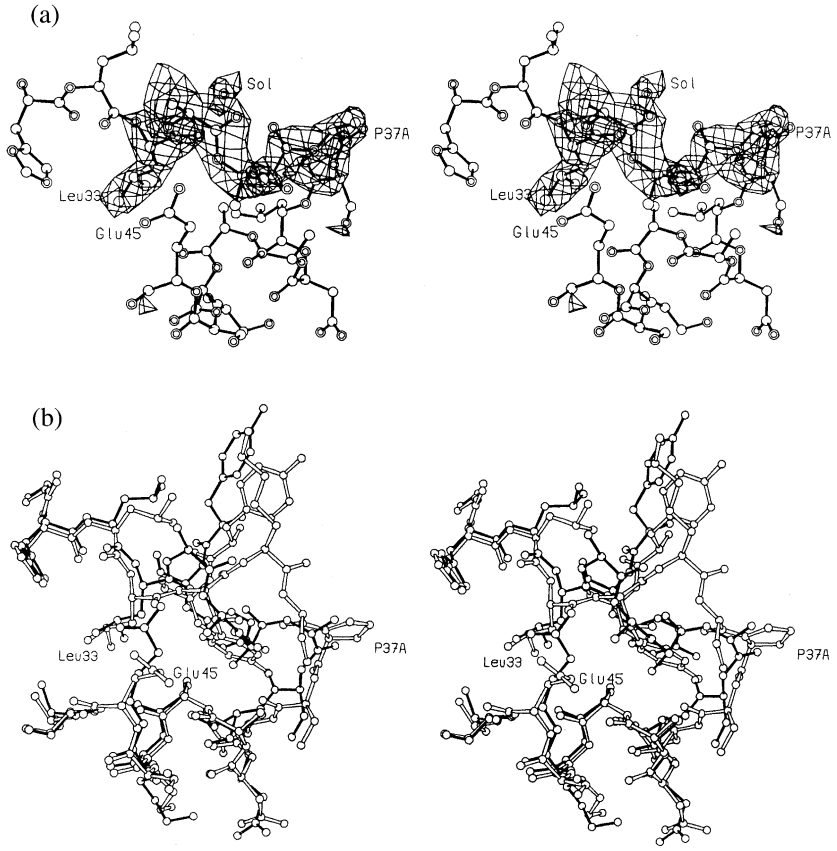

Fig. 3. (a) Difference electron density map in the vicinity of residues 33–38 in the mutant A34–37. The model of the structure with residues 33–38 deleted was refined for 8 cycles. Structure factors, F_c, and phases were then calculated and were combined with the observed amplitudes, F_o, to calculate a map with coefficients $(F_o - F_c)$. The map is contoured at 3σ and is to 2.2 Å resolution. (b) Superposition of the structure of mutant A34–37 (solid bonds) on WT (open bonds) showing the large change in conformation in the vicinity of the mutation. The overall superposition of the two structures is based on the backbone atoms within the amino-terminal domain, i.e. residues 15–70.

the (φ, ψ) values for Gly56 are in the 'left-handed' region of the Ramachandran diagram, common for glycine but rare for alanine.

The crystal structure of the mutant with 3 substitutions (i.e. mutant A53–57($C^{54}G^{56}$), sequence ^{53}ACAGA^{57}) was determined. In this case the mutated residues are all solvent-exposed and, even though Asn53 and Asn55 participate in a crystal contact, the mutant crystallized isomorphously with wildtype. The refined structure (Table 3) is very similar to that of wildtype with overall root-mean-square discrepancy between the backbone atoms of 0.22 Å. Within the region of the mutations (residues 50–59) the discrepancy was 0.19 Å.

3.3. Combination of two distant polyalanine helices

To investigate the combination of polyalanine substitutions at separate sites, replacements within α-helix 126–134 were combined with those in α-helix 39–50. These helices are approximately 40 Å apart (Fig. 2).

Table 4
Change in backbone dihedral angles in the vicinity of residues 32–40 for mutant A34–37

Wildtype			Mutant A34–37		
Amino acid	φ	ψ	Amino acid	φ	ψ
Leu32	−64	131	Leu	−68	134
Leu33	−95	−48	Leu	−99	80
Thr34	−164	150	Ala	−55	−22
Lys35	−89	−1	Ala	−91	−15
Ser36	−76	148	Ala	−96	120
Pro37	−73	−19	Ala	−108	8
Ser38	−81	125	Ser	−124	−106
Leu39	−66	−22	Leu	172	−41
Asn40	−78	−37	Asn	−58	−41

First, the mutant with the sequence ^{53}ACAGA^{57}, was combined with the previously described variant E128A/V131A/N132A [22], to give A53–57(C^{54}G^{56})/A128–132. The stability of this variant was close to WT* (Table 2). The mutant also gave crystals isomorphous with wildtype and a very similar structure (overall backbone discrepancy of 0.26 Å).

An alternative construction gave mutant A40–49(K^{43}L^{46})/A127–134(L^{133}) which combines the sequence segments ^{39}LAAAKAALAAAI^{50} and ^{126}WAAAAAALA^{134}. This protein is 5.2 °C more stable than wildtype at pH 3.0 (Fig. 4a) and is additive in terms of its constituent parts.

The mutant was crystallized isomorphously with wildtype, although with a decrease of 1.7 Å in the *c* cell dimension (Table 3). This is presumably due to the fact that although helices 39–50 and 126–134 are at opposite ends of the lysozyme molecule, helix 39–50 of one molecule forms a crystal contact with helix 126–134 of an adjacent molecule. Associated with the adjustment in the crystal contact, and the reduction in the cell dimension, there is a change of 5° in the hinge-bending angle between the two domains. Otherwise, the mutant structure is well conserved with root-mean-square difference in coordinates for the amino-terminal domain (i.e. residues 15–60) and carboxy-terminal domain (residues 80–160) of 0.27 and 0.26 Å, respectively. Within the carboxy-terminal domain the backbone of Lys135 moves by 1.2 Å, presumably due to a combination of the effects of the multiple mutations plus changes in the crystal contacts in which this residue participates. Within the amino-terminal domain there is no residue that has a root-mean-square coordinate shift larger than 0.5 Å.

None of the residues substituted here are thought to be directly involved in catalysis although Asn132, Lys135, Ser136, Arg138 and Gln141 participate in substrate binding [48].

4. Discussion

4.1. Tolerance to polyalanine substitutions within different types of secondary structure

It has previously been shown that multiple alanine substitutions in different helices of T4 lysozyme are well tolerated. Alanine substitutions at solvent-exposed sites generally change stability little, and, so long as favorable interactions are not lost, can often increase the stability of the protein [22,26–28]. Alanine substitutions at buried sites are usually destabilizing [33,49]. The thermodynamic measurements, per se, do not show whether these differences in stability are due to changes in the folded or the unfolded state.

The present results provide further support for these general principles. They also show that polyalanine substitutions in different α-helices can be combined together. The mutant A40–49(K^{43}L^{46})/A127–134(L^{133}) combines a total of 9 alanine substitutions in two different helices and actually increases the melting temperature of the protein by 5.2 °C. The mutant includes the two sequence segments ^{40}AAAKAALAAA^{49} and ^{127}AAAAAALA^{134}, i.e. a total of 15 alanines within a combined span of 18 residues. Nevertheless the protein folds normally, based on halo size at 37 °C is fully active, could be crystallized, and retained a 3D structure very similar to wildtype. In this case the two α-helices were far apart in the 3D structure and the effects on stability were additive.

In a second example, mutant A127–141(L^{133}W^{138}) included 10 alanine substitutions within two consecutive α-helices (126–134 and 137–141) without significantly changing stability (Table 2). The sequence of this mutant, i.e.

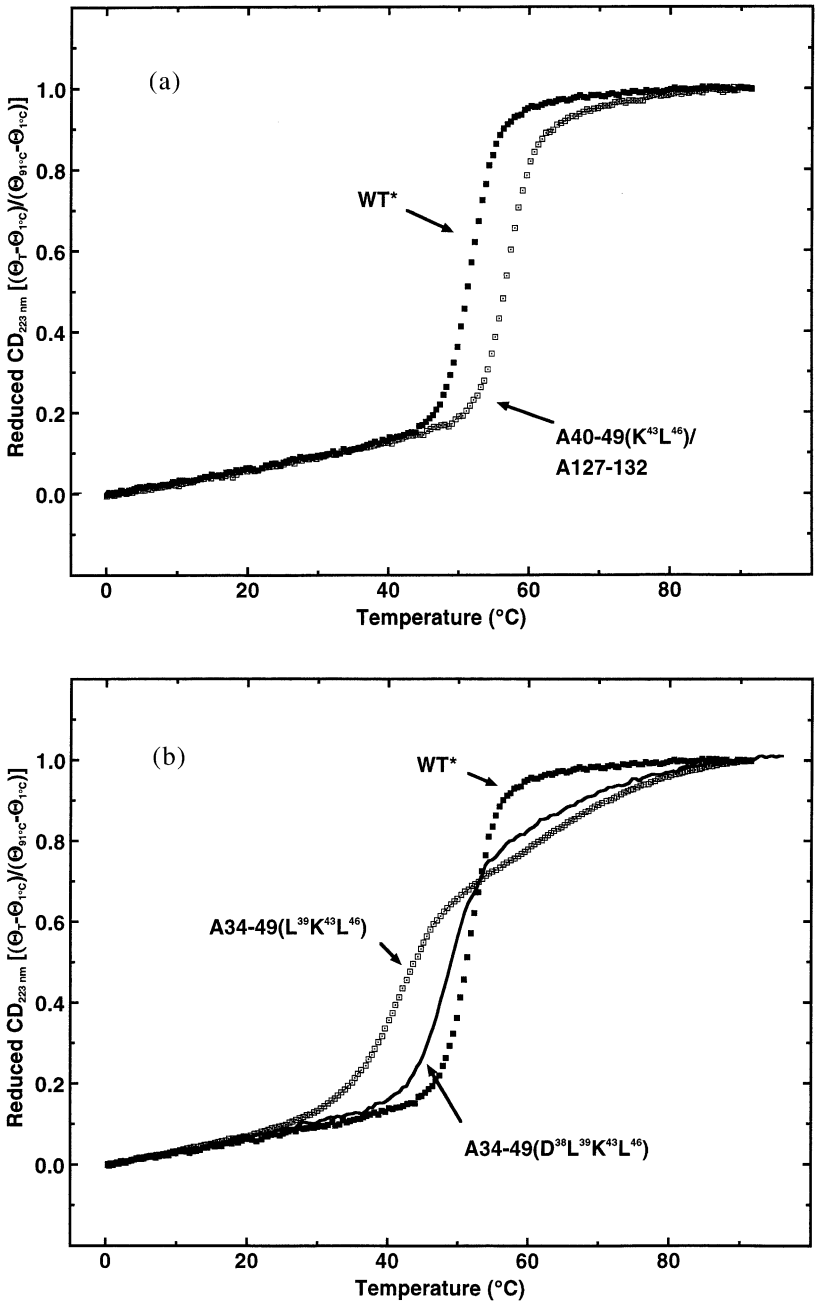

Fig. 4. Thermal unfolding transitions followed by changes in circular dichroism, of representative lysozyme mutants. (a) Comparison of standard, cysteine-free wildtype lysozyme (WT*) with mutant A40–49($K^{43}L^{46}$)/A127–132, which is 5.2° more stable. (b) Mutant A34–49($L^{39}K^{43}L^{46}$) has the sequence ^{34}AAAAALAAAKAALAAA^{49} and begins to unfold at a temperature less than that of wildtype, but seems to retain some helical character at a temperature when WT* is fully unfolded. If Ala38 in this polyalanine mutant is replaced with aspartic acid, to give the mutant A34–49($D^{38}L^{39}K^{43}L^{46}$), the unfolding transition reverts closer to that of wildtype.

^{127}AAAAAALAAAAWAAA^{141}, included 13 alanines within a contiguous span of 15 residues. This mutant includes the 2 substitutions K135A and S136A within the turn connecting the two helices. This and a number of other mutants with substitutions in turn or loop regions, including A53–57($C^{54}G^{56}$), A53–58, A52–58, A34–37 and A34–37/A128–132, all suggest that such substitutions can be tolerated (Table 2), although not quite so well as within α-helices. In the crystal structures of the mutants A53–57($C^{54}G^{56}$) and A53–57($C^{54}G^{56}$)/A128–132, both the overall structure and the local conformation of the backbone were well conserved suggesting that changes in stability that did occur were due to a loss of interactions within the substituted sidechains. In the crystal structure of mutant A34–37, however, a change to a more α-like conformation in the vicinity of the substitution was observed (Fig. 3b). This suggests that multiple alanines tend to adopt a helical conformation, consistent with their high α-helix propensity ([28] and references therein). In other words, multiple alanine substitutions within turn or loop regions will tend to adopt a conformation that differs from the conformation in native lysozyme (an extreme example of this is discussed below).

4.2. A helix initiation signal

It was previously shown that α-helix 39–50 can be stabilized by alanine substitutions, as in the polyalanine mutant A40–49($K^{43}L^{46}$) (Table 2) [27]. The loop that precedes the helix was also seen in the mutant A34–37 to be tolerant of alanine substitutions. When these two mutants were combined together, however, and the substitution Ser38→Ala was also included, to give the mutant A34–49($L^{39}K^{43}L^{46}$), the protein behaved in an unexpected manner. In the first place the stability was very much reduced ($\Delta T_m = -10.4$ °C) (Table 2). This was not anticipated because the two major constituent mutants, A40–49($K^{43}L^{46}$) and A34–37, were themselves either stabilizing or neutral. The remaining constituent mutant S38A was constructed as a single substitution, and, as expected, had little effect on stability since Ser38 is largely solvent-exposed. Also, the mutant S38D [50] is known to be slightly stabilizing (Table 2). The second unusual property of mutant A34–49($L^{39}K^{43}L^{46}$) is its atypical thermal transition (Fig. 4b) which suggests that the mutant protein begins to unfold at a temperature ~10 °C lower than wildtype, but then maintains substantial α-helical structure at a temperature well above that at which WT* is completely unfolded (Fig. 4b). This suggests that the high proportion of alanines present in A34–49($L^{39}K^{43}L^{46}$), i.e. ^{34}AAAAALAAAKAALAAA^{49}, tends to cause this region to form a single long α-helix. Such an extended α-helix is different from the native structure which consists of a loop followed by a helix.

In wildtype lysozyme Ser38 is at the N-cap position of α-helix 39–50 and its γ-hydroxyl accepts hydrogen bonds from the amide groups in the first turn of the α-helix [50]. It therefore seems that residue 38 might serve, at least in part, as an initiation signal for helix 39–50 (cf. [51]). According to this hypothesis, residues 34–49 in the alanine-rich mutant A34–39($L^{39}K^{43}L^{46}$) either form a single extended α-helix in the 'unfolded' state, or form such a helix during condensation of the structure, and so interfere with the normal folding process. This would rationalize the retention of α-helical structure under denaturing conditions, as judged by CD, and the substantial loss in stability of the folded versus the unfolded structure (Fig. 4b).

As a test of this hypothesis we performed random mutagenesis on residue 38 in the polyalanine background (i.e. in the context of the alanine-rich mutant A34–39($L^{39}K^{43}L^{46}$)). A small number of plaques gave halos indicative of functional lysozyme and on sequencing proved to contain the lysozyme in which Ala38 had been replaced by aspartic acid. Aspartic acid is a preferred helix-capping residue [52,53]. This single substitution, giving the mutant A34–49($D^{38}L^{39}K^{43}L^{46}$), caused the behavior of the protein to revert much closer, although not entirely, to wildtype (Fig. 4b). This strongly supports the idea that residue 38 provides at least part of a helix initiation signal that is necessary for the fold observed in T4 lysozyme.

We were not able to crystallize mutant A34–49($L^{39}K^{43}L^{46}$) but did obtain the structure of A34–49($D^{38}L^{39}K^{43}L^{46}$). In this structure the sidechain

of Asp38 appears to act as an N-cap for the residue 39–50 α-helix. However, an alanine at site 38 is not, of itself, substantially destabilizing (Table 2). Rather, it seems that in the context of the poly-alanine mutant A34–49($L^{39}K^{43}L^{46}$), Asp38 acts as a helix initiation signal.

In a typical protein, the information that codes for the 3D structure is presumably distributed over many amino acids. The present results suggest that in simplified sequences the key folding information may be restricted to a subset of critical residues, and so be more readily accessible to experimental analysis.

Acknowledgments

We sincerely thank Joan Wozniak, Joel Lindstrom and Sheila Snow for excellent technical assistance, Larry Weaver for help with the X-ray data collection of some of the crystal structures, Robert DuBose for help with the construction of some of the mutants, Hale Nicholson for access to unpublished data and Martin Sagermann for helpful discussion. This work was supported by a grant to B.W.M. from the National Institutes of Health (GM21967).

References

[1] B.W. Matthews, S.J. Remington, The three-dimensional structure of the lysozyme from bacteriophage T4, Proc. Natl. Acad. Sci. USA 71 (1974) 4178–4182.
[2] P.Y. Chou, G.D. Fasman, Empirical predictions of protein conformation, Ann. Rev. Biochem. 47 (1978) 251–276.
[3] B.W. Matthews, Comparison of the predicted and observed secondary structure of T4 phage lysozyme, Biochim. Biophys. Acta 405 (1975) 442–451.
[4] C. Schellman, The α_L conformation at the ends of helices, in: R. Jaenicke (Ed.), Protein Folding, Elsevier/North-Holland Biomedical Press, 1980, pp. 53–61.
[5] M. Sagermann, L.-G. Martensson, W.A. Baase, B.W. Matthews, A test of proposed rules for helix capping: implications for protein design, Prot. Sci. 11 (2002) 516–521.
[6] M. Elwell, J. Schellman, Phage T4 lysozyme physical properties and reversible unfolding, Biochim. Biophys. Acta 386 (1975) 309–323.
[7] M.L. Elwell, J.A. Schellman, Stability of phage T4 lysozymes: I. Native properties and thermal stability of wild-type and two mutant lysozymes, Biochim. Biophys. Acta 494 (1977) 367–383.
[8] M.L. Elwell, J.A. Schellman, Stability of phage T4 lysozymes: II. Unfolding with guanidinium chloride, Biochim. Biophys. Acta 580 (1979) 327–338.
[9] J.A. Schellman, Solvent denaturation, Biopolymers 17 (1978) 1305–1322.
[10] J.A. Schellman, M. Lindorfer, R. Hawkes, M. Grütter, Mutations and protein stability, Biopolymers 20 (1981) 1989–1999.
[11] R. Hawkes, M.G. Grütter, J.A. Schellman, Thermodynamic stability and point mutations of bacteriophage T4 lysozyme, J. Mol. Biol. 175 (1984) 195–212.
[12] W.J. Becktel, W.A. Baase, Thermal denaturation of bacteriophage T4 lysozyme at neutral pH, Biopolymers 26 (1987) 619–623.
[13] W.J. Becktel, J.A. Schellman, Protein stability curves, Biopolymers 26 (1987) 1859–1877.
[14] B.W. Matthews, Structural and genetic analysis of protein folding and stability, Curr. Opin. Struct. Biol. 3 (1993) 589–593.
[15] A. Fersht, Characterizing transition states in protein folding: an essential step in the puzzle, Curr. Opin. Struct. Biol. 5 (1994) 79–84.
[16] R.T. Sauer, M.E. Milla, C.D. Waldburger, B.M. Brown, J.F. Schildbach, Sequence determinants of folding and stability for the P22 Arc repressor dimer, FASEB J. 10 (1996) 42–48.
[17] K.A. Dill, H.S. Chan, From Levinthal to pathways to funnels, Nature Struct. Biol. 1 (1997) 10–19.
[18] D.V. Laurents, R.L. Baldwin, Protein folding: matching theory and experiment, Biophys. J. 75 (1998) 428–434.
[19] S.W. Golomb, R.E. Peile, R.A. Scholtz, Basic Concepts in Information Theory and Coding: The Adventures of Secret Agent 00111, Plenum Press, New York, 1994.
[20] B.C. Cunningham, J.A. Wells, High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis, Science 244 (1989) 1081–1084.
[21] J.U. Bowie, J.F. Reidhaar-Olson, W.A. Lim, R.T. Sauer, Deciphering the message in protein sequences: tolerance to amino acid substitutions, Science 247 (1990) 1306–1310.
[22] X.-J. Zhang, W.A. Baase, B.W. Matthews, Toward a simplification of the protein folding problem: a stabilizing polyalanine α-helix engineered in T4 lysozyme, Biochemistry 30 (1991) 2012–2017.
[23] M.E. Milla, B.B. Brown, R.T. Sauer, Protein stability effects of a complete set of alanine substitutions in Arc repressor, Nature Struct. Biol. 1 (1994) 518–523.
[24] M.-H. Yu, J.S. Weissman, P.S. Kim, Contribution of individual side-chains to the stability of BPTI examined by alanine-scanning mutagenesis, J. Mol. Biol. 249 (1995) 388–397.
[25] B.M. Brown, R.T. Sauer, Tolerance of Arc repressor to multiple-alanine substitutions, Proc. Natl. Acad. Sci. USA 97 (1999) 1983–1988.
[26] X.-J. Zhang, W.A. Baase, B.W. Matthews, Multiple alanine replacements within α-helix 126–134 of T4

lysozyme have independent, additive effects on both structure and stability, Prot. Sci. 1 (1992) 761–776.
[27] D.W. Heinz, W.A. Baase, B.W. Matthews, Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino acid sequence, Proc. Natl. Acad. Sci. USA 89 (1992) 3751–3755.
[28] M. Blaber, W.A. Baase, N. Gassner, B.W. Matthews, Alanine scanning mutagenesis of the α-helix 115–123 of phage T4 lysozyme: effects on structure, stability and the binding of solvent, J. Mol. Biol. 246 (1995) 317–330.
[29] M. Blaber, X.-J. Zhang, J.D. Lindstrom, S.D. Pepiot, W.A. Baase, B.W. Matthews, Determination of α-helix propensity within the context of a folded protein: sites 44 and 131 in bacteriophage T4 lysozyme, J. Mol. Biol. 235 (1994) 600–624.
[30] T.A. Kunkel, J.D. Roberts, R.A. Zakour, Rapid and efficient site-specific mutagenesis without phenotypic selection, Methods Enzymol. 154 (1987) 367–382.
[31] M. Matsumura, B.W. Matthews, Control of enzyme activity by an engineered disulfide bond, Science 243 (1989) 792–794.
[32] A.R. Poteete, S. Dao-pin, H. Nicholson, B.W. Matthews, Second-site revertants of an inactive T4 lysozyme mutant restore activity structuring the active site cleft, Biochemistry 30 (1991) 1425–1432.
[33] A.E. Eriksson, W.A. Baase, B.W. Matthews, Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences, J. Mol. Biol. 229 (1993) 747–769.
[34] J. Jancarik, S.-H. Kim, Sparse matrix sampling: a screening method for crystallization of proteins, J. Appl. Cryst. 24 (1991) 409–411.
[35] R. Hamlin, Multiwire area X-ray diffractometers, Methods Enzymol. 114 (1985) 416–452.
[36] X.-J. Zhang, B.W. Matthews, STRAT: a program to optimize X-ray data collection on an area detector system, J. Appl. Cryst. 26 (1993) 457–462.
[37] M.F. Schmid, L.H. Weaver, M.A. Holmes, et al., An oscillation data collection system for high-resolution protein crystallography, Acta Cryst. A 37 (1981) 701–710.
[38] M.G. Rossmann, Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting, J. Appl. Cryst. 12 (1979) 225–239.
[39] D.E. Tronrud, L.F. Ten Eyck, B.W. Matthews, An efficient general-purpose least-squares refinement program for macromolecular structures, Acta Cryst. A 43 (1987) 489–503.
[40] S. Dao-pin, T. Alber, W.A. Baase, J.A. Wozniak, B.W. Matthews, Structural and thermodynamic analysis of the packing of two α-helices in bacteriophage T4 lysozyme, J. Mol. Biol. 221 (1991) 647–667.
[41] X.-J. Zhang, J.A. Wozniak, B.W. Matthews, Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme, J. Mol. Biol. 250 (1995) 527–552.
[42] M.G. Rossmann, D.M. Blow, The detection of subunits within the crystallographic asymmetric unit, Acta Cryst. 15 (1962) 24–31.
[43] X.-J. Zhang, B.W. Matthews, Enhancement of the method of molecular replacement by incorporation of known structural information, Acta Cryst. D 50 (1994) 675–686.
[44] L.H. Weaver, B.W. Matthews, Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution, J. Mol. Biol. 193 (1987) 189–199.
[45] C.M. Wilmot, J.M. Thornton, β-turns and their distortions: a proposed new nomenclature, Prot. Eng. 3 (1990) 479–493.
[46] B.-L. Chen, W.A. Baase, H. Nicholson, J.A. Schellman, Folding kinetics of T4 lysozyme and nine mutants at 12 °C, Biochemistry 31 (1992) 1464–1476.
[47] D.E. Anderson, J. Lu, L. McIntosh, F.W. Dahlquist, The folding, stability and dynamics of T4 lysozyme: a perspective using nuclear magnetic resonance, NMR of Proteins, Macmillan Press, London, 1993, pp. 258–304.
[48] R. Kuroki, L.H. Weaver, B.W. Matthews, A covalent enzyme–substrate intermediate with saccharide distortion in a mutant T4 lysozyme, Science 262 (1993) 2030–2033.
[49] J. Xu, W.A. Baase, E. Baldwin, B.W. Matthews, The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect, Prot. Sci. 7 (1998) 158–177.
[50] H. Nicholson, W.J. Becktel, B.W. Matthews, Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles, Nature 336 (1988) 651–656.
[51] L.G. Presta, G.D. Rose, Helix signals in proteins, Science 240 (1988) 1632–1641.
[52] L. Serrano, A.R. Fersht, Capping and α-helix stability, Nature 342 (1989) 296–299.
[53] J.S. Richardson, D.C. Richardson, Amino acid preferences for specific locations at the ends of α helices, Science 240 (1988) 1648–1652.