Tolerance of a protein helix to multiple alanine and valine substitutions

Lydia M Gregoret[1] and Robert T Sauer[2]

Background: Protein stability is influenced by the intrinsic secondary structure propensities of the amino acids and by tertiary interactions, but which of these factors dominates is not known in most cases. We have used combinatorial mutagenesis to examine the effects of substituting a good helix-forming residue (alanine) and a poor helix-forming residue (valine) at many positions in an α helix of a native protein. This has allowed us to average over many molecular environments and assess to what extent the results reflect intrinsic helical propensities or are masked by tertiary effects.

Results: Alanine or valine residues were combinatorially substituted at 12 positions in α-helix 1 of λ repressor. Functional proteins were selected and sequenced to determine the degree to which each residue type was tolerated. On average, valine substitutions were accommodated slightly less well than alanine substitutions. On a positional basis, however, valine was tolerated as well as alanine at the majority of sites. In fact, alanine was preferred over valine statistically significantly only at four sites. Studies of mutant protein and peptide stabilities suggest that tertiary interactions mask the intrinsic secondary structure propensity differences at most of the remaining residue positions in this α helix.

Conclusions: At the majority of positions in α-helix 1 of λ repressor, tertiary interactions with other parts of the protein can be viewed as an environmental 'buffer' that help to diminish the helix destabilizing effects of valine mutations and allow these mutations to be tolerated at frequencies similar to alanine mutations.

Addresses: [1]Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Correspondence: Lydia M Gregoret
E-mail: gregoret@chemistry.ucsc.edu

Key words: combinatorial mutagenesis, helix propensity, protein stability, secondary structure

Received: **11 December 1997**
Revisions requested: **05 January 1998**
Revisions received: **23 January 1998**
Accepted: **30 January 1998**

Published: **27 February 1998**
http://biomednet.com/elecref/1359027800300119

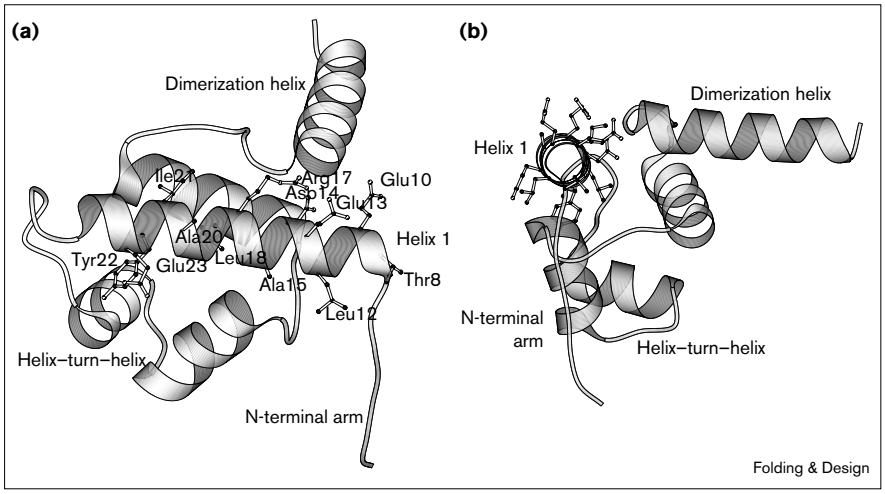
Folding & Design 27 February 1998, **3**:119–126

Introduction

Understanding the determinants of the stability of native proteins at the level of secondary and tertiary structure is a prerequisite for rational protein design or engineering. The interplay between the formation of local secondary structure and global tertiary contacts is also a crucial factor in the cooperativity of protein folding. The simplest elements of secondary structure in proteins are α helices. By varying the amino acids at certain, ideal positions in the α helices of proteins or peptides and measuring the accompanying changes in stability, scales have been devised that attempt to quantify the intrinsic α-helical propensities of different sidechains [1–7]. Although context is always a factor in such studies, the positions at which such changes are made are generally chosen to minimize the possibility of tertiary contacts and, in general, most helical propensity scales have similar relative rankings of amino acids. For example, alanine is always one of the best helix-forming residues, as are residues such as lysine, glutamine and methionine, whose sidechains have straight-chain aliphatic regions. In contrast, β-branched residues, such as valine, and amino acids with short sidechains that can form hydrogen bonds, such as asparagine and serine, are invariably poor helix formers. In energetic terms, the average difference between a good helix former like alanine and a poor helix former like valine is modest, ~0.5 kcal/mol, although this value can be as high as 1.0 kcal/mol [8] or as low as 0.3 kcal/mol [6] depending upon the scale chosen.

When amino acid substitutions are made at most α-helical positions in proteins, there can be a myriad of energetic effects caused by the perturbation of packing interactions or other types of tertiary contacts in addition to changes in helix propensity. The question then becomes whether the stability effects caused by differences in intrinsic secondary structure propensities are, on average, swamped out by these tertiary effects. To address this question, we have used combinatorial mutagenesis to examine the effects of substituting a good helix-forming residue (alanine) and a poor helix-forming residue (valine) at many positions in an α helix of a native protein. By studying multiple substitutions, small effects can be amplified. Thus, the extent to which it is possible to accumulate mutations of either type is a measure of the relative ease with which they may be accommodated in helical positions within a protein. More importantly, however, by probing many different positions

Figure 1



MOLSCRIPT [32] illustrations of λ repressor monomer with helix 1. **(a)** Helix 1 is in the foreground. Sidechains of the residues randomized in these experiments are shown. The N-terminal arm and helix–turn–helix are essential for binding to operator DNA. **(b)** End-on view of helix 1 showing the radial distribution of mutagenized residues. This view was achieved by a rotation of ~90° about the vertical axis of the configuration shown in (a).

simultaneously, the results are averaged over many individual molecular environments and we can then ask to what extent the results reflect intrinsic helical propensities or are masked by tertiary effects.

The combinatorial mutagenesis experiments reported here were performed in α-helix 1 of the N-terminal domain of λ repressor [9,10]. This 16-residue helix was chosen because it seems largely to serve a structural role in the protein (only the hydroxyl group of Tyr22 makes a DNA phosphate contact); the major surfaces required for operator-DNA binding or dimerization are formed by other parts of the N-terminal domain. In addition, previous studies of residue substitutions in helix 1 are available for comparison [11] and peptides corresponding to this helix show significant helix formation in solution, allowing the effects of mutations on the intrinsic stability of this helix to be studied in isolation [12].

Results

Combinatorial mutagenesis

12 structurally diverse residue positions in α-helix 1 of the N-terminal domain of λ repressor were chosen for mutagenesis. These include residues in the hydrophobic core (Ala15, Leu18, Ile21 and Tyr22), residues involved in an intrahelical salt bridge (Asp14 and Arg17), an N-cap residue (Thr8), and five surface residues (Glu10, Leu12, Glu13, Ala20 and Glu23). Figure 1 shows two views of these sidechains in the context of the native protein.

In one experiment, each of these positions was altered in a combinatorial manner using binomial mutagenesis so that the wild-type codon or a valine codon were equally probable [13]. The resulting 'valine' library consists of a pool of 2^{12} (4096) sequence variants. In a parallel experiment, the same positions were combinatorially mutated so that the wild-type codon or an alanine codon were equally probable. Because residues 15 and 20 are alanine in the wild-type protein, 1024 variants comprise this 'alanine' library.

Cross-streaking against phage λKH54 was used to screen 1700 transformants from the valine library and 1000 transformants from the alanine library for repressor function [13,14]. In this assay, biologically active repressor variants confer immunity to the host cell against the superinfecting phage. Plasmid DNA was isolated from immune clones and the genes for the N-terminal domain were sequenced. After correcting for out-of-frame and other error sequences, ~4% of the sequences in the valine library and 8% of the sequences in the alanine library were found to be active. This corresponds to ~160 active valine sequences and ~80 active alanine sequences. A number of repressor genes from non-immune clones were also sequenced for both libraries. Among inactive mutants, the average numbers of alanine and valine substitutions per sequence were 5.1 and 6.5, respectively. Because most mutants in both libraries are inactive, the average number of substitutions would be expected to be close to 5.0 for inactive alanine mutants and 6.0 for inactive valine mutants. The similarities between these observed and expected substitution values indicates that the ratios of mutant to wild-type codons in both combinatorial libraries must have been close to the expected 1:1 target values.

Sequences and residue tolerance

The sequences of functional variants from the valine and the alanine helix-1 libraries that were recovered from non-exhaustive sampling are shown in Table 1. The average number of substitutions per mutant sequence was 2.7 ± 1.0 for the valine library (25 active sequences and 12 mutagenized positions) and 2.9 ± 1.1 for the alanine library (36 active sequences and 10 mutagenized positions). If we

Table 1

Residues 8–23 of active helix-1 mutants.

Active alanine mutants	M	Active valine mutants	M
T E LED RL IYE		*T E LEDA RL AIYE*	
TQEQLEDARRLKAIY**A**	1	TQEQL**V**DARRLKAIYE	1
TQEQL**A**DARRLKAIYE	1	TQEQLEDARRLKAIY**V**	1
TQEQLE**A**ARRLKAIYE	1	TQEQLEDARRLKA**V**YE	1
AQEQLEDARRLKAIY**A**	2	TQEQLEDARRLK**V**IYE	1
TQEQL**A**DARRLKAIY**A**	2	**V**QEQLEDARRLKAIY**V**	2
TQ**A**QLEDAR**A**LKAIYE	2	**V**Q**V**QLEDARRLKAIYE	2
TQEQLEDAR**A**LKAIY**A**	2	TQEQLEDARRLK**VV**YE	2
TQEQLE**A**ARRLKA**A**YE	2	**V**QEQLEDARRLKA**V**YE	2
AQEQLEDAR**A**LKAIYE	2	TQ**V**QLEDARRLK**V**IYE	2
AQEQ**A**EDARRLKAIYE	2	TQEQLEDARRLK**V**IY**V**	2
TQ**A**QLEDARR**A**KAIYE	2	TQ**V**QLEDARRLK**V**IY**V**	3
TQEQLE**A**ARRLKAIY**A**	2	TQEQL**VV**ARRLK**V**IYE	3
TQEQ**A**EDAR**A**LKAIYE	2	TQEQLE**V**ARRLK**V**IY**V**	3
TQEQ**A**EDARRLKA**A**YE	2	TQ**V**QLEDARR**V**KA**V**YE	3
TQEQ**AA**DARR**A**KAIYE	3	TQEQ**V**EDARR**V**KAIY**V**	3
TQEQ**A**E**A**ARRLKAIY**A**	3	TQEQLE**V**ARR**V**KAIY**V**	3
AQEQLE**A**ARRLKAIY**A**	3	TQEQ**V**EDARRLK**V**IY**V**	3
AQEQ**A**EDARRLKAIY**A**	3	TQEQ**V**EDARR**V**K**V**IYE	3
TQEQ**A**EDARRLKA**A**Y**A**	3	**V**QEQLEDARRLK**V**IY**V**	3
AQEQLE**A**AR**A**LKAIYE	3	**V**Q**V**Q**V**EDARRLKAIY**V**	4
TQ**A**Q**A**EDARRLKAIY**A**	3	TQ**V**QLV**D**ARRLKA**V**Y**V**	4
TQEQ**AA**DARRLKAIY**A**	3	TQ**V**Q**V**EDAR**V**LKAIY**V**	4
TQ**A**QL**AA**ARRLKAIYE	3	TQ**V**QLEDARR**V**KA**V**Y**V**	4
TQ**A**QLEDAR**A**LKAIY**A**	3	TQEQ**VV**DARRLK**VV**YE	4
TQEQL**A**DARRLKA**A**Y**A**	3	TQ**V**Q**V**EDARR**V**K**V**IYE	4
AQ**A**QL**A**DARRLKA**A**YE	4	TQ**V**Q**V**EDARRLK**V**IY**V**	4
TQEQ**AAA**AR**A**LKAIYE	4		
AQ**A**QLE**A**AR**A**LKAIYE	4		
TQ**A**Q**AAA**ARRLKAIYE	4		
AQEQL**AA**AR**A**LKAIYE	4		
AQ**A**QLE**A**ARRLKAIY**A**	4		
TQEQ**v**E**A**AR**A**LKAIY**A**	4		
AQ**A**Q**A**EDARRLKAIY**A**	4		
TQ**A**QL**A**DARR**A**KAIY**A**	4		
TQ**A**Q**A**E**A**AR**A**LKAIYE	4		
TQ**A**Q**A**E**A**AR**A**LKAIYE	4		
AQ**A**QL**A**DAR**A**LKAIY**A**	5		
TQ**A**Q**AAA**AR**A**LKAIYE	5		

The sequences in italics show the wild-type residues at the positions that undergo mutations. Mutations are shown in bold type. M, the total number of mutations.

normalize these values by dividing by the expected number of substitutions if all mutations were functionally neutral (six for the valine library and five for the alanine library), then the fraction of allowed substitutions per position is 0.43 for the valine library and 0.58 for the alanine library. These results show that alanine is more easily accommodated than valine, but also reveal that the differences are modest and the distributions overlap substantially.

Table 2 and Figure 2 show the frequency at which valine and alanine were recovered in active variants at each mutagenized position in helix 1. At Glu23 and Ala20, substitutions were recovered at values similar to the input randomization frequency of 50%, as expected if substitutions at these positions are functionally neutral [13]. At

Table 2

Properties of mutated helix-1 positions.

Residue	Structural or functional role	ASA* (%)	Volume†	Alanine recovery (%)	Valine recovery (%)
Thr8	Helix N cap	81	3.0	32	19
Glu10	Solvent exposed	69	7.5	37	38
Leu12	Solvent exposed	56	3.9	39	31
Glu13	Solvent exposed	54	8.4	34	15
Asp14	Salt bridge to Arg17, hydrogen bond to Ser77	4	2.9	39	12
Ala15	Packs against helices 3 and 4	6	0.5	–	0
Arg17	Salt bridge to Asp14	47	6.2	34	4
Leu18	Packs against helices 3 and 4	0	2.5	8	23
Ala20	Solvent exposed	64	1.0	–	50
Ile21	Packs against helix 4–5 loop	12	3.1	13	27
Tyr22	Hydroxyl contacts phosphate; packs against helix 4 and helix 1–2 loop	5	6.5	0	0
Glu23	Solvent exposed	58	7.2	47	54

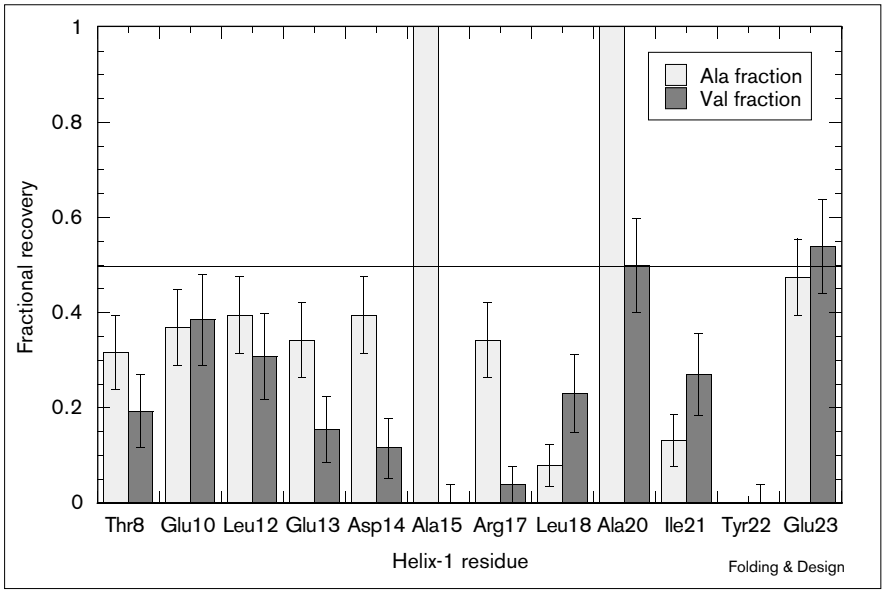
*ASA is the proportion of sidechain solvent-accessible surface area relative to the same sidechain in the central position of an extended tripeptide. †Volume in methylene group equivalents.

Thr8, Glu10, Leu12, Glu13, Asp14, Arg17, Leu18 and Ile21, valine and alanine substitutions were recovered at frequencies < 50%, suggesting that mutant substitutions at these residues are somewhat deleterious. Among these eight positions, alanine was clearly preferred over valine at Glu13, Asp14 and Arg17; valine was preferred over alanine at Leu18 and Ile21; and both residues were recovered at values within the expected sampling error at Thr8, Glu10 and Leu12. Only two positions were intolerant to substitutions; neither valine nor alanine was recovered in place of Tyr22 and valine was not an acceptable substitute for Ala15 in any of the functional mutants.

Stability of selected mutants in protein and peptide backgrounds

For studies of thermal stability, 12 mutant proteins with 1–4 valine or alanine substitutions at positions 8, 10, 12 and 23 were purified (Table 3). Most of the single mutants in this set were not recovered in the original activity screens and were constructed by cassette mutagenesis. All constructed mutants were found to be immune to superinfection. Table 3 lists the sequences and changes in T_m of each protein relative to wild type. The single substitutions reduce T_m by 2–5°C; the double substitutions at positions 8 and 23 reduced T_m by 8°C (double alanine) and 11°C (double valine); and the quadruple substitutions at positions 8, 10, 12 and 23 reduced T_m by 19°C. Two things are notable about these data. First, the ΔT_m values of the double and quadruple mutants are within a few degrees of those predicted from the effects of the single substitutions,

Figure 2



Fractional recovery of alanine and valine at mutagenized codons. If an alanine or valine mutation was functionally neutral, it should have a fractional recovery value of 0.5. The error bars represent 1 standard deviation unit in the expected sampling error as calculated using a Monte Carlo procedure [13]. Ala15 and Ala20 were not randomized in the alanine combinatorial mutagenesis experiment.

indicating that the mutant effects are largely additive. Second, alanine and valine substitutions at the same positions generally cause similar changes in T_m for both the single and multiple substitutions. For example, Figure 3 shows that the T8A/E10A/L12A/E23A and T8V/E10V/L12V/E23V proteins have similar melting profiles. Thus, in terms of thermal stability, alanine and valine seem to be tolerated almost equally well at positions 8, 10, 12 and 23. Figure 3 also shows that both quadruple mutants have melting temperatures some 19°C lower than that of wild type, indicating that the wild-type residues are preferred over alanines or valines at these positions.

The effects of the alanine and valine substitutions at positions 8, 10, 12 and 23 on the thermal stability of the N-terminal domain should reflect a combination of effects at the levels of secondary and tertiary structure. To discriminate

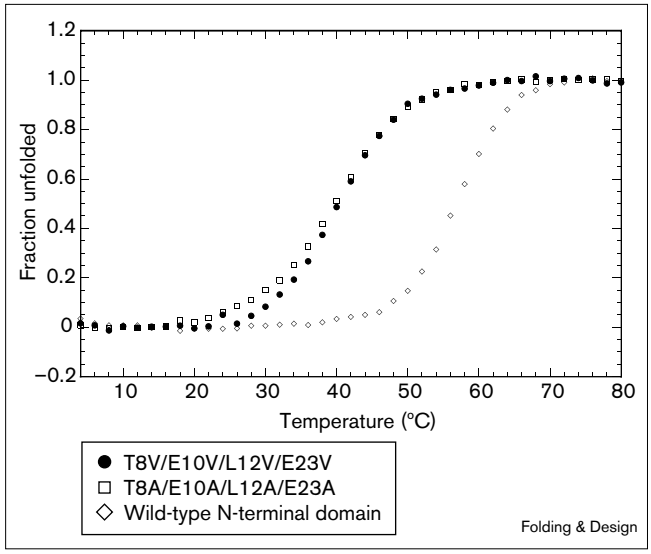
Table 3

Melting temperatures of proteins with helix-1 mutations.

Protein	T_m (°C)	ΔT_m (°C)
Wild type*	57	0
T8A	52	5
T8V	52	5
E10A	55	2
E10V	55	2
L12A	53	4
L12V	55	2
E23A	52	5
E23V	53	4
T8A/E23A	49	8
T8V/E23V	46	11
T8A/E10A/L12A/E23A	38	19
T8V/E10V/L12V/E23V	38	19

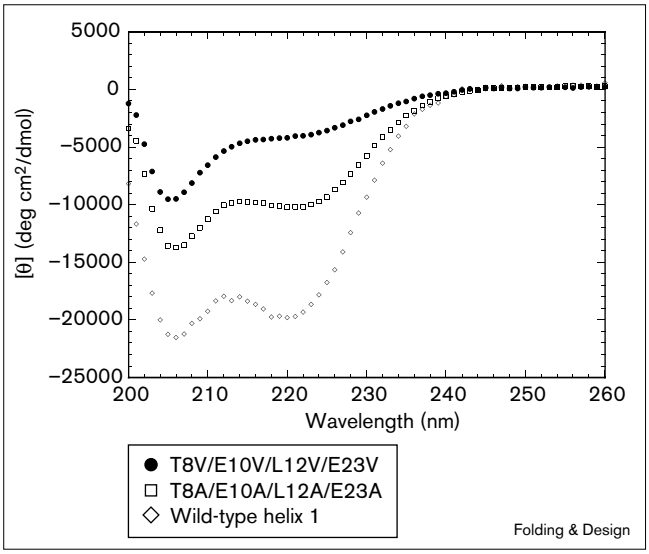
The errors on all T_m determinations were ± 1°C, with the exception of mutant T8V, which exhibited a poorly cooperative transition. The calculated T_m values were the same regardless of whether ΔC_p, the change in heat capacity upon unfolding, was fixed at 1400 kcal/mol deg or allowed to vary during the fitting procedure (see the Materials and methods section). *Residues 1–92 (the N-terminal domain) of λ repressor with a six-histidine tag.

Figure 3



Thermal denaturation of the T8V/E10V/L12V/E23V mutant, the T8A/E10A/L12A/E23A mutant, and the wild-type N-terminal domain of λ repressors. The fraction unfolded, f_u, is given by $f_u = K/(1 + K)$, where K was determined using a nonlinear least squares fit as described in the Materials and methods section.

Figure 4



CD spectra at 1°C of the T8V/E10V/L12V/E23V peptide, the T8A/E10A/L12A/E23A peptide, and the wild-type helix-1 peptide. Buffer conditions are described in the Materials and methods section.

between these two factors, peptides corresponding to residues 8–23 of the wild-type N-terminal domain, the T8A/E10A/L12A/E23A mutant, and the T8V/E10V/L12V/E23V mutant were synthesized with acetylated N termini and amidated C termini. The CD spectrum of the wild-type peptide has the highest helical content, the alanine-substituted peptide has reduced helicity, and the valine-substituted peptide shows a low level of helix formation (Figure 4). If we assume that a fully helical peptide would have a molar ellipticity of –34,000 deg cm^2/dmol at 222 nm, then the wild-type helix 1 peptide is ~60% helical, the quadruply substituted alanine peptide is ~30% helical, and the peptide with four valine substitutions is ~12% helical. (A peptide corresponding to residues 9–23 of λ repressor is 40% helical [12].) Thus, although the quadruple valine and alanine substitutions are tolerated equally well in the context of the tertiary structure of the intact N-terminal domain, this is not true in terms of secondary structure formation in isolation. In the peptide context, alanine is preferred over valine as expected from the intrinsic helical propensities of these two residues. In addition, although alanine is one of the best helix-forming residues, the presence of the wild-type residues at positions 8, 10, 12 and 23 promotes a significantly higher degree of helix formation than the presence of alanine at these positions.

Discussion

Sidechains in α-helix 1 help form the hydrophobic core and stabilize the native fold of the N-terminal domain of λ repressor by packing against portions of α helices 2, 3 and 4, and the loop regions preceding and following helix 4 [9]. We have used combinatorial mutagenesis to probe the tolerance of α-helix 1 to multiple alanine or valine substitutions. Alanine is a much better helix-forming residue than valine, and alanine mutations in helix 1, on average, were found to be accommodated more readily than valine mutations. It is also useful, however, to examine the distributions of allowed alanine or valine substitutions at individual positions, because average tolerance values can be biased by a few positions. On a positional basis, at seven of the 12 positions examined, either alanine and valine were accommodated within statistical error or valine was preferred (Figure 2). These data suggest that, at the majority of the positions examined, the tertiary effects of the alanine or valine substitutions mask the effects arising from stabilization of secondary structure. This is consistent with the findings of Pinker *et al.* [15] who demonstrated that the energetic effects of alanine substitutions in myoglobin were more highly correlated with the amount of lost buried surface area than with the free energy change in helix propensity upon mutation.

Our finding that helix 1 of λ repressor can accommodate as many as five alanine or four valine substitutions with retention of biological activity (Table 1) is not surprising given previous studies of helical tolerance. For example, Matthews and coworkers [16–18] demonstrated that multiple alanine substitutions could be accommodated in two helices in T4 lysozyme, we showed that α helices 2 and 3 of λ repressor were tolerant to as many as seven multiple alanine substitutions [13], and Johansson *et al.* [19] showed that 11 nearly consecutive valines are found in an α-helical conformation in the structure of pulmonary surfactant-associated polypeptide in non-polar solvents. We were, however, somewhat surprised to find that helix 1 of λ repressor, which has a largely structural involvement, is less tolerant to multiple alanine substitutions than repressor helices 2 and 3 [13], which mediate DNA recognition in addition to having a structural involvement. This might be interpreted as support for the idea that functional regions of proteins may not be optimized for stability [20,21] and are therefore more malleable in this respect.

Reidhaar-Olson and Sauer [11] previously examined the tolerance of α-helix 1 to amino-acid substitutions. We observe a greater degree of tolerance at residues 14, 17, 18 and 21 than was evident in their work, but both studies suggest that the majority of residue positions in helix 1, even those where sequence changes are tolerated, carry some amount of structural information. In the studies presented here, mutations at just two helix 1 positions (Ala20 or Glu23) were classified as being functionally neutral by the criteria that substitutions were recovered at frequencies close to 50% [13]. Positions 20 and 23 are both exposed on the surface. Two more positions (Ala15 and Tyr22) were completely intolerant to substitution. The

small Ala15 sidechain is overpacked, assigned the space of only 0.5 methylene groups by the Voronoi method [22], and completely buried (Table 2). Furthermore, it packs against one of the DNA recognition helices. Valine is apparently too large to be accommodated at this position in the core. The aromatic ring of Tyr22 is also largely buried and the hydroxyl group extends into solvent and contacts a DNA phosphate in the λ repressor–operator structure [10]. Hence, substitutions at position 22 would leave a large hole in the protein interior and also reduce operator affinity. At the remaining eight positions (Thr8, Glu10, Leu12, Glu13, Asp14, Arg17, Leu18 and Ile21), mutations were tolerated but the frequencies of recovery were < 50% indicating that the alanine or valine substitutions are accommodated at some structural or functional cost to the protein. In the entire set of mutations, the frequency of allowed substitutions showed a modest correlation with the fractional solvent accessibility of the wild-type sidechain (Table 2) for both the alanine substitutions (the correlation coefficient R = 0.67) and the valine substitutions (R = 0.54), consistent with the general idea that hydrophobic core positions are more important than surface positions in determining protein stability [23]. Little correlation was observed between the volume occupied by the wild-type residue and the substitution frequency in either the alanine or valine case.

Why, contrary to the expectations of secondary structure stabilization, is valine tolerated at some helical positions better than others? We note that the four positions at which alanine is clearly preferred over valine (Glu13, Asp14, Ala15 and Arg17) occur in the middle of helix 1, hinting at the possibility that helical propensity might be more important at the center and less critical towards the ends of this helix. Alternatively, there might simply be more opportunities for favorable tertiary contacts for valine near the beginning and end of helix 1. This is certainly the case at Leu18 and Ile21, which are largely buried, and where the preference for valine over alanine is likely to occur because valine can substitute better in core packing. At Thr8, Glu10, Leu12, Ala20 and Glu23, which are surface exposed, alanine and valine were tolerated at frequencies within sampling error. In addition, alanine and valine substitutions at four of these positions caused similar decreases in the thermal stabilities of single and multiple mutants. Thus, at these surface positions in the context of the native N-terminal domain, tertiary effects generally seem to outweigh secondary effects. Interestingly, this is not true in the isolated peptide model of helix 1. Thus, the T8A/E10A/L12A/E23A and T8V/E10V/L12V/E23V mutant proteins had the same thermal stability, whereas the T8A/E10A/L12A/E23A peptide showed a higher degree of helix formation than the T8V/E10V/L12V/E23V peptide. These data suggest that any tertiary contacts that help ameliorate the effects of valine substitutions at positions 8, 10, 12 and 23 must be provided by or require regions of the N-terminal domain other than helix 1.

Tertiary interactions with other parts of the protein can be viewed as an environmental 'buffer' that helps diminish the effects of helix destabilizing mutations in helix 1. Indeed, in protein G, it has recently been shown that the same 11-residue sequence can be forced into either a helical or a strand–hairpin conformation depending on the surrounding tertiary context [24]. It has also been shown that helical propensity is not conserved among helical peptides dissected from analogous helical regions of evolutionarily related proteins [25], indicating that tertiary interactions are required to stabilize some sequences in a particular secondary structure. A better understanding of the magnitude and extent of tertiary influence on secondary structure is essential for the development of the next generation of protein structure prediction methods.

Materials and methods

Combinatorial mutagenesis and cloning

Plasmid pLG100, a derivative of pWL105 [14,26], encodes residues 1–92 of λ repressor with a C-terminal tail of six histidines under transcriptional control of a *tac* promoter. The region of the gene encoding helix 1 (residues 8–23) is flanked by unique *Hpa*I and *Sac*I sites. Different mutagenic oligonucleotide cassettes were constructed for combinatorial alanine and valine substitutions. The coding strand of each cassette was synthesized by standard phosphoramidite chemistry using reagents and an Applied Biosystems DNA synthesizer. At positions where the wild-type codon and the alanine (or valine) codon differed by one base, a 1:1 mixture of the two bases was used in synthesis. At the codons that differed by two bases, a resin-splitting strategy [27] was used to achieve a 1:1 ratio of mutant to wild-type codons. The column containing the oligonucleotide was removed from the synthesizer, dried for 5 min under vacuum, and then opened. The resin was removed, divided in half, and repackaged into two new columns. Three bases corresponding to the wild-type codon were added to one column, and three bases coding for the alanine or valine were added to the other column. The resin was then recombined until the next position that required resin splitting was reached, at which point the procedure was repeated. The non-coding strands of each cassette were synthesized enzymatically using the Klenow fragment of DNA polymerase (New England Biolabs) and a short oligonucleotide primer. The double-stranded oligonucleotide was digested with *Sac*I (New England Biolabs) to create the proper overhang at one end and was then ligated to the *Sac*I/*Hpa*I backbone fragment of pLG100. This material was transformed into *Escherichia coli* strain X90 by electroporation and transformants were selected for ampicillin resistance. Some single alanine or valine substitution mutants were constructed by standard methods of cassette mutagenesis.

Activity screen

λ repressor N-terminal domain variants were classified as active if they conferred immunity to the host cell against phage λ KH54 in a cross-streaking assay [13]. The assay was performed at 30°C under conditions in which the *tac* promoter mediated expression of the protein was induced with isopropyl-β-D-thiogalactoside (IPTG). The activities of all functional mutants were verified by restreaking.

Solvent accessibility and volume calculations

Solvent accessibility was computed using the program ACCESS by Scott Presnell, which implements the Lee and Richards algorithm [28] for determining contact and accessible surface areas. The surface areas were normalized by those in a reference extended tripeptide [29].

Sidechain volumes were computed using the method of Voronoi polyhedra as implemented in the program CALC-VOLUME by Gerstein *et al.* [22]. For normalization, the volume of one methylene group equivalent was taken to be 22.5 Å^3. Calculations were performed on the λ-repressor dimer [10] (with DNA removed) and the values were averaged.

Protein purification

Proteins were purified using Nickel-NTA affinity chromatography (Qiagen Inc.). Typically, a 1 l bacterial culture of the appropriate strain was grown at 37°C in Luria–Bertani broth to an absorbance at 600 nm of 0.6–1.0. Protein expression was then induced by addition of IPTG (0.2 g/l) and growth was continued for 3 h. Cells were harvested by centrifugation at 4000 × g and resuspended in 20 ml of Buffer A (6 M guanidine hydrochloride, 100 mM NaH_2PO_4, 10 mM Tris, pH 8.0). After stirring at room temperature for 1 h, the lysate was cleared by centrifugation at 15,000 × g and the supernatant was loaded on a 1 ml Ni-NTA column equilibrated with Buffer A. The column was washed with 100 ml of Buffer A and the histidine-tagged protein was eluted with 20 ml of 6 M guanidine hydrochloride and 0.2 M acetic acid. Purified proteins were dialyzed into 10 mM Tris-HCl (pH 8.0), 2 mM $CaCl_2$, 0.1 mM EDTA, 1.4 mM β-mercaptoethanol and 5% glycerol. Protein concentration was determined from the absorbance at 275 nm using a 1 cm cuvette; an absorbance of 1 corresponds to an N-terminal domain concentration of 0.633 mg/ml.

Peptide synthesis

The peptides containing multiple alanine or valine mutations in helix 1 were synthesized, purified by reverse phase HPLC, and analyzed for amino acid composition by the MIT Biopolymers Laboratory. The wild-type helix 1 peptide was synthesized in the laboratory of Glenn Millhauser at the University of California, Santa Cruz. Peptide concentrations were determined from the absorbance at 275 nm in 6 M guanidinium hydrochloride using an extinction coefficient of 1450 M^{-1} cm^{-1}. The peptide CD spectra showed no dependence on concentration over a 10-fold concentration range, indicating that the peptides are monomeric.

Circular dichroism and thermal melts

Proteins and peptides were dialyzed into or dissolved in buffer containing 50 mM potassium phosphate buffer (pH 7.0) and 100 mM KCl, and aggregated material was removed by centrifugation at 13,000 × g at 4°C. CD spectra were collected using an Aviv 60DS spectropolarimeter. For proteins, a concentration of 3 μM and a cuvette with a 1 cm pathlength were used. At this concentration, the N-terminal domain of λ repressor is ~98% monomeric [30]. For peptides, the concentration was 200 μM and the pathlength was 1 mm. As estimated from the CD ellipticity at 222 nm, most mutant proteins had α-helical contents similar to wild type, but the T8A/E10A/L12A/E23A and T8V/E10V/L12V/E23V mutant proteins displayed apparent reductions in α helicity of ~20%. Both these quadruple mutants, however, aggregated to some extent and we were unable to determine whether the apparent reduction in helicity was caused by the formation of a non-helical aggregate.

Thermal melts were performed with 1°C temperature increments and a 1 min equilibration at each temperature using a Hewlett-Packard temperature controller or using a programmable water bath set to heat at a rate of 0.3 deg/min. Data were fitted using a Macintosh version of the program NONLIN [31] to the equation:

$$y = \frac{K}{(1+K)}\,[(y_f - m_f T) - (y_u - m_u T)] + (y_u + m_u T) \qquad (1)$$

where $K = \exp(-\Delta G/RT)$, where $\Delta G = [(T_m - T)/T_m]\Delta H - (T_m - T)\Delta C_p + T_m \Delta C_p \ln(T_m/T)$; y_f and y_u are the intercepts of the folded (lower) and unfolded (upper) baselines, respectively; m_f and m_u are the slopes of these baselines; T_m is the temperature of 50% denaturation; ΔH is the enthalpy of denaturation at the T_m and ΔC_p is the heat capacity change upon denaturation.

Acknowledgements

We thank Alan Davidson and Glenn Millhauser for advice and comments on the manuscript, Steven Lendaris and Bob Kalpin for preliminary data on single-substitution mutants, Chris Stenland for synthesizing peptides, and the Sauer and Kim labs for helpful suggestions. This work was supported in part by grants from the NIH (AI-15706), from the Jane Coffin Childs Foundation, and from the University of California, Santa Cruz.

References

1. Lyu, P.C., Liff, M.I., Marky, L.A. & Kallenbach, N.R. (1990). Sidechain contributions to the stability of alpha-helical structure in peptides. *Science* **250**, 669-673.
2. O'Neil, K.T. & DeGrado, W.F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly-occurring amino acids. *Science* **250**, 646-651.
3. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T.M. & Baldwin, R.L. (1990). Relative helix-forming tendencies of nonpolar amino acids. *Nature* **344**, 268-270.
4. Horovitz, A., Matthews, J.M. & Fersht, A.R. (1992). Alpha-helix stability in proteins. II. Factors that influence stability at an internal position. *J. Mol. Biol.* **227**, 560-568.
5. Blaber, M., Zhang, X.-J. & Matthews, B.W. (1993). Structural basis of amino acid alpha helix propensity. *Science* **260**, 1637-1640.
6. Blaber, M., Zhang, X.-J., Lindstrom, J.D., Pepiot, S.D., Baase, W.A. & Matthews, B.W. (1994). Determination of α-helix propensity within the context of a folded protein. *J. Mol. Biol.* **235**, 600-624.
7. Myers, J.K., Pace, C.N. & Scholtz, J.M. (1997). A direct comparison of helix propensity in proteins and peptides. *Proc. Natl Acad. Sci. USA* **94**, 2833-2837.
8. Rohl, C.A., Chakrabartty, A. & Baldwin, R.L. (1996). Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **5**, 2623-2637.
9. Pabo, C.O. & Lewis, M. (1982). The operator-binding domain of λ repressor: structure and DNA recognition. *Nature* **298**, 443-447.
10. Beamer, L.J. & Pabo, C.O. (1992). Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. Mol. Biol.* **227**, 177-196.
11. Reidhaar-Olson, J.F. & Sauer, R.T. (1990). Functionally acceptable substitutions in two α-helical regions of λ repressor. *Proteins* **7**, 306-316.
12. Marqusee, S. & Sauer, R.T. (1994). Contributions of a salt bridge-hydrogen bond network to the stability of secondary and tertiary structure in λ-repressor. *Protein Sci.* **3**, 2217-2225.
13. Gregoret, L.M. & Sauer, R.T. (1993). Additivity of mutant effects assessed by binomial mutagenesis. *Proc. Natl Acad. Sci. USA* **90**, 4246-4250.
14. Lim, W.A. & Sauer, R.T. (1991). The role of internal packing interactions in determining the structure and stability of a protein. *J. Mol. Biol.* **219**, 359-376.
15. Pinker, R.J., Lin, L., Rose, G.D. & Kallenbach, N.R. (1993). Effects of alanine substitutions in α-helices of sperm whale myoglobin on protein stability. *Protein Sci.* **2**, 1099-1105.
16. Zhang, X.-J., Baase, W.A. & Matthews, B.W. (1991). Toward a simplification of the protein folding problem: a stabilizing polyalanine α-helix engineered in T4 lysozyme. *Biochemistry* **30**, 2012-2017.
17. Heinz, D.W., Baase, W.A. & Matthews, B.W. (1992). Folding and function of a T4 lysozyme containing ten consecutive alanines illustrates the redundancy of information in amino acid sequence. *Proc. Natl Acad. Sci. USA* **89**, 3751-3755.
18. Zhang, X.-J., Baase, W.A. & Matthews, B.W. (1992). Multiple alanine replacements within α-helix 126-134 of T4 lysozyme have independent, additive effects on both structure and stability. *Protein Sci.* **1**, 761-776.
19. Johansson, J., Szyperski, T., Curstedt, T. & Wüthrich, K. (1994). The NMR structure of the pulmonary surfactant-associated polypeptide SP-C in an apolar solvent contains a valyl-rich α-helix. *Biochemistry* **33**, 6015-6023.
20. Meiering, E.M., Serrano, L. & Fersht, A.R. (1992). Effect of active site residues in barnase on activity and stability. *J. Mol. Biol.* **225**, 585-589.
21. Shoichet, B.K., Baase, W.A., Kuroki, R. & Matthews, B.W. (1995). A relationship between protein stability and protein function. *Proc. Natl Acad. Sci. USA* **92**, 452-456.
22. Gerstein, M.B., Tsai, J. & Levitt, M. (1995). The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.* **249**, 955-966.

23. Bowie, J.U., Reidhaar-Olson, J.F., Lim, W.A. & Sauer, R.T. (1990). Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **247**, 1306-1310.
24. Minor, D.L., Jr. & Kim, P.S. (1996). Context-dependent secondary structure formation of a designed protein sequence. *Nature* **380**, 730-734.
25. Muñoz, V., Blanco, F.J. & Serrano, L. (1995). The distribution of alpha-helix propensity along the polypeptide chain is not conserved in proteins from the same family. *Protein Sci.* **4**, 1577-1586.
26. Lim, W.A., Farruggio, D.C. & Sauer, R.T. (1992). Structural and energetic consequences of disruptive mutations in a protein core. *Biochemistry* **31**, 4324-4333.
27. Cormack, B.P. & Struhl, K. (1993). Regional codon randomization: defining a TATA-binding protein surface required for RNA polymerase III transcription. *Science* **262**, 244-248.
28. Lee, B. & Richards, F.M. (1971). The interpretation of proteins structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.
29. Spolar, R.S., Livingstone, J.R. & Record, M.T.J. (1992). Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry* **31**, 3947-3955.
30. Weiss, M.A., Pabo, C.O., Karplus, M. & Sauer, R.T. (1987). Dimerization of the operator binding domain of phage lambda repressor. *Biochemistry* **26**, 897-904.
31. Santoro, M.M. & Bolen, D.W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27**, 8062-8068.
32. Kraulis, P.J. (1991). MOLSCRIPT - a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/fad – for further information, see the explanation on the contents pages.