A Genetic and Biochemical Analysis of the Bacteriophage λ Cro Protein

by
Andrew Arnold Pakula
B. S., Biol. and Chem., Tufts University
(1980)

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the
Massachusetts Institute of Technology

September, 1988

Signature of Author

____

Certified by

____
Robert T. Sauer, Thesis Supervisor

Accepted by

____
Robert T. Sauer, Chairman, Biology Graduate Committee

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

SEP 15 1988

LIBRARIES

A Genetic and Biochemical Analysis of the Bacteriophage λ Cro Protein

by
Andrew Arnold Pakula

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

Abstract

This thesis describes the use of amino acid substitution mutations to probe the sequence determinants of stability in the λ Cro protein. I first isolated a large number of defective Cro proteins using random mutagenesis of a cloned *cro* gene and a genetic screen. The thermal stability of mutant Cro proteins is reflected in their resistance to proteolysis *in vivo*, thus examination of intracellular protein concentrations allowed me to correlate the severity of the destabilizing effects of an amino acid substitution with its chemical nature and its position in the three-dimensional structure of Cro. The largest class of destabilizing mutations consisted of those that affected buried residues. Some of these introduced polar or charged residues into the hydrophobic core, while other substitutions are more conservative and replace one hydrophobic side chain with another. Substitutions of buried residues presumably destabilize Cro by disrupting hydrophobic packing interactions. A smaller class of destabilizing substitutions alters residues that are solvent exposed in the wild type structure. Some of these substitutions replaced glycines at positions where their less restricted backbone dihedral angles are required, some introduced prolines into α-helices, and others disrupted charge-stabilized hydrogen bonds.

To investigate the determinants of thermal stability further, I isolated second site substitutions in Cro that relieve the proteolytic sensitivity *in vivo* of an unstable mutant Cro protein. Among these second site suppressors were two that increased thermal stability dramatically. Others reduce proteolytic sensitivity without increasing thermal stability, presumably by altering DNA binding properties, or by increasing the tendency of the protein to form protease-resistant aggregates.

The two strongly stabilizing substitutions isolated as second site suppressors both affect solvent exposed residues. One of these replaces tyrosine at position 26 with a cysteine. The other replaces glutamine at position 16 with leucine. To investigate the mechanisms by which these substitutions stabilized Cro, I constructed additional amino acid substitutions at each of these residue positions. The hydrophobicities of the substituted amino acid side chains were observed to be related to their effects on stability, however, the direction of these effects were opposite for the two positions: increased stability resulted from increased hydrophobicity at position 16, and decreased hydrophobicity at position 26. We explain these observations in relation to the energetically favorable consequences of removing hydrophobic surface area from exposure to solvent. The side chain of residue 16 is more solvent exposed in the unfolded state than in the folded, and thus hydrophobic residues at this position favor folding. The side chain of residue 26 is hyperexposed in the folded structure. I propose that hydrophobic residues at this position destabilize the folded state because they are less exposed to solvent in the unfolded state. Either or both of these mechanisms for increasing protein stability might be suitable for use in protein stabilization strategies.

Acknowledgements

I am grateful to the many people who helped me in large and small ways, technical, theoretical, and personal. Michael Lichten, Hillary Nelson, and Drew Vershon were particularly generous with technical assistance. Thanks to Chris Kaiser for being ready for an argument at any time.

Thanks especially to Bob Sauer. I could not have asked for a better thesis advisor.

Table of Contents

Abstract 2
Acknowledgements 4
Table of Contents 5
List of Figures 6
List of Tables 7
Chapter I - The Determinants of Protein Stability 8
Introduction 9
Hydrophobic Interactions 11
Van der Waals Interactions 12
Electrostatic Interactions 13
Hydrogen Bonds 14
Conformational Entropy 15
Effects on the Unfolded State 16
References 18
Chapter II - λ Cro Mutations: Effects on Activity and Intracellular Degradation 22
Introduction 23
Materials and Methods 23
Results 26
Discussion 29
References 34
Chapter III - Amino Acid Substitutions that Increase the Thermal Stability of the λ Cro Protein 51
Introduction 52
Materials and Methods 53
Results 58
Discussion 63
References 67
Chapter IV - The Hydrophobicity of Surface Residues Affects Protein Stability 82
Introduction 83
Materials and Methods 83
Results 85
Discussion 89
References 94

List of Figures

Chapter II.

Fig. 1. Mutations that affect expression of Cro 39

Fig. 2. Protein sequences of wild type Cro and mutant amino acid substitutions 41

Fig. 3. CRM levels of missense mutant proteins. 43

Fig. 4. Pulse-chase experiments 45

Chapter III.

Fig. 1. DNA and amino acid sequences of the synthetic *cro* gene 75

Fig. 2. Locations of suppressor substitutions in the Cro structure................77

Fig. 3. Intracellular turnover of mutant Cro proteins 79

Fig. 4. Thermal denaturation profiles of mutant Cro proteins 81

Chapter IV.

Fig. 1. Thermal denaturation profiles of Cro proteins altered at position 16......100

Fig. 2. Thermal stabilities of Cro proteins altered at residue 16 are related to residue hydrophobicity...102

Fig. 3. Thermal denaturation profiles for mutant Cro proteins altered at position 26...104

Fig. 4. Thermal stabilities of Cro proteins altered at residue 26 are related to residue hydrophobicity...106

Fig. 5. Thermal denaturation profiles for wild type, Leu16, Asp26, and Leu16/Asp26 Cro proteins..108

List of Tables

Chapter II.

Table 1. λ Cro Mutations 47

Chapter III.

Table 1. DNA and inferred amino acid sequence changes in the cro genes of revertant candidates 71

Table 2. Melting temperatures and intracellular half-lives of mutant Cro proteins 73

Chapter IV.

Table 1. Stability changes associated with residue substitutions at position 16 96

Table 2. Stability changes associated with residue substitutions at position 26 98

Chapter I

The Determinants of Protein Stability

Introduction:

There is currently a great deal of interest in the interactions that contribute to the folding and stabilization of protein structures. Ultimately, it is hoped that an understanding of these interactions will allow the prediction of protein structure from sequence and design *de novo* of proteins with useful activity and stability properties. At the present time, however, the realization of these goals seems far off. In fact, it is not yet possible to reliably predict the effects of single amino acid substitutions in proteins of known structure. The forces that are involved in protein stabilization are understood only in a general sense. Although it is clear that hydrophobic forces, electrostatic interactions, van der Waals interactions, and hydrogen bonds, all affect protein structure and stability, our ability to estimate the magnitude of their contributions in specific structural contexts is limited.

Comparison of the properties of proteins that differ by one or more amino acid substitutions provides one approach to the understanding of protein stability. A number of studies have involved the comparison of the amino acid sequences of homologous proteins that differ in thermal stability. This type of analysis suffers from the drawback that the proteins usually differ by many amino acid substitutions, and ascribing the stability changes to one or a few of these substitutions is difficult. The overall impression that one gains from this work is that increased protein stability generally results from the combination of a number of small stabilizing changes, and that many different types of interactions are involved, including hydrogen bonds, hydrophobic effects, ion pairs, and changes in helical propensity (Perutz and Raidt, 1975; Argos *et al* , 1979; Walker, *et al.*, 1980). The use of single amino acid substitutions, generated either by random or directed mutagenesis of a cloned gene, provides a more rigorous way of examining the importance of various types of interactions, and such studies have begun to yield useful results. In this introductory chapter, I will review the forces which stabilize proteins and some of the

general conclusions about protein stability that have emerged from comparative studies of evolutionarily related and mutant proteins.

Amino acid substitutions that affect stability represent a major fraction of temperature sensitive and defective mutations in proteins. In proteins of known structure, these "stability mutations" fall into several classes. Some introduce polar groups, holes, or bulky groups into the hydrophobic core; some disrupt tertiary hydrogen bonds; some introduce prolines into α–helices; and some replace glycines whose conformational flexibility is essential for turns. Mutations affecting the hydrophobic core comprise the largest of these groups and tend to have severely destabilizing effects. For example, in λ Cro (chapter 2), the N-terminal domain of λ repressor (Hecht *et al.*, 1983; 1984), and phage T4 lysozyme (Alber *et al.*, 1987), amino acid substitutions in the hydrophobic core represent 75-95% of the stability mutations. Single residue changes in the hydrophobic core are capable of completely disrupting protein folding. For example, Leu^{57} forms part of the hydrophobic core of the N-terminal domain of λ repressor (T_m=54°C). When this residue is replace by alanine or glycine, the melting temperature is reduced to 20°C for the Ala^{57} protein and 4°C for the Gly^{57} protein (Parsell and Sauer, 1988). In energetic terms, these reductions in melting temperature correspond to $\Delta\Delta G$ changes of 5.9 kcal/mol and 8.6 kcal/mol, respectively. Moreover, these mutations simply remove a portion of the core and should therefore be less deleterious than those that introduce polar or bulky groups.

For λ repressor, the folded form of the N-terminal domain is only about 5 kcal/mol more stable than the unfolded form at room temperature. Although this stabilization energy is modest it is not atypical. Stabilization energies for folded proteins commonly range from 3 to 15 kcal/mol. The net stability of a protein represents a small difference between large and opposing forces. The main factor opposing protein folding is the enormous loss of conformational entropy that occurs upon folding. For a protein of 100 amino acids, this loss of conformational entropy has been estimated to be in the range of 200-500 kcal/mol at

25°C (Creighton, 1984). Since most proteins are stably folded at this temperature, this large and unfavorable term must be balanced by favorable contributions from hydrogen bonds, van der Waals interactions, and electrostatic interactions in the native protein structure, and by the favorable entropic contribution of the hydrophobic effect.

Hydrophobic Interactions:

The hydrophobic effect (Kauzman, 1959) has been proposed to make the largest contribution to the stabilization of folded protein structures (Chothia, 1975). The basis of this phenomenon is the free energy cost of the increased structure that is imposed upon water at its interface with non-polar molecules. In an unfolded protein, water is presumed to be organized in cage-like structures around the hydrophobic side chains. Protein folding results in a large decrease in the non-polar surface area that is accessible to solvent, and thus much of the structured water is released. This increases the disorder of the solvent and provides a substantial driving force for protein folding. Estimates of the magnitude of the hydrophobic effect (200-300 kcal/mol for a protein of 100 amino acids; Creighton, 1984) have relied primarily on the free energies for the transfer of amino acid side chains from water to water vapor or to solvents such as ethanol, dioxane, N-methylacetamide, or cyclohexane (Nozaki and Tanford, 1971; Finney, 1975; Damodaran and Song, 1986; Radzicka and Wolfenden, 1988; Wolfenden *et al.*, 1981). The large number of solvents listed reflects a persistent uncertainty as to which, if any of these solvents, is a good representation of the nature of the protein interior. Nonetheless, in virtually all solvent systems, the transfer free energies of nonpolar side chains are proportional to their solvent accessible surface areas.

Recently, amino acid substitutions have been used to examine the contribution of individual side chains to the overall hydrophobic stabilization of proteins. Multiple substitutions at single positions have been constructed in kanamycin nucleotidyl transferase

(KNTase; Matsumura *et al.*, 1988a) and T4 lysozyme (Matsumura *et al.*, 1988b). In these studies, an essentially linear relationship was found between residue hydrophobicity and stabilization free energy. That is, the change in free energy of stabilization of the substituted proteins was roughly equal to the change in the transfer free energy for the relevant residues. These results suggest that the simple model of transfer free energy provides a reasonable estimate of the hydrophobic effect in proteins. However, recent work with the small ribonuclease, barnase, has shown that substitutions in the hydrophobic core result in decreases in stability that are considerably larger than those predicted based on free energies of amino acid side chain transfer (Kellis *et al.*, 1988). Similar discrepancies have also been observed in studies of substitutions in the hydrophobic core of the N-terminal domain of λ repressor (Parsell and Sauer, 1988). One factor that might account for the differences observed in the these studies is the ability of the protein interior to accommodate side chains of different sizes. For the T4 lysozyme studies, and probably also for KNTase, the altered residue is only partially buried and hence amino acid substitutions can probably be accommodated by local adjustments in packing interactions. In these cases, the change in stability should be chiefly due to changes in hydrophobic interactions. By contrast, replacing a tightly packed, completely buried residue in barnase or λ repressor may require leaving an energetically unfavorable hole in the hydrophobic core. Here, the reduced stability will result from the loss of hydrophobic interactions and from the cost of having a cavity in the protein interior.

Van der Waals Interactions:

Van der Waals forces or London dispersion forces are weak attractive forces that occur between all types of atoms. These forces are essentially electrostatic and arise because the short-range interactions between two atoms result in transient changes in electron distribution, such that both atoms take on transitory, and complementary, dipolar character. In the unfolded protein there will be van der Waals interactions with solvent,

and thus the contribution of these forces to protein stability should depend on the differences between the packing interactions made in the unfolded and folded states. The packing density in the interiors of folded proteins is extremely efficient, and as high as in crystals of small organic molecules (Chothia, 1975; Richards, 1977). This suggests that the energy of van der Waals interactions is optimized in the hydrophobic core. Based on the van der Waals energies calculated for crystal packing, the free energy of stabilization contributed by van der Waals interactions for a protein of 100 residues has been suggested to be close to 200 kcal/mol (Creighton, 1984).

Although many amino acid substitutions have been isolated that affect molecular packing in the hydrophobic core, the energetic effects of these mutations also include changes in hydrophobic interactions. Analysis of the energetic consequences of such modifications is further complicated by the observed flexibility of proteins, which sometimes allows altered packing interactions to compensate for changes in hydrophobic core residues (Alber *et al.*, 1988).

Electrostatic Interactions:

The association of two charged groups in aqueous solution is relatively weak, because charged groups are heavily solvated by water, and because water has a high dielectric constant. Electrostatic interactions are, however, stronger in non-polar solvents where these effects are reduced or absent. The theoretical evaluation of the energetic contributions of electrostatic attraction in proteins is complicated by several uncertainties: *(i)* What is the appropriate dielectric constant for protein interiors? *(ii)* What is the appropriate dielectric at the solvent-protein interface?, and *(iii)* Is the concept of the dielectric constant, which is defined only for bulk solvent, applicable for modeling interactions over small interatomic distances?.

Several studies of homologous proteins have suggested that ionic interactions can contribute to increased thermal stability (Perutz and Raidt, 1975; Walker *et al.*, 1980; Ruegg *et al.*, 1982). The energetic contributions of a buried ion pairs has been estimated at 2.5-3.0 kcal/mol (Fersht, 1971), while solvent exposed salt bridges in bovine pancreatic trypsin inhibitor and hemoglobin apparently contribute about 1 kcal/mol to overall stability (Brown *et al.*, 1978; Perutz, 1978). The interaction of charged groups with α-helix peptide dipoles has been shown to affect the stability of α-helices in solution (Shoemaker *et al.*, 1987; Mitchinson and Baldwin, 1986; Marqusee and Baldwin, 1988) and of ribonuclease reconstituted from the ribonuclease S-protein and various S-peptide variants (Mitchinson and Baldwin, 1986).

Hydrogen Bonds:

Since water can both donate and accept hydrogen bonds, any hydrogen bond made in a protein forms at the cost of breaking hydrogen bonds with solvent. Hence, it is unclear whether hydrogen bonds provide any net stabilization to the protein. In fact, initial studies of hydrogen bonds in small molecule model systems suggested that these interactions were unlikely to contribute substantial amounts of energy to the stability of the folded protein structure (Klotz and Franzen, 1962). However, recent studies have shown that hydrogen bonds in protein-substrate interactions can contribute 0.5 to 1.8 kcal/mol, while those involving a charged residue can contribute up to 6 kcal/mol (Fersht *et al.*, 1985).

Several hydrogen bonds mediated by side chains in T4 lysozyme have been studied using amino acid replacement (Grutter *et al.*, 1987; Alber *et al.*, 1987b; 1988). At Thr^{157}, a substantially buried position, it is clear that a side chain hydrogen bond contributes about 1.2 kcal/mol to the overall stability of T4 lysozyme. However, similar analysis of interactions made by the side chain of residue 86, a solvent exposed position, showed no

correlation between hydrogen bonding and increased stability (Alber *et al.*, 1988). The differences in the importance of hydrogen bonds at these positions is currently unclear. Perhaps the hydrogen bond mediated by residue 86 forms only in the crystal and not in solution. Alternatively, the hydrogen bond could be locally favorable, but not contribute to overall stability. For example, this hydrogen bond might not need to be broken for unfolding to occur.

It is important to keep in mind that, irrespective of the contribution of hydrogen bonding energy to protein stabilization, buried hydrogen bond donors and acceptors must be paired in the protein structure. Otherwise, the loss of favorable hydrogen bonds with water will detract from protein stability. The peptide backbone contains the majority of the hydrogen bonding groups of the protein. The hydrogen bonding potentials of these groups are generally satisfied via interactions within or between elements of secondary structure.

Conformational Entropy:

Covalent cross-links have been found to stabilize proteins in a number of instances, and thus the introduction of disulfide linkages has been a commonly employed strategy for attempting to increase protein stability through rational design (Katz and Kossiakoff, 1986; Wells and Powers, 1986; Sauer, *et al.*, 1986; Villafranca *et al.*, 1987; Pantoliano *et al.*, 1987; for review see Creighton, 1988; Wetzel *et al.*, 1988). The stabilization afforded by disulfides is thought to be due to a reduction in the number of conformations accessible to the unfolded protein, and thus to a reduced entropy of unfolding.

Amino acid substitutions may also affect protein stability by altering the local conformational freedom of the peptide backbone. For example, because glycine lacks a β-carbon it can assume many backbone dihedral angles that are energetically unfavorable for other amino acids. This property is extremely important because it allows glycine to be used in turns, where unusual dihedral angles are often required. However, glycines in α-

helices and β-strands have backbone angles that could be assumed by any residue. Such glycines may limit protein stability by increasing the entropy of the unfolded form. In fact, Gly→Ala substitutions in α-helices have been shown to stabilize proteins in several cases (Hecht *et al.*, 1986; Imanaka *et al.*, 1986; Matthews *et al.*, 1987). Presumably, glycine is a poor helix former relative to residues such as alanine for this same reason (Chou and Fasman, 1978). Proline, because of its pyrolidine ring, has extremely limited backbone flexibility about its N-C_α bond (the phi dihedral angle is restricted to -60 degrees). Hence, for residue positions where this phi angle is acceptable, substitutions of the form Xxx→Pro may stabilize proteins by decreasing the entropy change associated with unfolding (Matthews *et al.*, 1987). Proline, like glycine, is a poor helix forming residue (Chou and Fasman, 1978) but for a different reason. If proline occurs in the middle or at the C-terminal end of an α-helix, there are steric clashes between the pyrolidine side chain and the β-carbon of the previous residue (Schimmel and Flory, 1968) and one of the helical hydrogen bonds is lost because proline does not have a peptide -NH group. Proline can occur at the N-terminal end of an α-helix because there are no steric clashes with the preceding residue and the peptide -NH of these residues of the helix are not involved in intrahelical hydrogen bonds.

Effects on the Unfolded State:

With the exception of the glycine and proline substitutions discussed above, it is generally held that amino acid substitutions are likely to exert their influence on protein stability by affecting interactions that occur in the native protein. The unfolded protein is viewed essentially as a random coil, whose properties are relatively insensitive to amino acid substitution. This view is supported by the finding that the stabilizing or destabilizing effects of most substitutions can generally be rationalized in terms of effects on the folded protein (Alber *et al.*, 1987a) and by the finding that the equilibrium changes caused by many desabilizing mutations is accounted for by changes in the rate of unfolding (Matthews

and Hurle, 1987). However, Shortle has recently challenged this view and has suggested that a number of amino acid substitution in Staphylococcal nuclease exert their affects on stability through changes in the properties of the unfolded state (Shortle and Meeker, 1986). Several amino acid substitutions in tryptophan synthase have also been proposed, based on kinetic data, to alter stability via effects on the unfolded state (Beasty *et al.*, 1986; Matthews and Hurle, 1987). Thus, the unfolded state needs to be kept in mind when the effects of amino acid substitutions on protein stability are considered.

References:

Alber, T., Sun, D.-P., Nye, J. A., Muchmore, D. C., Matthews, B. W. Biochemistry 26:3754-3758, 1987a.

Alber T., Sun, D.-P., Wilson, K., Wozniak, J. A., Cook, S. P., Matthews, B. W. Nature 330:41-46, 1987b.

Alber T., Bell, J. A., Sun, D.-P., Nicholson, H., Wozniak, J. A., Cook, S. P., Matthews, B. W. Science, 239:631-635, 1988.

Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frank, G., Tratschin, J. D. Biochemistry 18:5698, 1979.

Beasty, A.M., Hurle, M.R., Manz, J.T., Stackhouse, T., Onuffer, J.J., Matthews, C.R. Biochemistry 25:2965-2974, 1986.

Brown, L. R., DeMarco, A., Richarz, R., Wagner, G., Wuthrich, K. Eur. J. Biochemistry 88:87-95, 1978.

Chothia, C. Nature 254:304-308, 1975.

Chou, P. Y., Fasman, G. D. Ann. Rev. Biochemistry 47:251-276, 1978.

Creighton, T. E. Proteins: Structures and Molecular Properties. W. H. Freeman and Co., New York, 1984.

Creighton, T.E. Bioessays 8:57, 1988.

Damodaran, S., Song, K.B. J. Biol. Chem. 261:7220-7222, 1986.

Fersht, A. R. Cold Spring Harbor Symp. Quant. Biol. 36:71-73, 1971 .

Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y., Winter, G. Nature 314:235-238, 1985.

Finney, J.L. J. Mol. Biol. 96:721-732, 1975.

Grutter MG, G.T.M., Weaver LH, W.T.A., Matthews BW, J Mol Biol 197:315-329, 1987.

Hecht , M.H., Nelson, H.C.M., Sauer, R.T. Proc. Natl. Acad. Sci. 80:2676-2680, 1983.

Hecht, M.H., Sturtevant, J.M., Sauer, R.T. Proc. Natl. Acad. Sci. 81:5685-5689, 1984.

Hecht, M. H., Sturtevant, J. M., Sauer, R. T. Proteins 1:43-46, 1986.

Imanaka, T., Shibazaki, M., Takagi, M. Nature 324:695-697, 1986.

Katz, B.A., Kossiakoff, A. J Biol Chem 261:15480-5, 1986.

Kauzman, W. Adv. Prot. Chem. 14:1-63, 1959.

Kellis, J. T. Jr., Nyberg, K., Sali, D., Fersht, A. R. Nature 333: 784-6, 1988.

Klotz, I. M., Franzen, J. S. J. Am. Chem. Soc. 84:3461-3466, 1962.

Marqusee, S., Baldwin, R. L. Proc. Natl. Acad. Sci., 1988 (in press).

Matthews, B.W., Nicholson, H., Becktel, W.J. Proc Natl Acad Sci USA 84:6663-6667, 1987.

Matthews, C.R., Hurle, M.R. Bioessays 6:254-2577, 1987.

Matsumura, M., Yahanda, S., Yasumura, S., Yutani, K., Aiba, S. Eur J Biochem 171:715-720, 1988a.

Matsumura, M., Becktel, W.J., Matthews, B.W. Nature, 1988.

Mitchinson, C., Baldwin, R., L. Proteins 1:23-33, 1986.

Nozaki, Y., Tanford, C. J. Biol Chem. 246:2211-2217, 1971.

Pantoliano, M.W., Ladner, R.C., Bryan, P.N., Rollence, M.L., Wood, J.F., Poulos, T.L. Biochemistry 26:2077-2082, 1987.

Parsell, D. A., Sauer, R. T. J. Biol. Chem., 1988 (submitted).

Perutz, M. F. Science 201:1187-1191, 1978.

Perutz, M.F., Raidt, H. . Nature 255:256-259, 1975.

Radzicka, A., Wolfenden, R. Biochemistry 27:1664-1670, 1988.

Richards, F.M. Ann. Rev. Biophys. Bioeng. 6:151-176, 1977.

Ruegg, C., Ammer, D., Lerch, K. . J. Biol. Chem. 257:6420, 1982.

Sauer, R.T., Hehir, K., Stearman, R.S., Weiss, M.A., Jeitler-Nilsson, A., Suchanek, E.G., Pabo, C.O. Biochemistry 25:5992-8, 1986.

Schimmel, P. R., Flory, P. J. J. Mol. Biol. 34: 105-120, 1968.

Shoemaker, K.R., Kim, P.S., York, E.J., Stewart, J.M., Baldwin, R.L. Nature 326:563-567, 1987.

Shortle, D., Meeker, A.K. Proteins 1:81-89, 1986.

Villafranca, J.E., Howell, E.E., Oatley, S.J., Xuong, N.H., Kraut, J. Biochemistry 26:2182-2189, 1987.

Walker, J.E., Wonacott, A.J., Harris, J.I. Eur. J. Biochemistry 108:581-586, 1980.

Wells, J.A., Powers, D.B. 261:6564-6570, 1986.

Wetzel, R., Perry, L.J., Baase, W.A., Becktel, W.J. Proc Natl Acad Sci USA 85(2):401-5, 1988.

Wolfenden, R., Andersson, L., Cullis, P.M., Southgate, C.C.B. Biochemistry 20:849-855, 1981.

Chapter II

λ Cro Mutations: Effects on Activity and Intracellular Degradation

Introduction

The study of missense mutants provides a means of testing and adding to our understanding of the determinants of protein structure, stability and function. In this paper, we describe mutations that affect the intracellular level and activity of λ Cro, a small repressor with a known three-dimensional structure (1-3). We find that a Cro^- phenotype results from replacing any one of at least half of the 66 residues of Cro; some of the mutations change residues on the DNA binding surface, some change solvent-exposed residues distant from the proposed binding surface, and some affect residues in the hydrophobic core. Mutations that would be expected to destabilize the folded structure of Cro generally result in extreme sensitivity to intracellular proteolysis. In many cases, the resulting reduction in concentration is severe enough to account fully for the observed mutant phenotype.

Materials and Methods:

Plasmids, Phage and Bacteria: Three related Cro-producing plasmids - pAP100, pAP101 and pAP104 - were constructed for this work. Each plasmid is a pBR322 derivative and each bears the same fusion of the *lac* UV5 promoter to the λ *cro* gene. This fusion was derived from plasmid pTR214 (4). Plasmid pAP100 was constructed by inserting the 353 base pair *Eco*RI/*Nae*I fragment from pTR214 between the *Eco*RI and filled-in *Cla*I sites of pBR322 (5). Plasmid pAP101 was constructed by replacing the 391 base pair *Cla*I/*Bam*HI fragment of pAP100 with the 375 base pair *Eco*RI/*Bam*HI fragment from pBR322. pAP101 has an intact *tet* promoter and thus confers a significantly higher level of tetracycline resistance than pAP100. pAP104 is similar to pAP101 but contains the M13 origin of replication. pAP104 was constructed by replacing the small *Eco*RI/*Bam*HI

fragment of pZ150 (6) with that from plasmid pAP101. Strains containing pAP100, 101 or 104 express the same level of λ Cro as judged by the resistance of the strain to phage infection and by competitive radioimmunoassay. Thus, Cro^- mutations isolated in any of the three plasmid backgrounds can be directly compared.

λ200 is an immunity 21 bacteriophage that carries the *lac*Z gene under the control of the λ P_R promotor (7). Phage M13 app1 was constructed by deletion *in vitro* of the *lac* promoter and operator sequences of M13 mp8 (8). DNA between the unique *Ava*II and *Eco*RI sites was deleted, the DNA was ligated, and candidates were screened for those with a regenerated EcoRI site. The DNA of one such phage, M13 app1, was sequenced and shown to contain a deletion of M13 mp8 bases 5013-6232.

Escherichia coli strains used for this study were grown at 37° and include: MM294 (9); JM103 (10); GW5100 (JM103 cured of P1 lysogen; L. Marsh and G. Walker, unpublished); KL731 (11); US3 (Thi^-, His^-, *lac*Z^-, sup^o, *rec*A^-); LE30: (*mut*D5, *rps*L, *azi*, *gal*U95.; L. Enquist via D. Botstein); and DB6438 (DXIII, rif^r, *arg*Eam, *met*B, *mut*T198; J. Miller via D. Botstein). LB broth contains Bacto-tryptone (10 g/l), Bacto-yeast extract (5 g/l), NaCl (5 g/l), and 1 mM NaOH. LBT is LB broth plus 5 μg/ml tetracycline. LBA is supplemented with 200 μg/ml ampicillin.

Mutagenesis and Mutant Isolation: Plasmids were mutagenized *in vivo* with N-methyl-N'-nitroso-guanidine (MNNG) or 2-aminopurine (12) or by passage through mutator strains mutD (LE30) or mutT (DB6438). Purified pAP101 DNA was mutagenized *in vitro* with hydroxylamine (13).

Following mutagenesis *in vivo*, pAP100 or pAP101 DNA was isolated by the small-scale rapid boiling method (14). pAP104 contains an M13 origin of replication and packaged single stranded DNA transducing particles were obtained from supernatants of cultures of GW5100(pAP104) infected with M13 RV1 helper phage (6,15). The supernatants were sterilized by incubation at 65-70° for 30 minutes.

To identify plasmids bearing mutant *cro* genes, mutagenized plasmid DNA was introduced into strain US3/λ200 or its F^{+} derivative. Cells transformed by Cro^{-} plasmids form red (Lac^{+}) colonies after growth on MacConkey-lactose-tetracycline (5 μg/ml) plates. Mutagenized pAP100 or pAP101 DNA was introduced into US3/λ200 by transformation. To introduce pAP104 into US3/λ200 F^{+}, a freshly saturated culture was incubated with transducing particles (m.o.i.=0.1) for 30 min at 37° and then plated.

DNA Sequencing: For each Cro^{-} mutant candidate, the entire *cro* gene and *lac* promoter region were sequenced by the dideoxy method (16). For mutations isolated in pAP100 or pAP101, the *Eco*RI/*Hin*dIII fragment that contains the P_{lac}-*cro* fusion was subcloned into M13 mp8 (8) or M13 app1. M13 app1 is a useful cloning vector for restriction fragments containing the *lac* operator because these phages form blue plaques on $F^{+}Lac^{+}$ strains, such as KL731, when plated on X-gal indicator plates. This phenotype is presumed to result from titration of Lac repressor from the chromosomal *lac* operator. Cro^{-} mutations

in pAP104 backgrounds were sequenced using single-stranded plasmid DNA template obtained from transducing particles present in supernatants of M13 RV1 infected cultures of GW5100(pAP104) (6).

Assays of Cro Activity and Level: β-galactosidase assays were performed as described by Miller (12) on log phase cultures of US3/λ200 or its F^+ derivative bearing wild type or Cro^- plasmids. Competitive radioimmunoassays (RIA) were performed on lysates of the same strains essentially as described by Johnson (17). For western blot analysis, lysates were electrophoresed in 20% polyacrylamide gels containing $NaDodSO_4$ and urea (18), and were transferred to nitrocellulose electrophoretically (19). The nitrocellulose-bound Cro was detected by incubation with rabbit anti-Cro serum and subsequently with ^{125}I-Staph protein A (Amersham).

Pulse-Chase Experiments: Log phase cultures of US3/λ200 carrying wild type or Cro^- plasmids were pulse-labelled with L-[^{35}S]methionine (Amersham). Aliquots were removed 0.5, 10 and 60 minutes after addition of excess unlabelled L-methionine and lysates were prepared. Cro protein was immunoprecipitated, electrophoresed on acrylamide gels (18), and detected by autoradiography.

Results:

After random mutagenesis of a plasmid containing the λ *cro* gene, Cro^- mutations were isolated by screening for Lac^+ transformants of an indicator strain containing a λ P_R-

lacZ fusion. Cro represses transcription from the P_R promoter by binding at the O_R1 or O_R2 operator DNA sites (3), and thus β-galactosidase levels are inversely related to Cro activity in the indicator strain. For Cro^- mutants, high β-galactosidase levels indicate severe mutant defects.

Sequence Changes and Activities of Cro Mutants: The mutation in each Cro^- plasmid was identified by DNA sequencing, and was found to be a substitution, deletion or insertion of a single base pair. Table 1 lists the DNA sequence changes identified for 175 independent Cro^- mutant isolates and, where appropriate, indicates the inferred amino acid changes. A total of 79 different Cro^- mutations were isolated. Sixty-two are missense mutations; six are nonsense mutations; four affect translational start signals; five are *lac* promoter mutations; and two are frameshift mutations. Table 1 also lists the β-galactosidase levels in the λ P_R-*lacZ* fusion strain for each mutant.

The specificities of the mutations induced by the chemical mutagens and mutator strains are similar to those reported by other workers (20-22). Mutations that arose spontaneously during growth of the M13 origin plasmid, pAP104, as single stranded DNA were either G→T or G→C transversions. These mutations arose at a moderate frequency (10^{-5} to 10^{-4}), and may result from replication past apurinic sites in the single stranded DNA (23).

Promoter and Translational Start Mutations: Nine of the Cro^- mutations change bases in the *lac* UV5 promoter, in the Shine-Dalgarno ribosome binding sequence, or in the ATG translation initiation codon, as shown in Figure 1. The steady state level of Cro, as

measured by RIA, was reduced by all of these mutations (Table 1). Combining the steady state level data for these mutants with the corresponding β-galactosidase activity measurements establishes a reference curve for wild-type Cro. By comparison of the corresponding data for a mutant Cro protein, the relative specific activity of the mutant protein *in vivo* may be determined.

Missense Mutations: Figure 2 shows the wild type amino acid sequence of Cro and the missense mutations which, as a group, affect 32 of the 66 residue positions. The mutant sites are distributed throughout the protein sequence. We have not saturated all of the possible sites of missense mutation. Caruthers *et al.* (24) have constructed site-directed changes at three residue positions not represented in our collection and report that these changes confer a Cro-defective phenotype. Since many of the mutants in our collection were isolated only once and certain mutagenic specificities are underrepresented, it is likely that there are a number of other sites in Cro where amino acid substitutions will cause a defective phenotype.

The levels of cross reacting material (CRM) for each mutant were measured by competitive RIA. These data are listed in Table 1 and are represented schematically in Figure 3. All of the mutant proteins appear to be present in reduced amounts relative to wild type. Of the 62 missense mutants, 24 have CRM levels from 5-60% of the wild type level and the remaining 38 are present at levels below 5% of wild type. When mutant levels were estimated by Western blot analysis using high antisera concentrations, results comparable to the RIA data were obtained (data not shown).

The severely reduced levels of the mutant Cro proteins probably result from increased rates of cellular proteolysis. Figure 4 shows pulse-chase experiments for wild type Cro and three of the mutant proteins. Wild type Cro turns over with a half life of 30 to 60 minutes, whereas the Asp9→Tyr, Asp47→Gly, and Arg38→Leu mutant proteins turn over with half lives of 1 to 3 minutes. Higher levels of transcription do not prevent degradation of these mutant proteins. When mutant *cro* genes encoding the Asp47→Gly and Arg38→Leu substitutions were cloned under control of the highly efficient *tac*-promoter (25), the levels of the mutant proteins relative to wild-type were not increased.

Discussion:

In principle, there are three general factors that could contribute to the inability of any given mutant Cro protein to bind operator DNA. The folded form of the mutant protein could have decreased affinity for the operator; the mutant protein might be inactive because it is unfolded; or, the mutant protein might be rapidly degraded by cellular proteases. Our data do not address the first possibilities directly but do indicate that proteolysis plays an important role in the phenotypic defects displayed by most of the Cro$^-$ missense mutants. This raises several questions. Which cellular proteases degrade the mutant Cro proteins? What determines whether a given Cro mutant will be hypersensitive or relatively resistant to degradation? How do missense proteins in other systems avoid cellular proteolysis? In the discussion that follows, we consider possible answers to these questions.

It is widely recognized that nonsense fragments, missense mutants, and other abnormal proteins can be sensitive to proteolysis *in vivo* (26-31). Moreover, many of these proteins are stabilized in strains of *E. coli* bearing *lon*$^-$, *htp*R$^-$, or *hfl*$^-$ mutations (32-34). However, as judged by pulse-chase experiments, these cellular mutations do not prevent rapid turn-over of the mutant Cro proteins data not shown). These data suggest

that the mutant Cro proteins are degraded by a proteolytic system, which has not yet been characterized genetically. We are currently attempting to obtain cellular mutants in which the mutant Cro proteins are not degraded (D. Parsell, A.P, D. Botstein & R.T.S., unpublished).

What properties of the mutant Cro proteins make them susceptible to proteolysis? In our view, protein stability is the most likely determinant of proteolytic sensitivity, since denatured proteins, but not their folded counterparts, are good substrates for most proteases. By this model, the mutant Cro proteins with the lowest thermodynamic stabilities should be the ones that are most sensitive to proteolysis. As we argue below, this appears to be the case.

Based on their positions in the wild type structure (1,2), about two-thirds of the missense substitutions would be expected to reduce Cro stability. Thirty-six of the forty mutants in this "stability" class have intracellular levels between 0 and 5% of wild type, and three have levels between 5 and 10% of wild type. Most "stability" mutations (30 of 40) affect residues in the tightly packed protein interior. Here, some substitutions (e.g. Phe14→Leu) would remove favorable packing and hydrophobic interactions, whereas others (e.g. Ala33→Val) probably introduce unfavorable interactions. Some of the "stability" substitutions (5 of 40) disrupt charge stabilized hydrogen bonds on the surface of Cro; for example, the Asp47→Tyr and Ser49→Asn mutations affect the hydrogen bond that links the wild type side chains at these positions. For Gly48→Ala and four similar substitutions, stability appears to be reduced because a glycine in a tight-turn has been replaced with a residue that cannot readily assume the positive ϕ, ψ dihedral angles required for the turn (35).

Most of the "stability" mutants have levels that are too low to determine whether the mutant protein has significant activity. However, some proteins in the "stability" class seem to have specific activities that are near wild type. For example, strains containing the Asp47→Gly protein have CRM levels of 1% and β-galactosidase levels of 1239 units, while strains containing wild type Cro expressed from the z4 mutant promoter have CRM levels of 1.5% and β-gal levels of 1292 units. Other mutant proteins with specific activities near wild type are Asp9→Tyr, Leu23→Phe, Gly48→Glu, and Glu53→Lys.

About one-third (22/62) of the Cro^- mutations affect DNA-binding residues in the α2 and α3 helices, and the C-terminal β-region of Cro (1,2). These mutations could decrease operator affinity by removing favorable interactions with the DNA or introducing unfavorable interactions. For example, mutations in or near the α3 recognition helix (e.g. Tyr26→Asp, Gln27→His, Ser28→Asn, Lys32→Thr, Arg38→Gln) change residues that have been proposed to contact bases in the major groove of the operator site. Other mutations (e.g. Gln16→His in α helix 2 and Lys56→Gln in the C-terminal β-region) probably disrupt hydrogen bonds or salt bridges with the phosphate oxygens of the DNA backbone.

We will refer to the mutations described immediately above as "DNA binding" mutations. The mutant proteins in this class appear to have reduced affinity for operator DNA since their specific activities *in vivo* are lower than that of wild type Cro. The "DNA binding" mutant proteins are also present at reduced intracellular levels, but their levels are generally higher than those of the "stability" mutants. Most (18/22) of the "DNA binding" mutants have levels between 5 and 60% of wild type whereas only 4/40 of the "stability" mutants have levels this high. Why are the levels of the "DNA binding" mutant proteins reduced at all? There are several possibilities: *(i)* These proteins may be slightly less stable

than wild type. If Cro were marginally stable, even small changes in stability could increase its turn-over rate. *(ii)* If Cro were less susceptible to proteolysis when bound to DNA, then mutations that decreased DNA binding could lead to increased proteolysis. Non-operator binding would be an essential component of such a mechanism since Cro bound to operator DNA represents an insignificant fraction of the total Cro in the cell. *(iii)* It is conceivable that wild type Cro preferentially inhibits the expression of cellular genes whose products participate in protein degradation. In such a case, the mutant strains would have increased proteolytic levels or activities that could be responsible for the increased degradation of the mutant proteins.

Missense mutants of different proteins show radically different patterns of sensitivity to intracellular proteolysis. Almost all of the λ Cro mutants are degraded to some extent. However, the λ repressor mutants studied previously in this laboratory were not degraded (36-38). To account for these differences and the known properties of many of the mutant proteins, at least three classes of missense mutation must be postulated. The first class includes mutations that do not have significant effects on the thermodynamic stability of the protein. About half of the λ repressor mutants have this property. These missense proteins are protease resistant because they are stably folded. The second class includes mutations which destabilize the native structure and thereby increase the fraction of protein molecules in an unfolded, protease-sensitive state. This model seems reasonable for most of the λ Cro mutations since there is a good correlation between substitutions that would be expected to destabilize the protein and mutant proteins whose intracellular levels are severely reduced. The third class also includes mutant proteins with reduced thermodynamic stability but in this case the unfolded chains escape proteolysis. Many of the λ repressor mutants fall in this category. Since the unfolded chains of these mutant

repressors aggregate during denaturation *in vitro* (38), it is possible that a similar process of aggregation accounts for their protease resistance in the cell.

Acknowledgements: We thank Deborah Lucas for technical assistance; Tom Roberts, Michael Berman, Susan Gottesman, Hans Cheng, Carol Gross and Brian Seed for providing plasmids and strains; Sandy Johnson for antiserum; and Brian Matthews for helpful discussion and for providing the coordinates of the λ Cro structure. This work was supported by NIH grant AI-16892. AP was supported by an NIH predoctoral training grant and VY was supported by a grant from the M.I.T. UROP program.

References:

(1) Anderson, W.F., Ohlendorf, D.H., Takeda, Y. & Matthews, B.W. (1981) Nature 290, 754-758.

(2) Ohlendorf, D. H., Anderson, W.F., Fisher, R.G., Takeda, Y. & Matthews, B.W. (1982) Nature 298, 718-723.

(3) Johnson, A., Meyer, B.J. & Ptashne, M. (1978) Proc. Natl. Acad. Sci. USA 75, 1783-1787.

(4) Roberts, T.M., Kacich, R. & Ptashne, M. (1979) Proc. Natl. Acad. Sci. USA 76, 760-764.

(5) Boyer, H.W., Betlach, M., Bolivar, F., Rodriguez, R.L., Heyneker, H.L., Shine, J. & Goodman, H.M. (1977) in Recombinant Molecules: Impact on Science and Society, (eds., R.F. Beers, Jr. & E.G. Basser), Raven, New York, pp 9-20.

(6) Zagursky, R.J. & Berman, M.L. (1984) Gene 27, 183-191.

(7) Meyer, B. J., Maurer, R. & Ptashne, M. (1980) J. Mol. Biol. 139, 163-194.

(8) Messing, J. & Vieira, J. (1982) Gene 19, 269-276.

(9) Backman, K. & Ptashne, M. (1978) Cell 13, 65-71.

(10) Messing, J., Crea, R. & Seeburg, P.H. (1981) Nucl. Acid Res. 9, 309-321.

(11) Low, K.B. (1972) Bacteriol. Rev. 36, 587-607.

(12) Miller, J.H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor, New York.

(13) Davis, R.W., Botstein, D. & Roth, J.R. (1980) in A Manual for Genetic Engineering, Advanced Bacterial Genetics, Cold Spring Harbor Press, New York.

(14) Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor.

(15) Levinson, A., Silver, D. & Seed, B. (1984) J. Mol. Appl. Genet 2, 507-517.

(16) Sanger, F., Nicklens, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.

(17) Johnson, A. D. (1980) Ph.D. Thesis, Harvard University.

(18) Ito, K., Date, T., and Wickner, W., 1980. J. Biol. Chem. 255: 2123-2130.

(19) Tsang, V.C.W.,Peralta, J.M. and Simons, A.R. (1983) in Methods in Enzymology 92, 377-391, Academic Press, New York.

(20) Freese, E., Bautz, E. & Freese, E. B. (1961) Proc. Natl. Acad. Sci. USA 47, 845-855.

(21) Miller, J. F. (1983) Ann. Rev. Genet. 17, 215-238.

(22) Cox, E. C. (1976) Ann. Rev. Genet. 10, 135-156.

(23) Schaaper, R. M. & Loeb, L. A. (1981) Proc. Natl. Acad. Sci. USA 78, 1773-1777.

(24) Caruthers, M. H., Barone, A. D., Beltman, J., Bracco, L. P., Dodds, D. R., Dubendorff, J. W., Eisenbeis, S. J., Gayle, R. B., Prosser, K., Rosendahl, M. S., Sutton, J. & Tang, J.-Y. (1985) J. Cell. Biochem., in press.

(25) Amman, E., Brosius, J., & Ptashne, M. (1983) Gene 25, 167-178.

(26) Goldberg, A. L. & St. John, A. C. (1976) Ann. Rev. Biochem. 45, 747-803.

(27) Goldberg, A. L., & Dice, J. F. (1974) Ann. Rev. Biochem. 43, 835-869.

(28) Schlotmann, M. & Beyreuther, K. (1979) Eur. J. Biochem. 95, 39-49.

(29) Zipser, D. & Bhavsar, P. (1976) J. Bact. 127, 1538-1542.

(30) Gilbert, H. J. & Drabble, W. T. (1980) J. Gen. Microbiol. 117, 33-45.

(31) Greeb, J., Atkins, J. F., & Loper, J. C. (1971) J. Bact. 106, 421-431.

(32) Gottesman, S. & Zipser, D. (1979) J. Bacteriol. 133, 844-851.

(33) Baker, T. A., Grossman, A. D., and Gross, C. A. (1984) Proc. Natl. Acad. Sci. USA 81, 6779-6783.

(34) Hoyt, M.A., Knight, D.M., Das, A., Miller, H. I., and Echols, H. (1982) Cell 31, 565-573.

(35) Ramachandran, G.N., Venkatachalam, C.M., and Krimm, S. (1966) Biophys. J. 6, 849-872.

(36) Hecht, M. H., Nelson, H.C.M., & Sauer, R. T. (1983) Proc. Natl. Acad. Sci USA 74, 2676-2680.

(37) Nelson, H. C. M., Hecht, M. H. & Sauer, R. T. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 441-450.

(38) Hecht, M. H., Sturtevant, J. M. & Sauer, R. T. (1984) Proc. Natl. Acad. Sci USA 81, 5685-5689.

(39) Hsiang, M. W., Cole, R. D., Takeda, Y. & Echols, H. (1977) Nature 270, 275-277.

(40) Roberts, T. M., Shimatake, H., Brady, D., Rosenberg, M. (1977) Nature 270, 274-275.

Fig. 1. DNA sequence of the promoter region and 5' non-coding sequences. Mutations and their allele names are shown below the wild type sequence. The numbering is relative to the start point of translation. The presumptive start point of transcription is denoted by an arrow. SD Cro is the ribosome binding site of the λ *cro* gene.

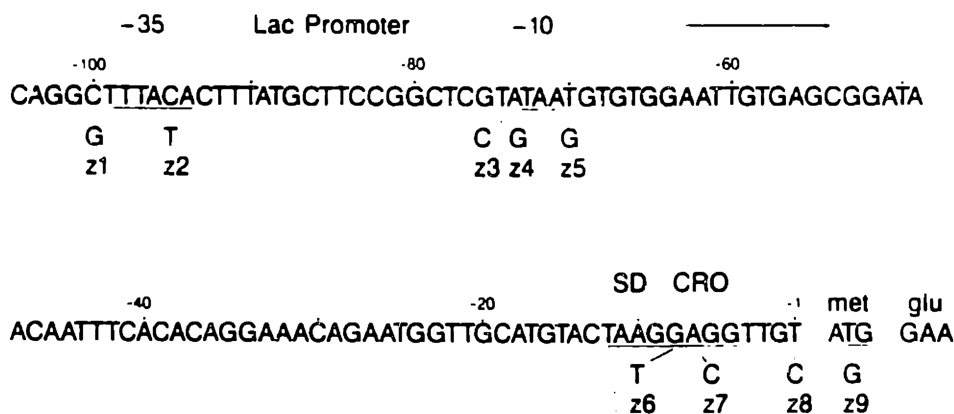
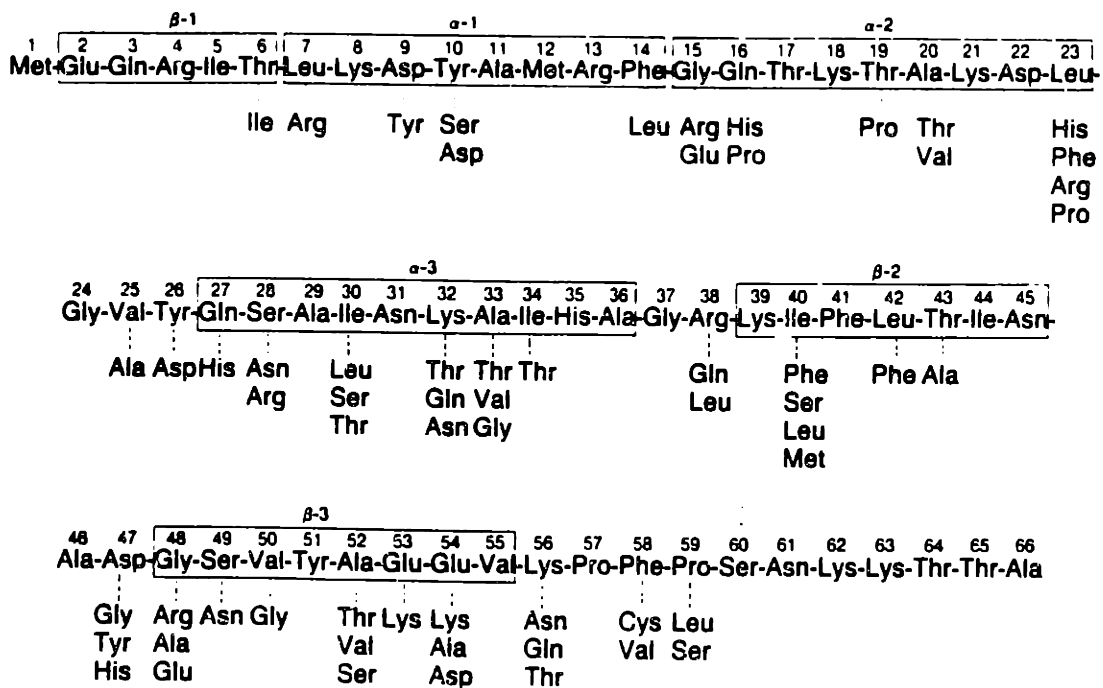
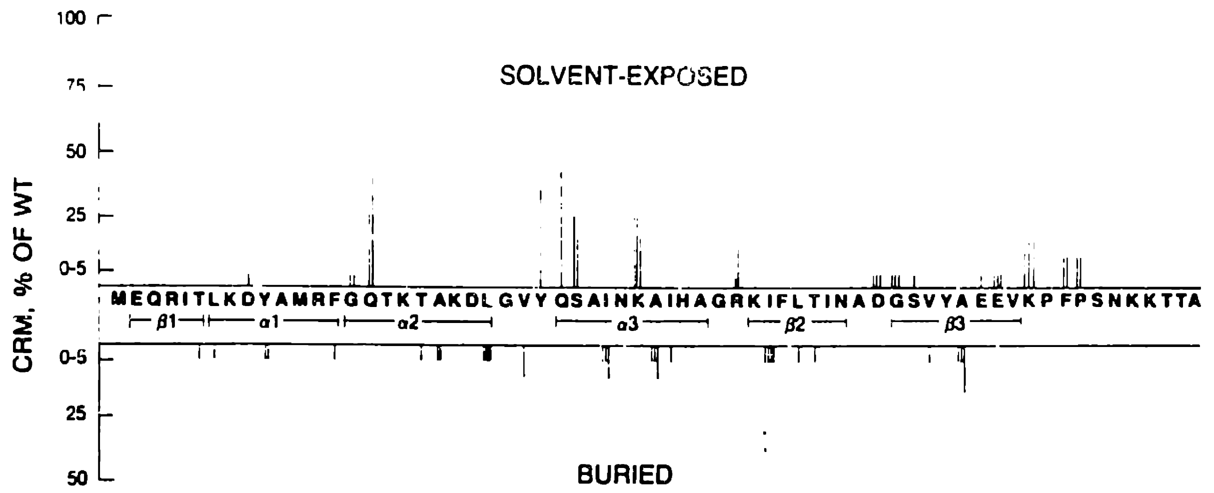

-35
Lac Promoter
-10
-100
-80
-60
CAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATA
G
T
C
G
G
z1
z2
z3
z4
z5
SD
CRO
-40
-20
-1
met
glu
ACAATTTCACACAGGAAACAGAATGGTTGCATGTACTAAGGAGGTTGT
ATG
GAA
T
C
C
G
z6
z7
z8
z9

Fig. 2. Protein sequence of wild type Cro (39,40) and positions of mutant amino acid substitutions. α-helix and β-sheet regions in wild type Cro (1) are boxed.



β-1
α-1
α-2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Met-Glu-Gln-Arg-Ile-Thr-Leu-Lys-Asp-Tyr-Ala-Met-Arg-Phe-Gly-Gln-Thr-Lys-Thr-Ala-Lys-Asp-Leu-
Ile Arg Tyr Ser Leu Arg His Pro Thr His
Asp Glu Pro Val Phe
Arg
Pro
α-3
β-2
24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
Gly-Val-Tyr-Gln-Ser-Ala-Ile-Asn-Lys-Ala-Ile-His-Ala-Gly-Arg-Lys-Ile-Phe-Leu-Thr-Ile-Asn-
Ala Asp His Asn Leu Thr Thr Thr Gln Phe Phe Ala
Arg Ser Gln Val Leu Ser
Thr Asn Gly Leu
Met
β-3
46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66
Ala-Asp-Gly-Ser-Val-Tyr-Ala-Glu-Glu-Val-Lys-Pro-Phe-Pro-Ser-Asn-Lys-Lys-Thr-Thr-Ala
Gly Arg Asn Gly Thr Lys Lys Asn Cys Leu
Tyr Ala Val Ala Gln Val Ser
His Glu Ser Asp Thr

Fig. 3. CRM levels of missense mutant proteins. Wild type residues are indicated by the single letter amino acid code. CRM levels for substitutions affecting solvent exposed residues are shown above and those affecting buried residues below the sequence.



SOLVENT-EXPOSED
BURIED
CRM, % OF WT
100
75
50
25
0-5
0-5
25
50
MEQRITLKDYAMRFGQTKTAKDLGVYQSAINKAIHAGRKIFLTINADGSVYAEEVKPFPSNKKTTA
β1
α1
α2
α3
β2
β3

Fig. 4. Pulse-chase experiments. Cells were labeled with ^{35}S-Met for 0.5 minutes. The time of the unlabeled methionine chase (in minutes) is indicated below each lane. Lane C shows a 0.5 minute chase for a control strain carrying pBR322 instead of a Cro plasmid.

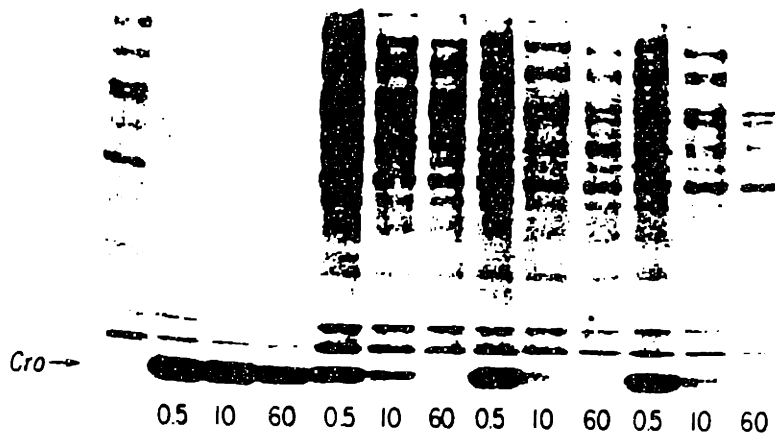

C
Wild Type
Asp 9 – Tyr
Asp 47 – Gly
Arg38 – Leu
Cro
0.5
10
60
0.5
10
60
0.5
10
60
0.5
10
60

Table 1. The number of isolates of a particular mutation obtained with each mutagen is superscripted. CRM levels for mutant Cro proteins are expressed as percentage of wild type levels of Cro. A, 2-aminopurine; H, hydroxylamine; N, MNNG; M, MNNG-mutagenized packaged single stranded plasmid introduced into tester strain via transduction; T, MutT; D, MutD; S, spontaneous mutagenesis arising during growth of the m13 origin plasmid pAP104 as single stranded DNA; N.D., not determined; TAA, ochre termination codon; TGA, opal termination codon.

Table 1. λ Cro Mutations

Allele	Amino Acid Change	Mutagens	Base Change	β-Gal activ.	CRM % WT
WT	-	-	-	38	100
z1	-	A^{1}	C→G(-100)	106	7
z2	-	$H^{2}M^{2}$	C→T(-95)	797	3
z3	-	A^{1}	T→C(-75)	86	13
z4	-	T^{1}	T→G(-75)	1292	1
z5	-	T^{1}	T→G(-70)	2162	<1
z6	-	A^{1}	G→T(-8)	1085	2
z7	-	T^{1}	A→C(-7)	132	6
z8	-	A^{1}	T→C(-1)	45	11
z9	-	T^{1}	T→G(2)	2252	<1
Qoc3	Gln^{3}→TAA	H^{1}	C→T(7)	2278	<1
f1	Frameshift	T^{1}	+T(9)	N.D.	N.D.
TI6	Thr^{6}→Ile	$H^{1}N^{4}$	C→T(17)	1563	2
LR7	Leu^{7}→Arg	T^{1}	T→G(20)	2408	<2
DY9	Asp^{9}→Tyr	S^{1}	G→T(25)	1238	3
YD10	Tyr^{10}→Asp	T^{2}	T→G(28)	2378	<1
YS10	Tyr^{10}→Ser	T^{1}	A→C(29)	2385	<1
FL14	Phe^{14}→Leu	T^{1}	T→G(42)	207	2
GR15	Gly^{15}→Arg	$D^{1}H^{1}$	G→A(43)	1490	1
GE15	Gly^{15}→Glu	N^{3}	G→A(44)	2169	<1
Qoc16	Gln^{16}→TAA	H^{1}	C→T(46)	2091	<1
QP16	Gln^{16}→Pro	T^{1}	A→C(47)	1774	41
QH16	Gln^{16}→His	T^{2}	A→C(48)	2022	28

TP19	Thr^{19}→Pro	T^{1}	A→C(55)	2127	<1
AT20	Ala^{20}→Thr	D^{1}H^{1}N^{2}	G→A(58)	2529	3
AV20	Ala^{20}→Val	A^{1}H^{3}N^{3}	C→T(59)	2245	<1
LF23	Leu^{23}→Phe	N^{5}	C→T(67)	509	2
LP23	Leu^{23}→Pro	A^{1}	T→C(68)	2366	<1
LR23	Leu^{23}→Arg	T^{1}	T→G(68)	2353	<1
LH23	Leu^{23}→His	D^{1}	T→A(68)	2125	3
VA25	Val^{25}→Ala	M^{1}	T→C(74)	937	9
YD26	Tyr^{26}→Asp	T^{1}	T→G(76)	2018	54
Qoc27	Gln^{27}→TAA	H^{1}	C→T(79)	2090	<1
QH27	Gln^{27}→His	T^{2}	A→C(81)	1934	44
SR28	Ser^{28}→Arg	T^{2}	A→C(82)	2246	13
SN28	Ser^{28}→Asn	N^{1}	G→A(83)	1860	25
IL30	Ile^{30}→Leu	T^{1}	A→C(88)	1203	9
IT30	Ile^{30}→Thr	A^{1}	T→C(89)	2384	<1
IS30	Ile^{30}→Ser	T^{2}	T→G(89)	2371	<2
KQ32	Lys^{32}→Gln	T^{1}	A→C(94)	1194	42
KT32	Lys^{32}→Thr	T^{3}	A→C(95)	2158	22
KN32	Lys^{32}→Asn	S^{2}	G→T(96)	1089	12
AT33	Ala^{33}→Thr	H^{2}N^{5}	G→A(97)	2231	<4
AV33	Ala^{33}→Val	A^{1}N^{4}	C→T(98)	2252	<1
AG33	Ala^{33}→Gly	D^{1}T^{1}	C→G(98)	2039	6
IT34	Ile^{34}→Thr	A^{1}	T→C(101)	1972	<2
Rop38	Arg^{38}→TGA	D^{1}N^{2}	C→T(112)	2017	<1
RL38	Arg^{38}→Leu	S^{1}	G→T(113)	2212	2

RQ38	Arg^{38}→Gln	N^{1}	G→A(113)	2298	18
IF40	Ile^{40}→Phe	D^{3}	A→T(118)	2247	<4
IL40	Ile^{40}→Leu	T^{3}	A→C(118)	2282	<4
f2	Frameshift	$D^{2}T^{1}$	-T(119)	2157	<1
f3	Frameshift	D^{2}	+T(119)	2018	<1
IS40	Ile^{40}→Ser	T^{1}	T→G(119)	2503	<5
IM40	Ile^{40}→Met	T^{2}	T→G(120)	2376	2
Lop42	Leu^{42}→TGA	T^{1}	T→G(125)	2403	<1
LF42	Leu^{42}→Phe	T^{7}	A→C(126)	1542	<2
LF42'	Leu^{42}→Phe	T^{1}	A→T(126)	N.D.	N.D.
TA43	Thr^{43}→Ala	A^{1}	A→G(127)	145	2
DY47	Asp^{47}→Tyr	S^{2}	G→T(139)	2037	<1
DH47	Asp^{47}→His	M^{1}	G→C(139)	1584	5
DG47	Asp^{47}→Gly	N^{3}	A→G(140)	1239	1
Gop48	Gly^{48}→TGA	S^{1}	G→T(142)	2229	<1
GR48	Gly^{48}→Arg	$D^{1}M^{3}N^{1}$	G→A(142)	1072	<1
GE48	Gly^{48}→Glu	N^{8}	G→A(143)	531	1
GA48	Gly^{48}→Ala	M^{2}	G→C(143)	1126	5
SN49	Ser^{49}→Asn	M^{2}	G→A(146)	1450	<7
VG50	Val^{50}→Gly	T^{1}	T→G(149)	2384	<2
AT52	Ala^{52}→Thr	$H^{1}N^{1}$	G→A(154)	2177	<4
AS52	Ala^{52}→Ser	M^{1}	G→T(154)	1798	2
AV52	Ala^{52}→Val	$D^{1}N^{1}$	C→T(155)	2107	12
EK53	Glu^{53}→Lys	M^{1}	G→A(157)	398	5
EK54	Glu^{54}→Lys	$D^{1}H^{1}M^{1}N^{1}$	G→A(160)	2216	2

EA54	Glu^{54}→Ala	T^{1}	A→C(161)	2329	2
ED54	Glu^{54}→Asp	S^{1}	G→C(162)	2220	<2
KQ56	Lys^{56}→Gln	T^{2}	A→C(166)	2269	15
KT56	Lys^{56}→Thr	T^{5}	A→C(167)	2188	41
KN56	Lys^{56}→Asn	D^{1}	G→C(168)	2117	29
KN56'	Lys^{56}→Asn	T^{1}	G→T(168)	N.D.	N.D.
FV58	Phe^{58}→Val	T^{3}	T→G(172)	2389	10
FC58	Phe^{58}→Cys	T^{3}	T→G(173)	2250	8
PS59	Pro^{59}→Ser	N^{2}	C→T(175)	2230	6
PL59	Pro^{59}→Leu	D^{1}	C→T(176)	2076	6

Chapter III

Amino Acid Substitutions that Increase the Thermal Stability of the λ Cro Protein

Introduction

The forces that contribute to protein stability have been understood in a general way for some time. Hydrophobic interactions, electrostatic interactions, van der Waals interactions and hydrogen bonds all play important roles in stabilizing proteins, but it is still extremely difficult to assess the quantitative importance of specific interactions within the context of a folded protein. As a result, it has usually not been possible to predict accurately, from inspection of a protein structure, the effect of any particular amino acid substitution. This inability clearly presents a major obstacle to achieving the goals of rational protein modification and design. To examine the effects of residue substitutions on protein structure and stability, several groups have used genetic analysis to study proteins with known structures. One approach has been to use random mutagenesis to create a large pool of mutant genes, and a genetic screen or selection to identify individual amino acid changes that cause loss of activity *in vivo*. Among the mutants identified in this way are those that reduce protein stability. This class of substitutions may identify residues that are important for determining and/or stabilizing protein structure (1-6). With such mutations in hand, reversion analysis can then be used to identify intragenic mutations that restore activity. These may provide information about the nature of the mutant defect or identify alternative means by which protein stability can be improved (3,7-9). The advantage of using random mutagenesis for this type of analysis is that large numbers of amino acid changes can be surveyed in a rapid and relatively unbiased way, and interesting, useful, or surprising substitutions may be chosen for further analysis.

We have been studying the bacteriophage λ Cro protein, a small, sequence-specific DNA binding protein which acts as a dimer (10, 11). Its crystal structure has been determined (12) and the structure of its complex with operator DNA has been predicted based on model building studies (13). In previous studies, we isolated a large collection of

randomly generated Cro-defective mutations and observed that substitutions of structurally important residues, such as those in the hydrophobic interior of Cro, generally resulted in proteins that were rapidly degraded *in vivo* (5). This observation led us to suggest that reduced stability of the folded structure of Cro was the cause of the increased proteolytic susceptibility of this class of mutants. Thus, we suspected that the rate of intracellular turnover could serve as a convenient indicator of thermal stability, and that substitutions that increase thermal stability could be isolated as second site suppressors of mutations that cause proteolytic sensitivity.

In this paper, we describe the isolation of intragenic mutations that suppress the rapid degradation of an unstable mutant Cro protein. Among these second-site mutations are two substitutions that dramatically increase the thermal stability of the revertant proteins. The extent to which these substitutions stabilize Cro is somewhat surprising since they affect solvent-exposed residues, which as a class are usually thought to play only minor roles in the maintenance of protein structure.

Materials and Methods:

Plasmids and Bacteria: *Escherichia coli* strain X90 (15) bears an F' episome carrying the Lac repressor-overproducing allele *lacI*^{Q}. Strain X9T was derived from X90 by selection for resistance to a bacteriophage (probably T1) that contaminated the laboratory fermenters (14). *E. coli* strain DB4729 is *endo*I$^-$, r_K^-, m_K^+, *thi*$^-$, *su2*$^+$, Δ(*srl-recA*) (obtained from D. Botstein). Plasmid pAP104 (5) is a pBR322 derived plasmid that carries the λ *cro* gene under *lac* promoter control, confers resistance to ampicillin and tetracycline, and contains both an M13 and colE1 origin of replication. The presence of the M13 origin

causes single stranded plasmid DNA to be made and packaged upon infection with an M13 helper phage, such as RV1. Plasmid pAP5M is a derivative of pAP104, in which the *lac* promoter has been replaced by the *tac* promoter as described (15). Plasmid pAP114 is a derivative of pAP5M in which the sequences between the *EcoRI* and *Nde*I sites of pBR322 were deleted and the *cro* sequences from the *Bgl*II site to the end of the gene have been replaced by the sequences from the synthetic gene shown in Fig 1. Plasmid pAP119 was derived from pAP114 by introducing the *Nco*I site shown in Fig 1. This was accomplished using a mismatch primer and the oligonucleotide directed mutagenesis kit from Amersham. The synthetic *cro* gene was constructed using automated DNA synthesis and standard cloning procedures. This gene encodes the wild type amino acid sequence and incorporates a number of restriction enzyme recognition sites which simplify recombination *in vitro* and cassette mutagenesis.

Affinity Purification of Antibody: A λ Cro-Sepharose column was prepared using CNBr-activated Sepharose 4B essentially as described by the supplier (Pharmacia). Crude rabbit serum containing antibodies raised against wild type Cro was loaded onto the column, which was washed with PBS (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4) to remove unbound serum proteins. Bound antibodies were eluted with 3 M KSCN and, following dialysis, were stored in PBS with 3 mM NaN_3.

Mutagenesis and Mutant Screen: To isolate intragenic mutations that suppress the rapid degradation of an unstable Cro mutant, we used antibodies to identify colonies in which the Cro antigen was present at higher steady state concentrations than in the original mutant colonies. Strain DB4729 carrying plasmid pAP104 bearing the IL30 *cro* gene (5) shows little or no antibody reactivity because the the mutant IL30 protein is rapidly degraded and thus present at a low concentration. The *cro* gene of this plasmid was mutagenized *in vitro* using a modification of the procedure of Myers and Maniatis (16).

Single stranded pAP104 DNA was prepared from transducing particles purified after infection of strain X90/pAP104(IL30) with the phage M13 RV1 (17). The resulting DNA includes approximately equal amounts of single stranded plasmid DNA and helper phage DNA, but the latter does not adversely affect the mutagenesis procedure. Single stranded DNA was next subjected to one of three mutagenic regimens: *(i)* Incubation in 1 M $NaNO_2$, 0.2 M NaOAc [pH 4.3] for 5 h at room temperature; *(ii)* incubation in 0.13 mM $KMNO_4$ for 1 h at room temperature, and; *(iii)* incubation in 25 mM sodium citrate/citric acid buffer [pH 4.8], 100 mM NaCl at 70 °C, for 3 h. Following mutagenesis, DNA was recovered by EtOH precipitation and washed twice with 70% EtOH. An oligonucleotide primer which is complementary to a region 3' to the *cro* gene was then annealed to the single stranded DNA and used to prime second strand synthesis by reverse transcriptase (Bethesda Research Laboratories). Partially double-stranded DNA was recovered by EtOH precipitation and was digested with restriction enzymes *Eco*RI and *Hin*dIII, each of which cuts once in pAP104, to yield a small double stranded fragment that includes the entire *cro* gene and *lac* promoter region. This fragment was isolated by agarose gel electrophoresis and ligated to the unmutagenized backbone fragment of pAP104.

The antibody screening method is a modification of the cDNA library screening procedure described by Helfman *et al.* (18). Mutagenized DNA was introduced into strain DB4729 by transformation and cells were plated on nitrocellulose filters overlaid on LB agar plates containing 100 μg/ml ampicillin. Colonies were then replicated onto a second filter, and regrown for several hours. One filter of each pair was treated with 0.1 N NaOH to lyse the bacteria. After neutralization with 2 M Tris-HCl [pH 7.0], filters were incubated for several hours in TBS buffer (50 mM Tris-HCl [pH 7.5], 150 mM NaCl) containing 5% nonfat dry milk, 0.1 M $MgCl_2$, 1 μg/ml DNAase I, 40 μg/ml hen egg lysozyme, and 3 mM NaN_3. Filters were washed, wiped clean of visible cell debris, and incubated for one hour in TBS containing 5% nonfat dry milk, 0.2% Tween 20, and affinity purified rabbit anti-

Cro serum at a 1:500 dilution. Filters were then washed extensively in TBS with 0.2% Tween-20, incubated for 1 h in TBS containing 5% nonfat dry milk, 0.2% Tween-20, and ^{125}I-Protein A (1:500 dilution; Amersham), and washed again. Bound radioactive protein A was visualized by autoradiography. Antibody reactive colonies were located on the unprocessed duplicate filters by comparison with autoradiograms, and were purified twice by restreaking for single colonies and retesting as in the initial screen.

DNA Sequencing: For each antibody reactive isolate, the sequence of the entire *cro* gene and *lac* promoter region was determined using the dideoxy method (19). Single stranded template DNA was obtained from transducing particles in supernatants of M13 RV1 infected cultures of X90 or X9T carrying the plasmid of interest. In some cases, the Sequenase dideoxy sequencing kit (US Biochemicals) was used.

Placing Genes Under *Tac* Promoter Control and Construction of Single Mutants: The suppressor mutations were isolated in plasmid pAP104, in which transcription of the *cro* gene is directed by the *lac* promoter. To express large quantities of the mutant or revertant proteins for purification, we placed the variant *cro* genes under the control of the stronger *tac* promoter (15). For those suppressor mutations located after the naturally occurring *Bgl*II site in the *cro* gene, we isolated the small *Bgl*II-*Hin*dIII fragment from pAP104 and cloned this in place of the corresponding fragment of pAP5M. This places the IL30 mutation and the suppressor mutation in a *cro* gene under *tac* promoter control. To construct *tac*-promoted genes carrying the suppressor mutations alone, we synthesized the appropriate double stranded cassettes and inserted them into pAP114. For the QL16 suppressor mutation, a cassette was prepared and inserted between the *Nco*I and *Bgl*II sites of pAP119.

Pulse-Chase Experiments: Logarithmic phase cultures of DB4729 carrying plasmids with wild type or mutant *cro* genes were pulse labelled with ^{35}S-methionine (100

μCi) for one minute. Unlabeled methionine was then added to a concentration of 1.4 mg/ml to quench incorporation of the radioactive amino acid. Thereafter, portions were periodically removed, placed on ice, and N-ethylmaleimide was immediately added at a concentration of 2 mM to prevent further proteolysis. Cells were pelleted by centrifugation and lysed in 1% SDS. The lysates were resuspended in immunoprecipitation dilution buffer (1.25% Triton X-100, 190 mM NaCl, 60 mM Tris-HCl [pH 7.5], 6 mM EDTA), insoluble material was removed by centrifugation, and affinity purified rabbit anti-Cro antibody was added in molar excess over Cro. After a 30 min incubation, a suspension of fixed Staph A cells (Calbiochem) was added, the mixture was incubated for an additional 30 min, and the cells and bound antibody and Cro were collected by centrifugation. The pellet was resuspended in wash buffer (50 mM Tris-HCl [pH 7.5], 5 mM EDTA, 150 mM NaCl) and centrifuged again. Labelled proteins were stripped from the Staph A cells by heating at 90 °C for 3 min in electrophoresis loading buffer (0.125 M Tris-HCl [pH 6.8], 4% SDS, 20% glycerol, 0.14 M β–mercaptoethanol). Immunoprecipitated proteins were separated by electrophoresis in a 20% polyacrylamide gel containing SDS and urea (20). Following treatment with Autofluor autoradiographic intensifier (National Diagnostics), the gel was dried and the amount of ^{35}S-labeled Cro was quantified by autoradiography and densitometry

Protein Purification: Cultures of strains X90 or X9T carrying plasmids pAP5M, pAP114, or pAP119 were grown at 37 °C in 2 liters of LB broth with 100 μg/ml ampicillin to an optical density of 1.5 at 600 nm. Expression of the wild type or mutant Cro protein from the *tac* promoter was induced by the addition of 0.01% IPTG (w/v). The induced cultures were grown for an additional 3 h, and cells were harvested by centrifugation and lysed by sonication. Cell lysis and all subsequent manipulations were performed at 4 °C. Polyethylenimine was added to a final concentration of 0.6% (v/v) and the precipitate was collected by centrifugation. Proteins were precipitated from the supernatant with 90%

saturated $(NH_4)_2SO_4$, dialyzed into SB50 buffer (50 mM Tris-HCl [pH 7.0], 0.1 mM EDTA, 5% glycerol, 1.4 mM β-mercaptoethanol, 50 mM KCl) and loaded onto a 2.5 x 10 cm DEAE Sephacel column (Pharmacia) equilibrated in SB50. The flow-through from this column was loaded directly onto a 1.5 x 8 cm Cellex P column (Bio-Rad Laboratories) equilibrated in SB50. This column was washed thoroughly and proteins were eluted with a gradient from SB50 to SB plus 1 M KCl. Fractions containing Cro were identified by SDS gel electrophoresis. The peak fractions were pooled, concentrated by precipitation with 90% saturated $(NH_4)_2SO_4$, and chromatographed on a 2.5 x 48 cm column of Sephadex G50 F (Pharmacia).

Thermal Denaturation: Wild type and mutant Cro proteins at a concentration of 120 μg/ml (in 10 mM KPi [pH 7.0], 0.1 M KCl) were heated at a rate of 0.8 °C per minute. Protein unfolding was monitored by recording the changes in ellipticity at 222 nm with a AVIV model 60S circular dichroism (CD) spectropolarimeter. The ellipticity was found to remain constant for long periods when the temperature was held near the T_m, and the melts were found to be essentially reversible.

DNA binding: The gel shift assay was performed essentially as described (21,22) using a radiolabelled fragment from pKB252 (23) generated by digestion with *Bgl*II and *Nsi*I. This fragment contains the λ O_R1,O_R2 and O_R3 operators. The O_R3 site is the strongest binding site for Cro (11). Binding assays were performed at room temperature in a buffer consisting of 10 mM Tris-HCl [pH 7.0], 200 mM potassium glutamate, 1 mM EDTA, 100 μg/ml bovine serum albumin, and 100 μg/ml sonicated salmon sperm DNA.

RESULTS:

Isolation of suppressor mutations: Our aim, in this study, was to identify amino acid changes that increased the thermal stability of the λ Cro protein. One might

expect that such changes would arise in a genetic screen for second site suppressor mutations that restore activity to a mutant protein with reduced stability. However, the revertant mutations that we isolated in this way appeared to restore activity by improving DNA binding without increasing thermal stability (unpublished results). Enhanced activity mutations of this type had previously been isolated and characterized in λ repressor (24). As a result, we decided to attempt to isolate stabilizing substitutions based on the anticipation that such substitutions would suppress the intracellular proteolytic sensitivity of an unstable mutant Cro protein. To isolate such mutations, a mutant *cro* gene was subjected to random mutagenesis, and colonies containing increased intracellular Cro levels were identified using an antibody screen (see Methods). A variant Cro protein bearing the IL30 (Ile^{30}→Leu) substitution was used for these studies. The destabilizing effect of this amino acid change is presumably due to disrupted packing interactions, since the Ile^{30} side chain forms part of the hydrophobic core of Cro (12). The IL30 Cro protein is present in the cell at about 10% of the wild type level (5), and is both less thermally stable than wild type Cro *in vitro* and degraded more rapidly than wild type Cro *in vivo* (see below).

Locations of amino acid substitutions. The *cro* genes and *lac* promoter regions of fourteen independent revertant candidates were analyzed by DNA sequencing. Each of these isolates retained the IL30 mutation and had acquired additional sequence changes as shown in Table 1. In all, we found four revertants with changes in upstream sequences; nine genes encoding single second-site mutations in the *cro* gene; and two genes encoding second and third site changes in the *cro* gene. We presumed that base changes in the regions upstream from the *cro* gene increase expression of the IL30 Cro protein and did not pursue their study. Similarly, we did not study the isolates containing second and third site revertant mutations further. In the nine genes containing single second-site changes, there were five different reverting substitutions.

The identity of the five second-site amino acid substitutions and their locations in the three-dimensional structure of the Cro dimer are shown in Fig. 2. The AV11 suppressor mutation ($Ala^{11} \rightarrow Val$) changes a residue that is buried in the wild type crystal structure; the remaining suppressors change solvent exposed side chains. Inspection of the Cro structure indicates that the side chain of a valine at position 11 could contact the mutant leucine side-chain at position 30 (12). Hence, the AV11 substitution might correct a packing defect caused by the IL30 substitution. In this case, the effect of the AV11 change on stability would be expected to differ, depending on the identity of the residue present at position 30; in genetic terms, the suppression would be allele specific.

Each of the remaining four suppressor mutations affect surface residues at the end of an α-helix. The QL16 change ($Gln^{16} \rightarrow Leu$) is located at the beginning of α-helix 2; the YC26 and QP27 changes ($Tyr^{26} \rightarrow Cys$ and $Gln^{27} \rightarrow Pro$, respectively) are located at the N-terminal end of α-helix 3; and the AS36 change ($Ala^{36} \rightarrow Ser$) is located near the C-terminal end of α-helix 3. Moreover, three of these substitutions replace residues that are believed to make DNA-binding contacts (13). The Gln^{16} side chain is thought to contact the sugar-phosphate backbone of the operator, while the side chains of Tyr^{26} and Gln^{27} are thought to make sequence specific major groove contacts.

Intracellular half-lives: The rates of intracellular degradation for wild type Cro, IL30 Cro, and each revertant protein were measured by pulse-chase experiments. These data are shown in Fig 3; the half-life of each protein is listed in Table 2. Each revertant protein has a substantially longer half-life than the IL30 parent. Moreover, in each case, the half-life of the revertant protein is as long, or longer, than that of wild type Cro. For example, the QP27/IL30 protein shows essentially no turnover during the 3 h chase period of this experiment, whereas the half-lives of the IL30 and wild type proteins are approximately 11 and 63 minutes, respectively.

Protein Purification: To examine the effects of the suppressor mutations on the physical properties of Cro, we purified wild type Cro, the IL30 protein, and several of the revertant proteins. We also sought to purify several proteins bearing suppressor mutations without the IL30 mutation. Genes containing the QL16, YC26, and AS36 alleles in otherwise wild-type backgrounds were constructed in a straightforward manner using cassette mutagenesis (see Methods). However, despite repeated attempts, we were unable to introduce the QP27 change into a wild type *cro* gene; the QP27 allele was recovered, but always in combination with additional changes, predominantly frameshifts. We suggest that this substitution causes Cro to be deleterious to the cell.

In all, seven variant Cro proteins were purified. These included the IL30, IL30/YC26, IL30/QP27, IL30/AS36, QL16, YC26, and AS36 proteins. During purification, each of these proteins behaved essentially like wild type with the exception of the QP27/IL30 double mutant, which tended to precipitate at high concentrations. We were able to obtain 10 to 120 mg quantities of each protein at purity levels of 95% or better.

Thermal Stability: The stabilities of wild type Cro and the purified variants to denaturation were determined by experiments in which thermal unfolding was monitored by the change in CD ellipticity at 222 nm. Fig. 4 shows the resulting melting curves. Table 2 lists the melting temperature (T_m) for each protein , along with the change in T_m brought about by the suppressor mutation (ΔT_m). Wild type Cro is half-denatured at 40 °C, under the conditions used here. The IL30 protein is less thermally stable with a melting temperature 5 °C below the wild type value. Thus, these data are consistent with our original surmise that the cellular instability of the IL30 mutant is caused by thermal instability.

Two of the revertant substitutions - YC26 and QL16 - cause large increases in the thermal stability of Cro. The effect of the YC26 substitution was measured in an otherwise

wild type Cro protein, and in Cro with the IL30 substitution. The increase in T_m does not seem to depend on the residue at position 30; in the IL30 background, the YC26 change causes an 11 °C increase in stability; in the wild type background, a 12 °C increase is observed (Fig. 4 and Table 2). The QL16 substitution also causes a dramatic stabilization of Cro. The QL16 change increases the T_m of wild type Cro by 14 °C; its effect in combination with IL30 was not determined. The QP27 and AS36 changes cause only small increases (3 °C and 1 °C, respectively) in Cro stability.

The increased thermal stability afforded by the YC26 and QL16 substitutions presents a plausible mechanism by which these amino acid changes prevent proteolytic degradation of the IL30 mutant. Specifically, the IL30 mutant is thought to be proteolytically susceptible in the cell because it is unfolded to a greater extent than wild type. By increasing stability, the YC26 or QL16 changes will reduce the fraction of the IL30 protein that is unfolded and, as a consequence, reduce its susceptibility to proteolysis. In accord with this model, a strong relationship between increasing thermal stability and decreasing proteolytic susceptibility *in vivo* has recently been reported for a protein that is related to Cro, the N-terminal domain of λ repressor (25).

It is important to note, however, that we do not observe a direct relationship between stability and proteolytic susceptibility for the mutant proteins. For example, the QP27/IL30 protein, has a longer half-life than that of YC26/IL30, whose melting temperature is 9 degrees higher. Similarly, the IL30/YC26 and wild type proteins have the same half-lives but melting temperatures that differ by 7 degrees. These observations presumably reflect the involvement of properties, other than intrinsic thermal stability, that affect proteolytic sensitivity. These may include binding to non-specific DNA (25) and/or aggregation of the revertant protein. These mechanisms may play some role in suppression

of degradation by QP27 and AS36, as these substitutions cause only small changes in melting temperature.

DNA binding: The antibody screen employed in this study identifies revertant Cro proteins solely by their increased intracellular level and does not specifically require or preclude any effect of the substitutions on DNA binding activity. In fact, three of the five second site suppressor substitutions (QL16, YC26 and QP27) replace residues that have been proposed to contribute to the affinity and/or specificity of the Cro-operator DNA complex (13). We examined the effects of these substitutions on sequence-specific DNA binding activity *in vitro* using the gel mobility shift assay (21, 22). Under the conditions used, half-maximal binding of wild type Cro to O_R3 occurs at a concentration of 1×10^{-10} M. The QL16 and QP27 substitutions reduce the apparent affinity of Cro for operator DNA by 1,500 and 150 fold, respectively (data not shown). These severe effects support the idea that these residues play key roles in operator recognition. The YC26 substitution, on the other hand, has only a small (2-8 fold) effect on binding, suggesting that any contact made between Tyr^{26} and operator DNA is either relatively unimportant or that a comparable contact can be made by Cys^{26}.

One of the mechanisms by which susceptibility to intracellular proteolysis might be reduced is increased non-specific binding affinity. The effects of the QL16, YC26, QP27 and AS36 substitutions on the formation of non-specific DNA binding complexes were investigated using the gel shift assay, but, no significant changes were observed.

Discussion

We have used an antibody screen to identify randomly generated second site mutations that confer increased resistance to intracellular proteolysis upon an unstable Cro

mutant. At least three classes of mechanisms could conceivably account for the increased resistance to proteolysis *in vivo* . (*i*) Decreasing the fraction of Cro that is present in the unfolded state would be expected to increase resistance to proteolysis of the revertant proteins because unfolded proteins are preferentially digested by proteases. This could be accomplished directly by raising the protein's intrinsic thermal stability or indirectly by increasing the binding of the folded protein to bulk cellular DNA. (*ii*) Structural or chemical features of the folded or unfolded protein that serve as recognition signals for one or more proteases may be altered. (*iii*) The formation of protein aggregates or complexes with other cellular components could physically prevent access by proteases.

Two of the revertant substitutions, YC26 and QL16 ($Tyr^{26}{\rightarrow}Cys$ and $Gln^{16}{\rightarrow}Leu$, respectively), increase the thermal stability of Cro to a surprising extent. The side chains of both the wild type Gln^{16} and Tyr^{26} residues are substantially exposed to solvent in the wild type structure (12) and, although such residues are not thought to play major roles in thermal stability, the increases in melting temperature caused by these substitutions are among the largest that have been reported for single amino acid substitutions. Hence, it is apparent that side chains at some surface residue positions can contribute significantly to protein stability. Inspection of the structure of wild type Cro shows that the mutant side chains could be positioned to contact nearby side chains, but it is not clear whether such interactions occur in the mutant proteins, or if they do occur, whether they would be sufficient to account for the observed stabilization. The λ repressor and Cro proteins have a very similar tertiary structure in the DNA-binding regions where Gln^{16} (α-helix 2) and Tyr^{26} (α-helix 3) are located (26). Interestingly, a mutation in λ repressor, $Gln^{33}{\rightarrow}Tyr$, at the position analogous to Gln^{16} of Cro, also increases thermal stability (6). Perhaps a particular structural feature common to this homologous region of the two proteins is responsible for the ability of substitutions at position 16 in Cro and 33 in λ repressor to increase stability.

It seems likely that the changes in stability afforded by the YC26 and QL16 substitutions are responsible for the increased resistance to proteolysis *in vivo* as these substitutions do not appear to have other effects that would reduce protease sensitivity. For example, they do not appear to cause increased DNA binding, or aggregation in solution, even at high concentration. Moreover, these proteins are soluble in crude lysates and show no tendency to copurify with other cellular components. The AS36 and QP27 substitutions do not, however, increase thermal stability dramatically. These mutations may reduce proteolytic sensitivity by one of the other mechanisms mentioned above. The AS36 substitution partially restores the activity of the IL30 protein *in vivo*, suggesting that DNA binding is altered, although this is not apparent in our assay *in vitro*. Aggregation might be involved in the reduced proteolytic susceptibility of IL30/QP27, since this protein displays a tendency to precipitate at high concentration *in vitro*.

The screening method that we have used to isolate the suppressor mutations is unusual because it does not require restoration of activity *in vivo*. A similar approach could be used to isolate suppressor mutations in other unstable proteins when no other convenient phenotype exists. The ability to screen for a phenotype that does not depend on activity is also desirable when investigating issues of protein structure and stability. For example, *cro* mutations like QL16 that increase stability but adversely affect specific DNA binding, would not be identified in an activity-based screen or selection. This may be a particularly important consideration for small proteins that make extensive contacts with large molecules. For Cro, the fraction of surface area that is involved in DNA binding is relatively large, so that sites where stabilizing changes may be made, without affecting DNA binding could be relatively rare. Another type of problem that may be approached using a mutant screen that does not involve activity, is the relationship between protein flexibility and activity. Among the class of mutations that increase a protein's stability, but

reduce its activity, might be those that restrict functionally important motions or conformational shifts.

The use of the random genetic approach and the isolation of suppressor mutations in *cro* has provided a number of interesting and unexpected results for further study. The striking increases in thermal stability conferred by two solvent exposed amino acid substitutions may provide new insight into the ways in which surface residues can influence protein stability. Since substitution of solvent-exposed residues generally does not increase or decrease thermal stability, these exceptional positions would not likely have been chosen for study using directed mutagenesis. Although further investigation of these mutations will require the construction of specific amino acid substitutions at these sites, their identification as important positions demonstrates the utility of the genetic approach in the study of protein structure.

Acknowledgements: We thank Peter Kim for assistance with CD measurements, and Jim Bowie for strain X9T and for helpful discussions.

References

1 Hecht , M.H., Nelson, H.C.M., Sauer, R.T. Mutations in λ repressor's amino-terminal domain: Implications for protein stability and DNA binding. Proc. Natl. Acad. Sci. 80:2676-2680, 1983.

2 Shortle, D., Lin, B. Genetic Analysis of Staphylococcal Nuclease: Identification of Three Intragenic "Global" Suppressors of Nuclease-minus Mutations. Genetics 110:539-555, 1985.

3 Shortle, D., Meeker, A.K. Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation.. Proteins 1:81-89, 1986.

4 Alber, T., Sun, D.P., Nye, J.A., Muchmore, D.C., Matthews, B.W. Temperature-sensitive mutations of bacteriophage T4 lysozyme occur at sites with low mobility and low solvent accessibility in the folded protein.. Biochemistry 26:3754-3758, 1987.

5 Pakula, A.A., Young, V.B., Sauer, R.T. Bacteriophage λ Cro mutations: Effects on activity and intracellular degradation. Proc. Natl. Acad. Sci. 83:8829-8833, 1986.

6 Hecht, M.H., Sturtevant, J.M., Sauer, R.T. Effect of single amino acid replacements on the thermal stability of the NH_2-terminal domain of λ repressor. Proc. Natl. Acad. Sci. 81:5685-5689, 1984.

7 Yanofsky, C. Tryptophan biosynthesis in *Escherichia coli*: Genetic determination of the protein involved. J. Amer. Med. Assoc. 218:1026-1038, 1971.

8 Hecht, M. H.,Sauer R. T. Phage lambda repressor revertants. Amino acid substitutions that restore activity to mutant proteins. J. Mol. Biol. 186:53-63, 1985.

9 Hecht, M. H., Hehir K. M., Nelson H. C., Sturtevant, J. M, Sauer, R. T. Increasing and decreasing protein stability: effects of revertant substitutions on the thermal denaturation of phage lambda repressor. J. Cell. Biochem., 29:217-24, 1985.

10 Folkmanis A ; Takeda Y ; Simuth J ; Gussin G ; Echols H. Purification and properties of a DNA-binding protein with characteristics expected for the Cro protein of

bacteriophage lambda, a repressor essential for lytic growth. Proc. Natl. Acad. Sci. 73:2249-53, 1976.

11 Johnson, A., Meyer BJ , Ptashne, M. Mechanism of action of the cro protein of bacteriophage lambda. Proc. Natl. Acad. Sci. , 75:1783-7, 1978.

12 Anderson, W.F., Ohlendorf, D.H., Takeda, Y., Matthews, B.W. Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. Nature 290:754-758, 1981.

13 Ohlendorf, D.H., Anderson, W.F., Fisher, R.G., Takeda, Y., Matthews, B.W. The molecular basis of DNA-protein recognition inferred from the structure of cro repressor. Nature 298:718-723, 1982.

14 Bowie, J.U., Sauer, R.T. Identification of C-terminal extensions which protect proteins from intracellular proteolysis. (submitted).

15 Amman, E., Brosius, J., Ptashne, M. Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli.* Gene 25:167-178, 1983.

16 Myers, R, Maniatis, T. A general method for saturation mutagenesis of cloned DNA fragments. Science 229:242-247,1985.

17 Levinson, A., Silver, D. Seed, B. Minimal size plasmids containing an M13 origin for production of single-strand transducing particles.. J. Mol. Appl. Genet. 2:507-517, 1984.

18 Helfman, D. M., Feramisco, J. R., Riddes, J. C., Thomas, G. P., Hughes, S. H. Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression libraryProc. Natl. Acad. Sci.,80:31-35, 1983.

19 Sanger, F., Nicklen, S., Coulson, A.R. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. 74, 5463-5467, 1977.

20 Ito, K., Date, T., Wickner, W. Synthesis, assembly into the cytoplasmic membrane, and proteolytic processing of the precursor of coliphage m13 coat protein. J. Biol. Chem. 255:2123-2130, 1980.

21 Garner, M.M., Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: Application to components of the *Escherichia coli* lactose operon regulatory system. Nucl. Acids Res. 9:3047-3060, 1981.

22 Fried, M., Crothers, D.M. Equilibria and kinetic of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nucl. Acids Res. 9:6505-6525, 1981.

23 Backman, K.C., Ptashne, M. Maximizing gene expression on a plasmid using recombination *in vitro*. Cell 13:65-71, 1978.

24 Nelson, H.C.M., Sauer, R.T. Lambda repressor mutations that increase the affinity and specificity of operator binding. Cell 42:549-558, 1985.

25 Parsell, D.A., Sauer, R.T. The structural stability of a protein is an important determinant of its proteolytic susceptibility in *E. coli*, (submitted).

26 Ohlendorf DH ; Anderson WF ; Lewis M ; Pabo CO ; Matthews BW Comparison of the structures of cro and lambda repressor proteins from bacteriophage lambda. J. Mol. Biol. 169:757-769,1983.

Table 1: DNA and inferred amino acid sequence changes in the *cro* genes of revertant candidates. Four candidates were found to have aquired changes in the region upstream from the *cro* gene.

Revertant	Base and Amino Acid Change(s)		Isolates
AV11	Ala11→Val GCA→GTA		4
QL16	Gln16→Leu CAA→CTA		1
YC26	Tyr26→Cys TAT→TGT		1
QP27	Gln27→Pro CAA→CCA		1
AS36	Ala36→Ser GCC→TCC		1
AT29/TI65	Ala29→Thr GCG→ACG	Thr65→Ile ACA→ATA	1
AS36/AS29	Ala36→Ser GCC→TCC	Ala29→Ser GCG→TCG	1

Table 2. The melting temperatures, change in melting temperatures and intracellular half lives for wild type Cro and variant Cro proteins are listed. ΔT_m is the change in melting temperature resulting from the suppressor substitution. Thus, ΔT_m is relative to IL30 Cro for the doubly substituted proteins and relative to wild type Cro for the proteins with single mutations. ND-not determined.

Protein	T_m (°C)	ΔT_m (°C)	$T_{1/2}$(min)
IL30	35	0	11
IL30/YC26	47	+12	64
IL30/QP27	38	+3	>300
IL30/AS36	36	+1	84
IL30/QL16	ND	--	168
IL30/AV11	ND	--	126
Wild Type	40	0	63
QL16	54	+14	--
YC26	51	+11	--
AS36	41	+1	--

Figure 1. DNA and amino acid sequences of the synthetic *cro* gene. The translational initiation and stop codons are underlined and shown in bold characters. Restriction endonuclease recognition sequences are boxed. The transcriptional terminator from *trp*A consists of the base pairs between the *Sac*II and *Xba*I restriction sites.

NcoI DraI BalI BglII

CC**ATG**GAACAACGCATAACTTTAAAAGATTATGCAATGCGCTTTGGCCAAACCAAGACAGCTAAAGATCTC
GGTACCTTGTTGCGTATTGAAATTTTCTAATACGTTACGCGAAACCGGTTTGGTTCTGTCGATTTCTAGAG
Met Glu Gln Arg Ile Thr Leu Lys Asp Tyr Ala Met Arg Phe Gly Gln Thr Lys Thr Ala Lys Asp Leu
5 10 15 20

EagI SspI

GGCGTGTATCAAAGCGCGATCAACAAGGCCATCCATGCCGGCCGAAAAATATTTTTGACTATA
CCGCACATAGTTTCGCGCTAGTTGTTCCGGTAGGTACGGCCGGCTTTTTATAAAAACTGATAT
Gly Val Tyr Gln Ser Ala Ile Asn Lys Ala Ile His Ala Gly Arg Lys Ile Phe Leu Thr Ile
25 30 35 40

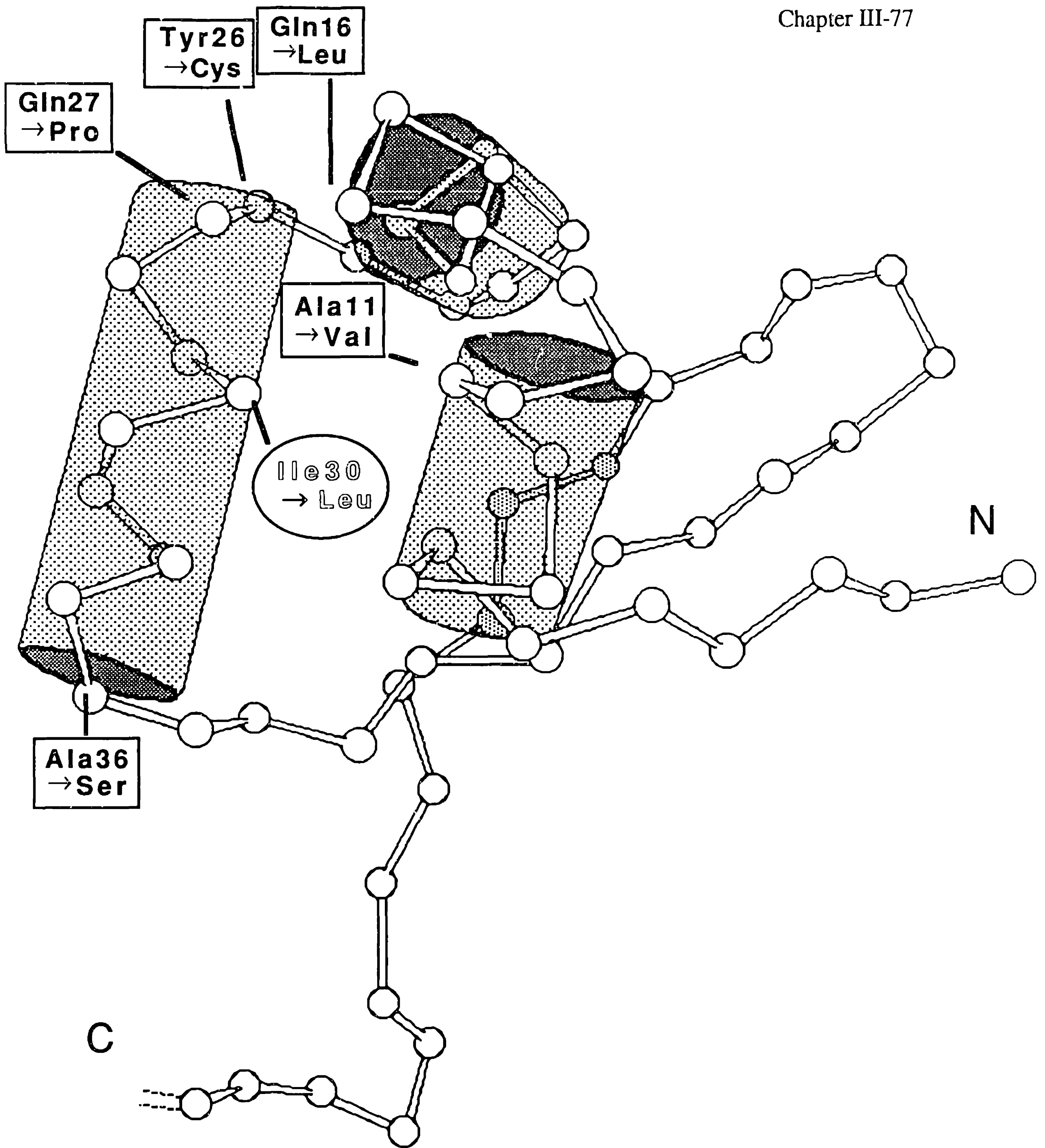
BamHI AccI AvaI

AACGCTGATGGATCCGTATACGCGGAAGAGGTAAAGCCCTTCCCGAGTAACAAAAAAACAACAGCA
TTGCGACTACCTAGGCATATGCGCCTTCTCCATTTCGGGAAGGGCTCATTGTTTTTTTGTTGTCGT
Asn Ala Asp Gly Ser Val Tyr Ala Glu Glu Val Lys Pro Phe Pro Ser Asn Lys Lys Thr Thr Ala
45 50 55 60 65

SmaI KpnI StyI ApaI SacII *trp*A transcriptional terminator XbaI

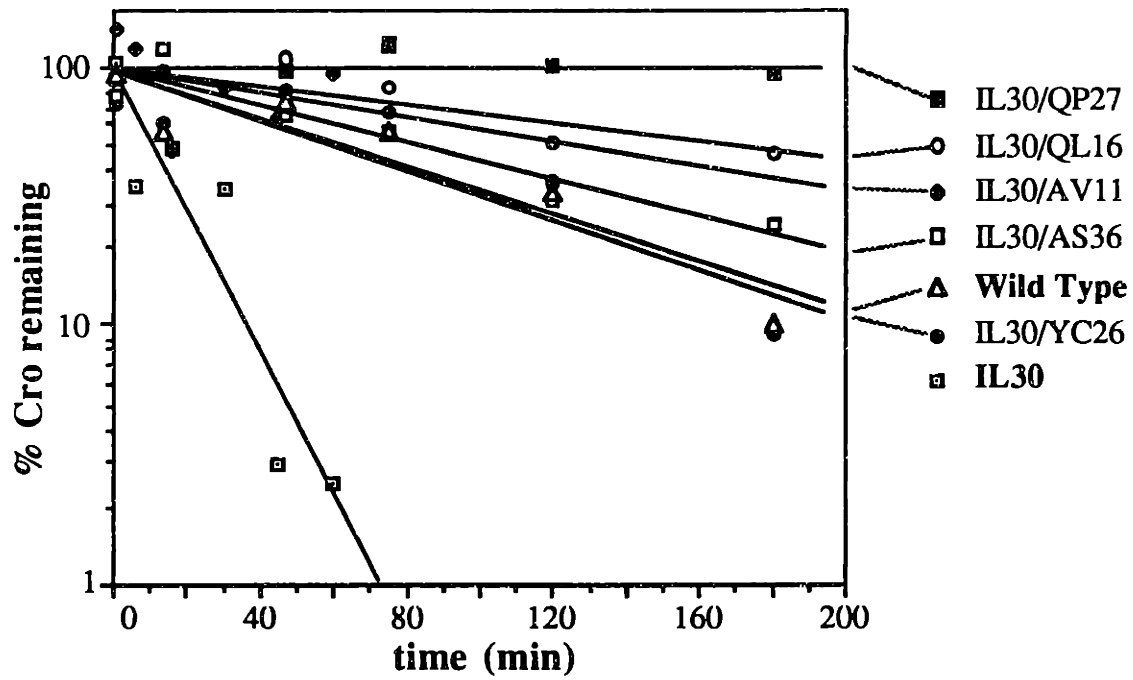
TAAATAACCCGGGTACCTTGGGCCCGCGGAGCCCGCCTAATGAGCGGGCTTTTTTTTCTAGA
ATTTATTGGGCCCATGGAACCCGGGCGCCTCGGGCGGATTACTCGCCCGAAAAAAAAAGATCTAT

Figure 2-The location of each suppressor substitution and the destabilizing Ile30→Leu substitution are shown in a diagramatic representation of a Cro monomer. The four most C-terminal amino acids are not shown, as they are disordered in the crystal structure. Figure adapted from (12).



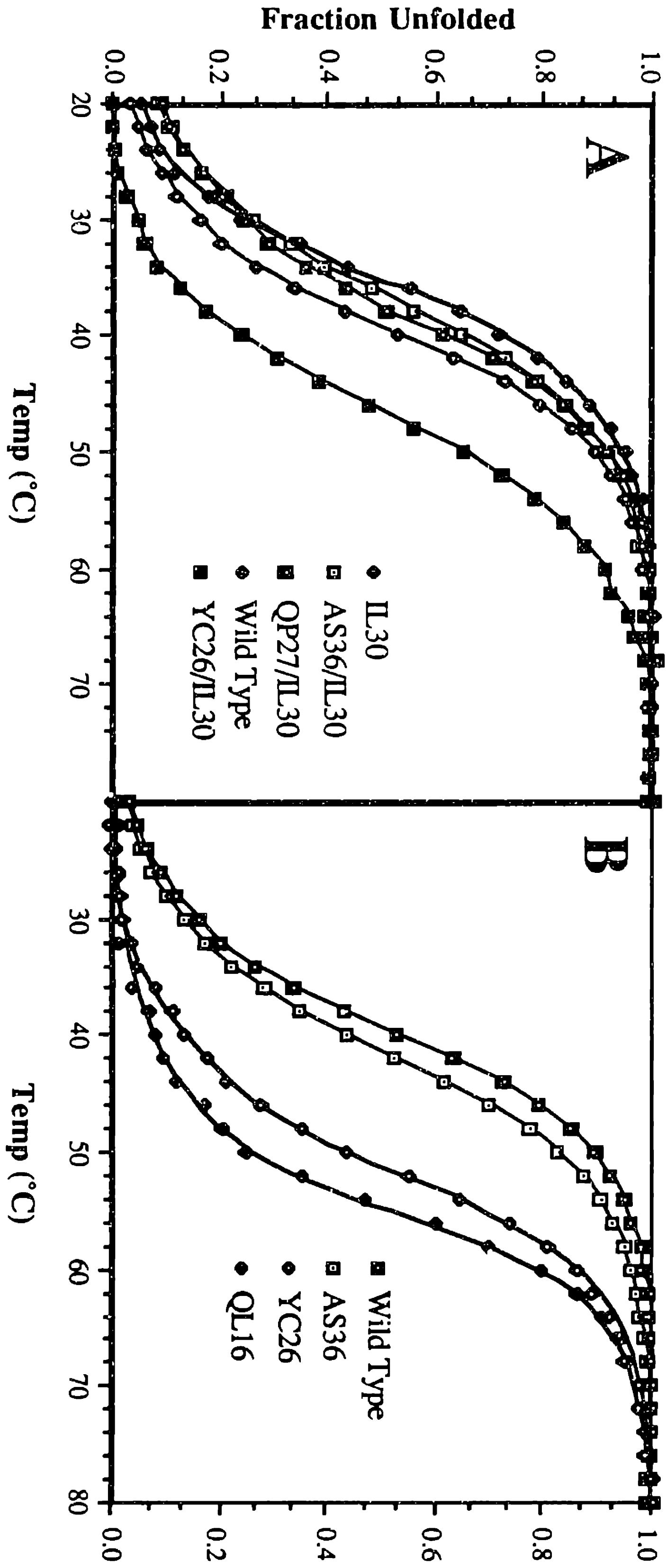
Gln16
→Leu
Tyr26
→Cys
Gln27
→Pro
Ala11
→Val
Ile30
→ Leu
N
Ala36
→Ser
C

Figure 3. Pulse-Chase experiment. The intracellular half-life of the IL30 Cro protein is about 11 minutes, and that of wild type is over 60 minutes. Each revertant protein has a half-life that is at least one hour.



100
10
1
% Cro remaining
0
40
80
120
160
200
time (min)
IL30/QP27
IL30/QL16
IL30/AV11
IL30/AS36
Wild Type
IL30/YC26
IL30

Figure 4. (A) Thermal denaturation profiles for wild type Cro, IL30 Cro, and Cro proteins with the IL30 substitution and one of three suppressor substitutions. (B) Thermal denaturation profiles for wild type Cro and Cro with the AS36, QL16, or YC26 substitutions.


Fraction Unfolded
A
B
IL30
AS36/IL30
QP27/IL30
Wild Type
YC26/IL30
Wild Type
AS36
YC26
QL16
Temp (°C)
Temp (°C)

Chapter IV

The Hydrophobicity of Surface Residues Affects Protein Stability

Introduction

In the previous chapter, we described the isolation of two amino acid substitutions, Gln^{16}→Leu and Tyr^{26}→Cys, that dramatically stabilize Cro to thermal denaturation. Both of these mutations affect residues that are located on the surface of Cro. To further investigate the mechanisms by which these substitutions stabilize the Cro structure, we constructed and purified additional mutant Cro proteins with substitutions at these positions and characterized their thermal stabilities. For position 16, increased residue hydrophobicity results in variant proteins with increased thermal stability. At residue position 26, the relationship between thermal stability and residue hydrophobicity is inverted; decreased residue hydrophobicity leads to increased thermal stability. We constructed a Cro protein containing the most stabilizing substitution at each of these two positions. The melting temperature of this doubly substituted protein is 28 degrees higher than that of wild type Cro, representing a change in the free energy of unfolding of more than 5 kcal/mol. We suggest that the construction of amino acid substitutions at solvent exposed positions is a viable approach to rational protein stabilization and discuss possible mechanisms for the stabilizing effects of the substitutions described here.

Materials and Methods

Nomenclature: Mutant *cro* genes and variant Cro proteins are referred to by the name(s) and superscripted position number(s) of any substituted amino acid residues. For example, a Cro protein in which the wild type Tyr^{26} has been replaced by cysteine, is named Cys^{26}. The wild type amino acids at positions 16, 26, and 30 in λ Cro are Gln, Tyr, and Ile, respectively.

Construction of Mutants: To construct additional substitutions at residues 16 and 26 of λ Cro, the codons for these amino acids were randomized in separate experiments. Oligonucleotide cassettes were synthesized using an Applied Biosystems automated DNA synthesizer. The sequence of the codon to be altered was replaced with an equal mixture of each of the four nucleotides on the sense strand, and inosines on the opposite strand (Reidhaar-Olson and Sauer, 1988). Plasmid pAP119 bears a synthetic *cro* gene into which a number of restriction endonuclease recognition sites have been engineered (Chapter 3). This plasmid contains both an M13 and colE1 origin of replication; the presence of the M13 origin causes single stranded plasmid DNA to be made and packaged upon infection with an M13 helper phage, such as RV1 (Levinson *et al.*, 1984). Cassettes were ligated to plasmid pAP119 backbones generated by restriction with *Nco*I and *Bgl*II or *Bgl*II and *Eag*I for mutagenesis of codons 16 and 26, respectively, and introduced into *E. coli* strain X9T - an *Escherichia coli* strain that bears an F' episome carrying the Lac repressor-overproducing allele *lacI^Q* (Bowie and Sauer, 1988). The DNA sequences of the mutant *cro* genes were determined using single stranded plasmid DNA template and the Sequenase dideoxy sequencing kit (US Biochemicals).

The Tyr^{26}→Asp defective mutation was originally isolated in a *cro* gene under *lac* promoter control (Chapter 2). To increase expression of the Asp^{26} protein, the mutant *cro* gene was placed under *tac* promoter control in plasmid pAP5M, as described previously (Chapter 3). A *cro* gene containing the Leu^{16} and Asp^{26} mutations was constructed by standard recombinant DNA techniques using the *Bgl*II recognition site that separates these two codons.

Chemical Modification: The single cysteine of Cys^{26}/Leu^{30} Cro (Chapter 3) was alkylated with either iodoacetamide or iodoacetate. Purified protein, at a concentration of 0.1 mM, was incubated for 30 min at 22 °C with a 100 fold molar excess of either alkylating reagent in a buffer consisting of 0.1 M KPi [pH 7.0], 0.1 M KCl. After

dialysis, the absence of remaining free thiols was confirmed by reaction with 5,5'-dithiobis (2-nitrobenzoic acid) (Ellman, 1959).

Protein Purification and Thermal Denaturation: Mutant proteins were purified from lysates of *E. coli* strain X9T bearing overproducing plasmids by ion exchange chromatography on DEAE Sephacel (Pharmacia) and Cellex P (Bio-Rad Laboratories), and gel filtration chromatography on Sephadex G50 (Pharmacia) as described (Chapter 3). Each of the mutant Cro proteins described here behaved essentially like wild type Cro at all steps of the purification procedure. The purification of the Leu^{16}, Cys^{26}, Leu^{30}, Cys^{26}/Leu^{30}, and wild type Cro proteins has been described (Chapter 3).

Thermal denaturation of mutant proteins and two alkylated derivatives of Cys^{26}/Leu^{30} Cro was performed at protein concentrations of 120 µg/ml in 10 mM KPi [pH 7.0], 0.1 M KCl. The temperature was changed at a rate of 0.8 °C per minute. Protein unfolding was monitored by recording changes in circular dichroism ellipticity at 222 nm as described previously (Chapter 3). The transition from the Cro dimer (N_2) to two unfolded monomers (2U) is a concerted reaction in which the folded monomer is presumably an unpopulated intermediate. As a result, the T_m will vary with Cro concentration. All unfolding reactions reported here were performed at the same concentration of wild type or mutant Cro. The equilibrium constant for unfolding is defined as $K_u=[U]^2/[N_2]$. $\Delta\Delta G_u$ values are calculated as $\Delta\Delta G_u=-RTln[K_u^{mutant}/K_u^{wt}]$ using K_u values measured or calculated from the van't Hoff equation at 45 °C.

Results:

Construction and purification of mutant Cro proteins: Mutant codons at residue positions 16 and 26 were generated in a plasmid-borne *cro* gene by cassette mutagenesis.

Three new amino acid substitutions at position 16, and six new substitutions at position 26 were isolated. In addition to wild type Cro protein, which contains Gln at residue 16 and Tyr at residue 26, variants containing Ser^{16}, His^{16}, Tyr^{16}, Leu^{16}, Gln^{26}, His^{26}, Leu^{26}, Cys^{26}, Asp^{26}, Trp^{26}, Val^{26}, and Phe^{26} have been purified and characterized. The behavior of each variant protein during purification was essentially indistinguishable from that of wild type Cro. The CD spectrum of each mutant Cro protein was also examined and found to be identical to that of wild type Cro. These facts suggest that none of the substitutions results in a gross alteration of the structure of the Cro protein.

Effects of Substitutions at Position 16 on Thermal Stability: Representative unfolding profiles for variant Cro proteins substituted at position 16 are shown in Fig. 1. The melting temperature (T_m) and the change in the free energy of unfolding relative to wild type ($\Delta\Delta G_u$) for these mutant proteins are listed in Table I. Replacement of the wild type Gln^{16} with either histidine or serine results in little or no change in the thermal stability of Cro. However, substitution with the hydrophobic amino acids, Tyr or Leu, results in substantial increases in stability. The melting temperature of the Tyr^{16} protein is 7 degrees greater than the T_m of wild type Cro; the change in the free energy of folding is 1.4 kcal/mol. Leu^{16} is the most stabilizing change that we have isolated at this position. Its T_m is 14 °C greater than that of wild type, and $\Delta\Delta G_u$ is 2.7 kcal/mol.

Fig. 2 shows that there is a rough correlation between the stabilizing effects of the residue 16 substitutions and the hydrophobicity of the side chain, as measured by free energies of transfer from ethanol to water. This suggests that the stabilizing effects of the Tyr^{16} and Leu^{16} substitutions may arise from new hydrophobic packing interactions. The side chain of Gln^{16} is partially buried in the crystal structure of wild type Cro. It seems reasonable to expect that Leu^{16} would also be partially buried as there are several nearby

aliphatic side chains against which the Leu^{16} side chain could pack. Examination of the structure suggests that a Tyr^{16} side chain could make some of the same contacts as Leu^{16} but, because of the planar nature of the aromatic ring, the packing would not be as efficient as in the Leu^{16} case.

Effects of Residue 26 Substitutions on Stability: The melting temperatures of proteins substituted at position 26 vary over a 17 °C range (Fig. 3, and Table 2). The most stable variant is Asp^{26} which has a melting temperature of 54 °C ($\Delta\Delta G_u$=2.5 kcal/mol). The only substitution at position 26 that renders the protein less thermally stable than wild type Cro is tryptophan; the melting temperature of the Trp^{26} protein is 2 °C lower than that of wild type Cro. To determine whether the stabilizing effect of a substitution at position 26 is specific to heat denaturation, we performed urea denaturation of the wild type (Tyr^{26}) and Asp^{26} Cro proteins. These proteins are half denatured at 1.8 M and 3.9 M urea respectively (not shown). The $\Delta\Delta G_u$ values measured from these experiments range from 3.1 to 3.4 kcal/mol over the range of 1.5 to 4.5 M urea. Hence, the substitution of Asp for Tyr at residue 26 stabilizes Cro to thermal denaturation and urea denaturation to roughly comparable extents.

Residue 26 is located in the turn between α-helices 2 and 3 of λ Cro, and the wild type Tyr^{26} side chain has a very high solvent accessibility (Anderson *et al.*, 1982). Figure 4 shows the relationship between $\Delta\Delta G_u$ for each position 26 substitution and the hydrophobicity of the relevant amino acid. The effect of residue hydrophobicity is exactly opposite of that at position 16. At residue 26, increasing thermal stability results from decreasing hydrophobicity.

One of the most stabilizing substitutions at residue 26 is Cys, raising the possibility that the enhanced stability could be caused by disulfide bond formation. However, we have found that alkylation of the mutant Cys^{26} side chain with a reagent that creates a

charged adduct (iodoacetic acid) results in an additional four degree increase in melting temperature, while alkylation with the uncharged amide derivative of this reagent (iodoacetamide) does not appreciably affect the thermal stability. These results show that disulfide formation is not involved in the increased stability of Cys^{26}, and further support the relation between decreasing hydrophobicity and increasing stability at residue 26.

Enhanced Stability of a Doubly Mutant Cro protein: The Leu^{16} and Asp^{26} substitutions each increase the melting temperature of Cro by about 14 °C and change the free energy of unfolding by about 2.5 kcal/mol. To examine the additivity of these two stabilizing substitutions, we constructed a *cro* gene containing both mutations and purified the doubly mutant protein (Leu^{16}/Asp^{26}). The thermal denaturation of this protein, as monitored by changes in ellipticity at 222 nm, is shown Fig. 5, along with denaturation profiles for the Leu^{16}, Asp^{26}, and wild type Cro proteins. The melting temperature of the doubly substituted protein is about 68 °C, an increase of 28°C relative to wild type Cro; the change in the free energy of unfolding is 5.5 kcal/mol. These increases represent nearly exact sums of the increases afforded by each of the two substitutions alone. Thus, the effects of the two substitutions are almost completely additive.

It is possible that the amino acid substitutions at positions 16 and 26 stabilize a different, more stable, conformation or λ Cro. However, the additivity of the stabilizing effects of the Leu^{16} and Asp^{26} substitutions suggests that the conformations of the singly and doubly mutant proteins are likely to be similar (Carter *et al.*, 1984). Moreover, this conformation is likely to be similar to wild type since: (*i*) The CD spectra of these proteins are identical to that of wild type; and (*ii*), the operator DNA binding affinities of the Cys^{26}, Trp^{26}, Phe^{26}, and Gln^{26} proteins are not severely reduced (data not shown). Any alteration in the three dimensional arrangement of the amino acid side chains near residue

26 would be expected to have deleterious effects on the DNA binding activity of Cro, as this position is in the helix-turn-helix DNA binding region of the protein (Anderson *et al.*, 1982).

Discussion

The Gln^{16}→Leu and Tyr^{26}→Cys substitutions were initially isolated using random mutagenesis followed by a colony screen to identify amino acid substitutions that could suppress the destabilizing effects of a mutation elsewhere in λ Cro (Chapter 3). Since the mechanisms of stabilization were not apparent from inspection of the structure of the wild type Cro protein, we constructed and characterized several different amino acid substitutions at each of these residue positions. The hydrophobicity of the side chains at both positions appears to play a role in their effects on stability, however, hydrophobicity has opposite consequences for stability at the two positions.

The hydrophobic effect, first proposed by Kauzman (1959), is believed to be a major stabilizing force for protein structure. Exposure of hydrophobic atoms to aqueous solvent causes water molecules to become structured around the side chain. When the hydrophobic atoms are moved to a non-polar environment, the release of the organized solvent gives rise to a favorable free energy of transfer. It should be noted that hydrophobic effects are only relevant for side chains whose solvent-accessibility changes upon folding. A hydrophobic group that is equally exposed to solvent in the native and denatured states should not contribute to, or detract from, the stability of the folded structure.

At position 16, residues with higher hydrophobicity appear to increase stability. This observation suggests that hydrophobic side chains at this position are less exposed to solvent in the folded structure than they are in the unfolded state of the protein. The side chain of Gln^{16} in wild type Cro is about 60% solvent inaccessible and inspection of the

crystal structure suggests that side chains of hydrophobic residues at this position could make hydrophobic contacts with the side chains of Ala^{11}, Ile^{30}, and Ile^{34} which are predominantly buried in the structure. Assuming that side chains in the denatured state are substantially exposed to solvent, the reduction in exposed hydrophobic surface area upon folding could account for the increased stability of the Cro proteins with hydrophobic residues at position 16. Contrary to the results for position 16, increasing hydrophobicity of the residue at position 26 results in decreased thermal stability. For example, Asp^{26} is the most stabilizing substitution at this position and Trp^{26} is the least stabilizing. This is the reverse of the traditional hydrophobic effect, and if interpreted in the same way, implies greater solvent accessibility in the folded form than in the denatured form.

Can the magnitudes of the stabilizing effects of substitutions at residue positions 16 and 26 be accounted for by hydrophobic considerations alone? Considering only effects due to the hydrophobic effect, the change in the free energy of unfolding ($\Delta\Delta G_u$) caused by a substitution from amino acid A to amino acid B can be estimated as:

$$\Delta\Delta G_u=(2)\Delta G_t^A(FA_u^A-FA_n^A) - (2)\Delta G_t^B(FA_u^B-FA_n^B) \qquad \text{eq. 1}$$

where ΔG_t is the free energy of transfer of a side chain from water to ethanol (Finney, 1975), and FA_u and FA_n are the fractional solvent accessibilities of the side chain in the native and denatured states, respectively. Fractional solvent accessibilities for side chain X are defined relative to the reference tripeptide Ala-X-Ala in an extended conformation as described by Lee and Richards (1971). The factor of 2 is introduced because the effects of an amino acid substitution will be exerted in each monomer of the Cro dimer. If we make the simplifying assumption that $(FA_u^A-FA_n^A)$ equals $(FA_u^B-FA_n^B)$, that is the two residues show comparable changes in fractional accessibility in going from the folded to the unfolded form, then:

$$\Delta\Delta G_u = (2)(\Delta\Delta G_t)(FA_n-FA_u) \qquad \text{eq. 2}$$

where $\Delta\Delta G_t$ represents the difference between the transfer free energies of the two relevant amino acid side chains. In the discussion that follows, we will assume that the fractional accessibility for all side chains in the unfolded state is 0.9. We will return to a discussion of this assumption below.

At position 16 in the native Cro structure, the wild type glutamine side chain has a fractional accessibility of 0.37. If we assume that the Leu^{16} side chain has the same accessibility, then the predicted $\Delta\Delta G_u$ for the $Gln^{16} \rightarrow Leu$ substitution is 2.4 kcal/mol. This value compares favorably with the experimentally determined stabilization of 2.7 kcal/mol. With the same assumptions, the predicted value of $\Delta\Delta G_u$ for Tyr^{16} is 2.9 kcal/mol. This is more than twice the observed value of 1.4 kcal/mol. However, inspection of the Cro structure shows that no more than half of the tyrosine side chain could be buried. To bring the observed and calculated values of $\Delta\Delta G_u$ into agreement, an FA_n value of 0.65 would be required for Tyr^{16}. Model building suggests that this is not an unreasonable value.

At position 26 in the native structure, the wild type tyrosine side chain has a solvent accessibility that is 1.4 times that expected for tyrosine in the reference state (Lee and Richards, 1971). This unusually high solvent exposure appears to result from the location of this residue at a turn. The backbone conformation in this part of Cro directs the Tyr^{26} side chain away from main chain atoms and from neighboring side chains. If we assume that any side chain will have the same fractional accessibility as the tyrosine in wild type Cro, then:

$$\Delta\Delta G_u = (2)(\Delta\Delta G_t)(1.4 - 0.9) = \Delta\Delta G_t$$

This predicts that the slope of a plot of $\Delta\Delta G_u$ vs. ΔG_t will be close to unity for residue substitutions at position 26. The line drawn through the data points in Figure 4 has a slope of 0.82, which is in good agreement with the expected slope. There is no reason to

expect an exact match. The $\Delta\Delta G_u$ values are calculated at 45°C, whereas the ΔG_t values are calculated at 20°C. Moreover, the assumptions with respect to fractional accessibility are approximate, at best. Hence, with the exception of Gln^{26}, which falls significantly off the line, the experimental data can be explained in a reasonable fashion by the simple model outlined above.

The correlation between the expected and observed values of $\Delta\Delta G_u$ is sufficiently good to suggest that hydrophobicity is the main factor in determining the relative stabilities of Cro proteins substituted at positions 16 and 26. Other factors such as hydrogen bonding, electrostatic interactions with helical dipoles, etc., may also be operative for some substitutions, but these contributions to stability are probably relatively minor. In the preceding calculations, we have assumed that the fractional solvent accessibility of a side chain in the unfolded state of Cro is 0.9. This value seems reasonable as a value of 1.0 is calculated for an extended peptide conformation and the denatured state is probably somewhat compact (Shortle, 1988**).

To date, most directed strategies for improving protein stability have attempted to reduce the entropy of protein unfolding by introducing disulfide cross-links or substitutions of the type Gly →Xxx and Xxx→Pro. In the cases where these attempts have been successful, stabilizations of 0.3 to 2.5 kcal/mol have been reported (Hecht *et al.*, 1984; Matsumura *et al.*, 1986; Matthews *et al.*, 1987; Stearman *et al.*, 1988). Our results suggest that the hydrophobicity of some solvent accessible residues can influence protein stability to a significant effect. It would clearly be useful if surface positions of these types could be identified in a systematic way. We suggest that two different classes of surface residues would be good targets for the rational manipulation of protein stability. The first class is comprised of partially buried side chains. At these positions, replacing hydrophilic side chains with hydrophobic side chains may lead to enhanced stability. It may be important, in these cases, to identify locations where the hydrophobic side chain could pack

against nearby nonpolar side chains. The second class is comprised of residues that are hyper-exposed to solvent . At these positions, the substitution of hydrophobic side chains with hydrophilic side chains should enhance stability.

Acknowledgements: We thank Brian Matthews for providing refined coordinates for λ Cro, and Jim Bowie for help with the solvent accessibility calculations.

References:

Anderson, W.F., Ohlendorf, D.H., Takeda, Y., Matthews, B.W. Nature 290:754-758, 1981.

Bowie, J. U., Sauer, R. T. 1988 (submitted).

Carter, P. J., Winter, G., Wilkinson, A. J., Fersht, A. R. Cell 38:835-840, 1984.

Ellman, G. L. Arch. Biochem. Biophys. 82:70, 1959.

Finney, J.L. J. Mol. Biol. 96:721-732, 1975.

Hecht, M.H., Sturtevant, J.M., Sauer, R.T. Proc. Natl. Acad. Sci. 81:5685-5689, 1984.

Kauzman, W. Adv. Prot. Chem. 14:1-63, 1959.

Lee, B., Richards, F. M. J. Mol. Biol. 55:379-400, 1971.

Levinson, A., Silver, D. Seed, B. J. Mol. Appl. Genet. 2:507-517, 1984.

Matsumura, M., Yasamura, S., Aiba, S. Nature 323:356-358, 1986.

Matthews, B. W., Nicholson, H., Becktel, W. J. Proc. Natl. Acad. Sci. 84:6663-6667, 1987.

Reidhaar-Olson, J. F., Sauer, R. T., Science 241:53-57, 1988.

Shortle, D. Biochemistry, 1988 (in press)**.

Stearman, R. S., Frankel, A. D., Freire, E., Sturtevant, J. M., Liu, B., Pabo, C. O., Biochem.,1988 (in press).

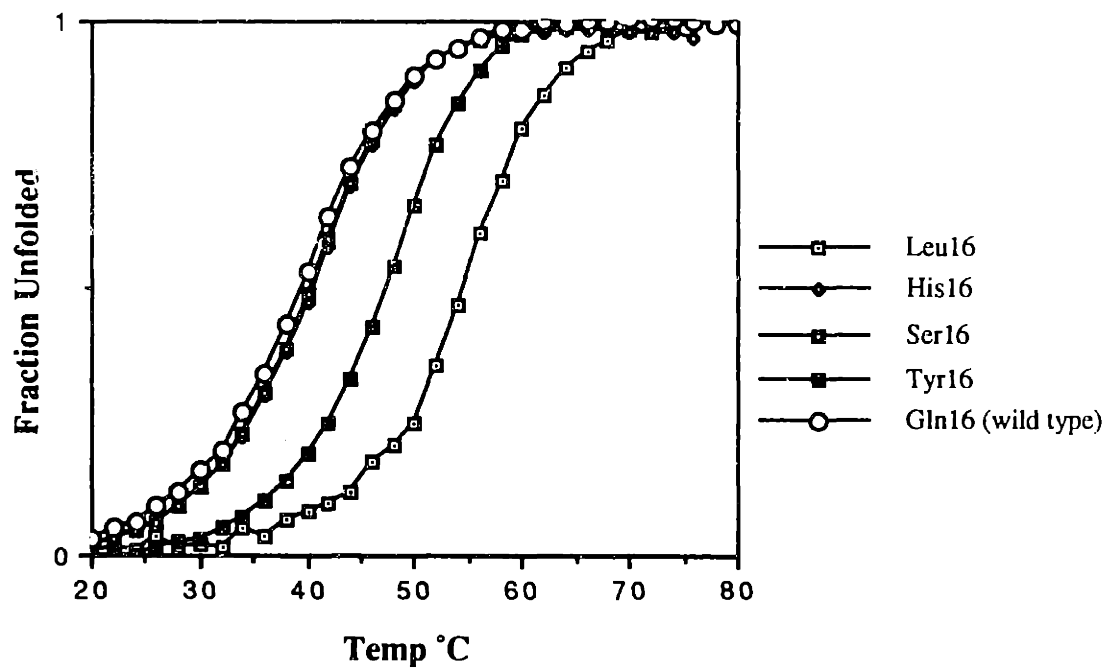
Table 1: Stability changes associated with residue substitutions at position 16. †$\Delta\Delta G_u$ is calculated at 45°C. [a] Data from Chapter 3.

Residue	T_m°C	$\Delta\Delta G_u$ (kcal/mol)†
Gln^{16} [a]	39.5	0.0← wild type
Ser^{16}	40.0	+0.1
His^{16}	40.5	+0.1
Tyr^{16}	47.0	+1.4
Leu^{16} [a]	54.0	+2.7

Table 2: Stability changes associated with residue substitutions at position 26. †$\Delta\Delta G_u$ is calculated at 45°C. [a] Data from Chapter 3.

Residue	T_m °C	$\Delta\Delta G_u$ (kcal/mol)†
Trp^{26}	37.5	-0.1
Tyr^{26} [a]	39.5	0.0 ← wild type
Phe^{26}	41.5	+0.4
Val^{26}	44.5	+0.8
Leu^{26}	46.0	+1.1
Gln^{26}	47.0	+1.3
His^{26}	49.5	+1.8
Cys^{26} [a]	52.0	+2.1
Asp^{26}	54.0	+2.5

Figure 1. Thermal denaturation profiles for the Leu^{16}, His^{16}, Ser^{16}, Tyr^{16}, and wild type Cro proteins. The stabilities of the His^{16} and Ser^{16} proteins are very similar to that of wild type, while the Tyr^{16} and Leu^{16} proteins have significantly higher melting temperatures.



Fraction Unfolded
1
0
20
30
40
50
60
70
80
Temp °C
Leu16
His16
Ser16
Tyr16
Gln16 (wild type)

Figure 2. The $\Delta\Delta G_u$ of unfolding for each substitution at residue 16 is plotted vs. the ΔG_t of transfer from ethanol to water for the relevant amino acid.

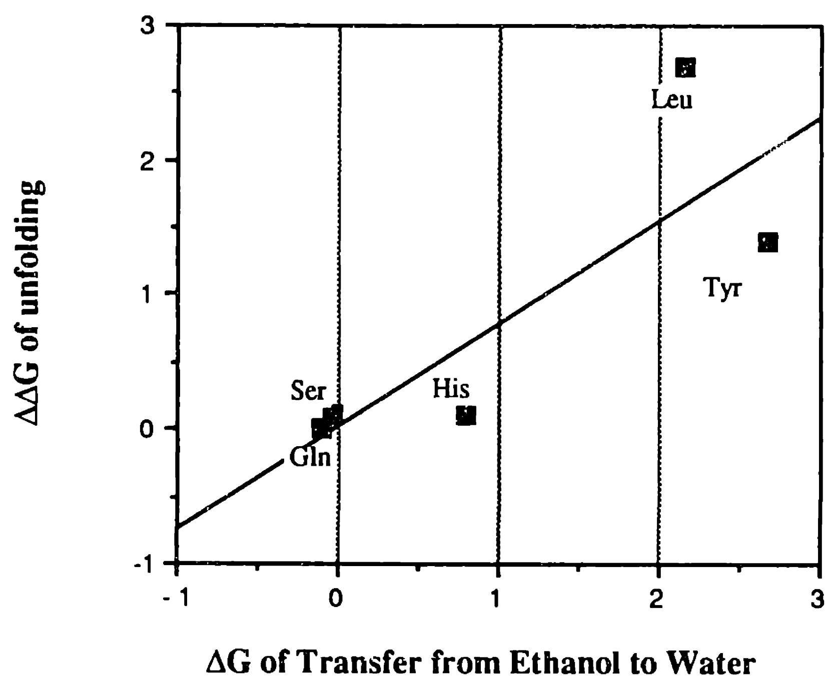
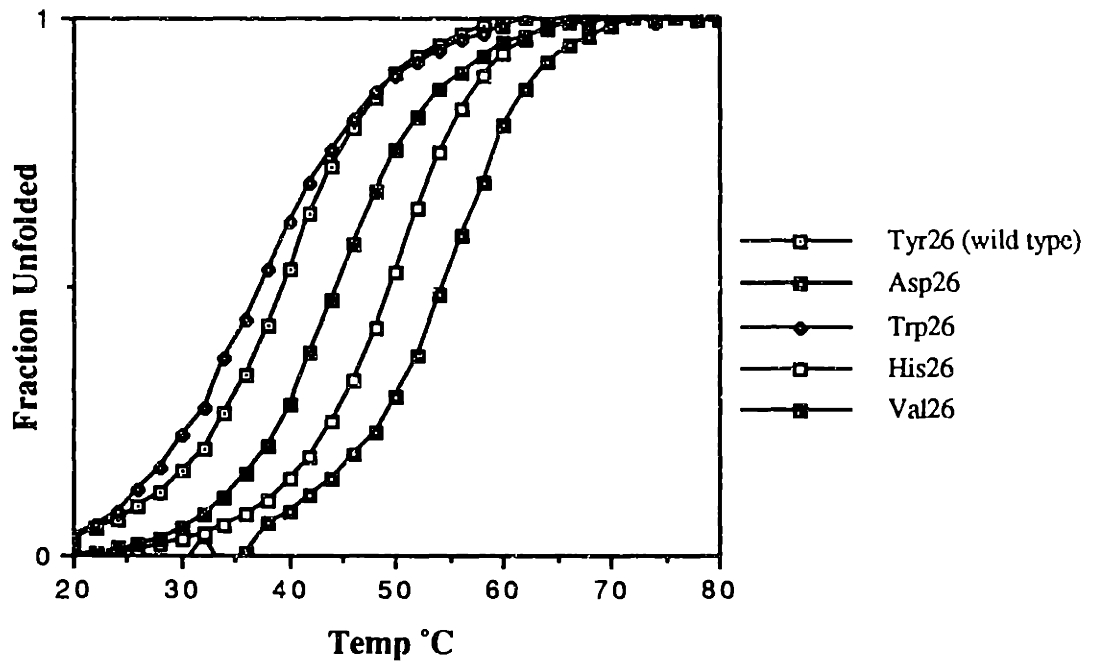
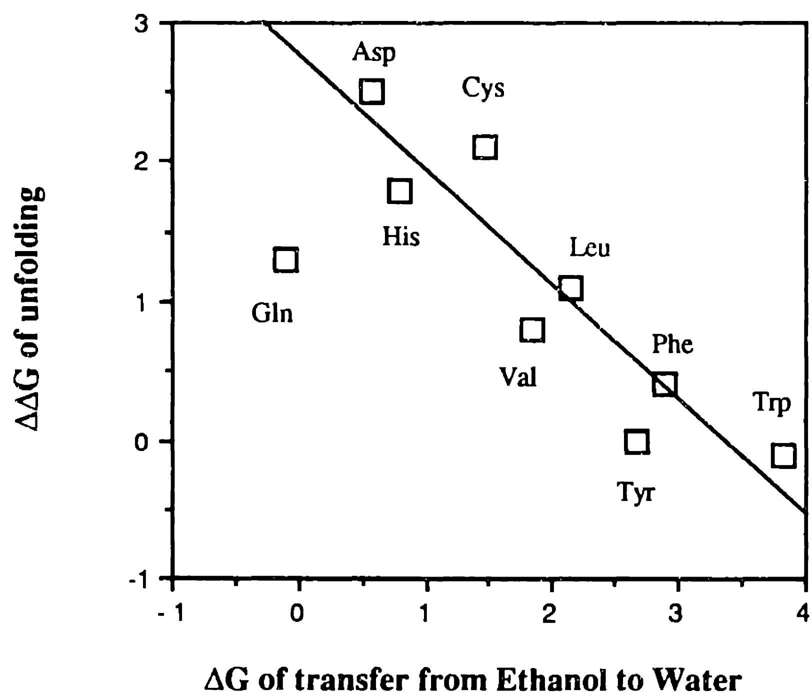

3
2
1
0
-1
- 1
0
1
2
3
Leu
Tyr
Ser
His
Gln
ΔΔG of unfolding
ΔG of Transfer from Ethanol to Water

Figure 3. Thermal denaturation profiles for the Trp^{26}, Val^{26}, His^{26}, Asp^{26} and wild type (Tyr^{26}) Cro proteins. The Trp^{26} Cro protein is less stable than wild type. Each of the other substitutions increases stability.



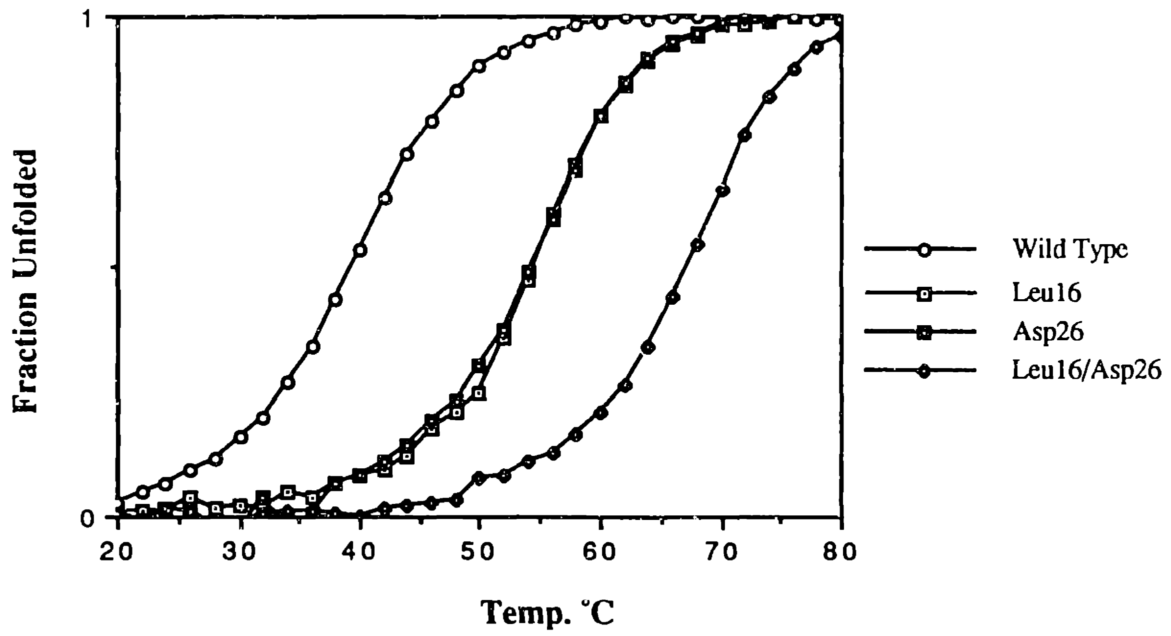
1
0
Fraction Unfolded
20
30
40
50
60
70
80
Temp °C
Tyr26 (wild type)
Asp26
Trp26
His26
Val26

Figure 4. The $\Delta\Delta G_u$ of unfolding for each substitution at residue 26 is plotted vs. the ΔG_t of transfer from ethanol to water for the relevant amino acid.



Asp
Cys
His
Gln
Leu
Val
Phe
Trp
Tyr
3
2
1
0
-1
- 1
0
1
2
3
4
ΔΔG of unfolding
ΔG of transfer from Ethanol to Water

Figure 5. Thermal denaturation profiles for the wild type, Leu^{16}, Asp^{26}, and Leu^{16}/Asp^{26} Cro proteins. The increase in stability that results from the combination of the Leu^{16} and Asp^{26} substitutions is approximately equal to the sum of the increases in stability caused by these two substitutions individually.


Fraction Unfolded
1
0
20
30
40
50
60
70
80
Temp. °C
Wild Type
Leu16
Asp26
Leu16/Asp26