Core Packing Defects in an Engineered Cro Monomer Corrected by Combinatorial Mutagenesis†

A. K. M. M. Mollah,‡ Michael A. Aleman,‡,§ Ronald A. Albright,‖,⊥ and Michael C. Mossing*,‡

Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, and Institute of Molecular Biology, Howard Hughes Medical Institute, and Department of Physics, University of Oregon, Eugene, Oregon 97403

Received August 18, 1995; Revised Manuscript Received October 26, 1995⊗

ABSTRACT: The crystal structure of an engineered monomer of the λ Cro repressor shows unexpected expansion of the hydrophobic core of the protein and disorder of the five C-terminal residues [Albright et al. (1996) *Biochemistry 35*, 735−742]. This structural information has guided the construction of a second generation of monomeric Cro proteins by combinatorial mutagenesis of selected core and C-terminal residues. Clones were identified in a library of randomized *cro* genes by a genetic screen for protein accumulation in *Escherichia coli*. Sequencing of candidate genes followed by purification and analysis of their product proteins has identified alternative arrangements of hydrophobic core residues which result in substantial increases in thermal stability. In contrast, residue replacements at the C-terminus have minor effects on stability but may increase protein expression levels.

Central to many current efforts in protein engineering is the principle that protein structure is determined by a minimum in free energy (Anfinsen, 1973). With this principle as our guide, progress in protein design and engineering depends on a thorough understanding of the relationship between the sequence, structure, and thermodynamic stability of proteins. Several recent efforts have been aimed at quantifying the relationship of hydrophobic packing to stability (Richards & Lim, 1993; Eriksson et al., 1992; Lim et al., 1994). We are interested in the relationship between structure and function in a small DNA-binding protein, the λ Cro repressor.

A family of simplified monomeric variants of Cro was constructed (Mossing & Sauer, 1990) in which the intermolecular antiparallel β-ribbon that forms the dimer interface was replaced by an intramolecular β-hairpin. A key feature of the design of these proteins was the maintenance of the contacts between phenylalanine 58 and residues in a hydrophobic pocket, as observed in the original Cro structure (Anderson et al., 1981). A critical role for F58 was identified in mutagenesis experiments. Although other residues in the dimer interface were functionally replaceable by many different amino acid residues, only a single functional replacement (by a tyrosine) of F58 was recovered (Mossing & Sauer, 1990). Each of the engineered monomers has a five amino acid insertion after position 56 of wild-type Cro. The degenerate codons 56a and 56b encoded all combinations of residues commonly found in tight β-hairpins (Sibanda et al., 1989). Codons 56c, 56d, and 56e encode duplicates of amino acids E54, V55, and K56 and replace amino acids normally provided by the opposite subunit in the dimer. One of the resulting monomers, K56-[DGEVK] (in which the β-turn consists of K56, D56a, G56b, and E56c), has been studied in detail (Mossing & Sauer, 1990). Its crystal structure has recently been solved by Albright et al. (1996).

In the particular case of K56-[DGEVK], the overall three-dimensional structure is very close to that predicted in the original design (Albright et al., 1996). It was surprising, however, to see that the hydrophobic core of the protein was expanded (see Figure 1b). In particular, Phe 58 is retracted by approximately 1 Å from the interior face of the α3 helix relative to its position in the wild-type Cro structure. The location of the engineered turn some 14 Å distant along a relatively extended stretch of β-strand from F58 would hardly predict this type of structural readjustment. The packing defects in the core of this protein may be an important clue in understanding the DNA-binding function of wild-type Cro, as a similar core expansion has been observed in a recent high-resolution structure of the complex of wt Cro and its operator site (R. A. Albright and B. W. Matthews, unpublished).

The last six amino acids in Cro are disordered both in the original crystal structure of dimeric Cro (Anderson et al., 1981) and in the monomer structure (Albright et al., 1996). There is evidence as well from NMR studies of the monomer that the C-terminus is highly mobile (M. C. Mossing, unpublished). Extensive site-directed mutagenesis studies of the C-terminus of wild-type Cro (Hubbard et al., 1990) have shown it to be critical for DNA binding. Investigation of the importance of this region for protein stability might allow one to disentangle the role of the C-terminus in maintaining protein stability from its role in DNA recognition. Stable truncated forms would allow the attribution of specific DNA-binding functions to the C-terminus.

This paper presents our efforts to overcome the packing defects present in the initial Cro monomer design. We have used combinatorial mutagenesis, guided by the structure of

† Supported by NIH Grant GM46513.

* To whom correspondence should be addressed. Phone: (219) 631-8412. E-mail: mossing.1@nd.edu.

‡ University of Notre Dame.

§ Present address: Vanderbilt University School of Medicine, Nashville, TN 37215.

‖ University of Oregon.

⊥ Present address: Department of Molecular Biophysics and Biochemistry, Yale University, Boyer Center for Molecular Medicine, Room 154, 295 Congress Ave., New Haven, CT 06510.

⊗ Abstract published in *Advance ACS Abstracts,* January 1, 1996.

0006-2960/96/0435-743$12.00/0

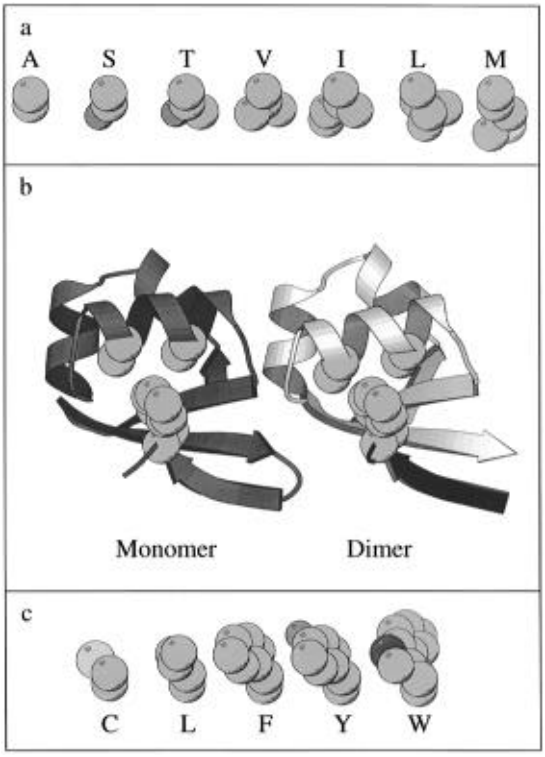

FIGURE 1: Expansion of the hydrophobic core of a Cro monomer and side chains selected to fill the gaps in the structure. (a) CPK representation of the side chains (including α carbons) encoded by the mutagenized codons at positions 29 and 33. Spheres indicate van der Waals surfaces of the non-hydrogen atoms in the side chain. All seven amino acids were specified by the degenerate codon at position 29 while only A, T, V, I, and M were specified at position 33. (b) Ribbon drawing of the structure of Cro mutant K56-[DGEVK] (left) and one-half of a wt Cro dimer (right). In the wt Cro structure, residues 1−56 of the A chain are represented as light ribbons, and residues 54−60 of the C chain are drawn as a dark ribbon. Side chains of residues 29, 33, and 58, which are involved in core packing, are shown in CPK representation. The views shown are the result of superimposing the two structures on the basis of the F58 side chains and then displacing the structures horizontally (c) CPK views of side chains substituted at position 58 as in (a) above. The figure was drawn using the program MOLSCRIPT (Kraulis, 1991) with monomer coordinates from Albright et al. (1996) and Cro coordinates from D. Tronrud and B. W. Matthews (personal communication).

the first generation monomer, and a genetic screen for protein expression level to construct a second generation of proteins with enhanced stability.

MATERIALS AND METHODS

Structure Analysis. The volumes of individual residues in the structure of Cro K56-[DGEVK] (Albright et al., 1996) were calculated by the Voroinoi polyhedra method of Richards (1974) as implemented in the VADAR package (Wishart et al., 1994). The average occupied volumes of residues for normalization were from the VADAR tabulation based on atomic radii of Shrake and Rupley (1973). Accessible surface areas were also calculated with VADAR using the method of Richmond (1984).

Mutagenesis and Cloning. Two primers corresponding to helix α3 (H3),[1] a 51-mer of sequence: 5′-GCT AAA GAT CTC GGC GTG TAT CAA AGC (AGT)(CT)(AG) ATC AAC AAG (AG)(CT)(ACGT) ATC CAT GCC-3′, and the C-terminus (CT), a 41-mer of sequence 5′-TG CCA AGC TTA T(CT)(AT) T(CT)(AT) CGG (CG)(ACT)A GGG CTT TAC TTC GCC GTC-3′, were synthesized for PCR mutagenesis. Bases in parentheses signify an equimolar mixture of the indicated bases at a given position. Primer H3 corresponds to the nontemplate strand of the mDG gene from codons 20 to 36. It maintains the *Bgl*II site of the mDG gene for cloning while allowing changes at codons 29 and 33. The degeneracy at codon 29 produces a mixture of Ala, Thr, Ser, Val, Leu, Ile, and Met codons at ratios of 2:2:2:2:2:1:1. Codon 33 encodes Ala, Thr, Val, Ile, and Met at ratios of 4:4:4:3:1. The CT primer corresponds to the template strand of the Cro gene from codon 65 (5′) to codon 57 (3′) of the Cro.mDG gene in the plasmid pUCro.mDG (Mossing & Sauer, 1990). It introduces changes in the sequence at codons 58, 60, and 61 (including possible stop codons) and specifies a unique stop codon at 62. The nucleotide mix selected for codon 58 gives an equal probability of Trp, Phe, Tyr, Leu, Cys, and stop codons. At codons 60 and 61 Lys, Arg, and stop codons are encoded with a frequency of 1:1:2, while at 62 the stop codon UAA was uniquely specified. At the 5′ end of primer CT is a *Hind*III site which is 39 bp upstream of the *Hind*III site in mDG gene. A third "reverse sequencing" primer corresponding to nucleotides 186−209 of pUC118 (Genbank U07649) was used to generate PCR products which included the entire 5′ tac promoter region as well as the complete N-terminus of the *cro* gene.

The plasmid pUCro.mDG (Mossing & Sauer, 1990) was used as the template for the PCR mutagenesis as illustrated in Figure 2. In the first reaction the CT and RS primers were used to obtain a 554 base pair fragment that has incorporated degenerate codons at positions 58, 60, and 61. A second reaction using primers H3 and CT yielded a 152 base pair fragment containing the changes at codons 29, 33, 58, 60, and 61. Individual cycles of amplification involved denaturation at 94 °C for 90 s, annealing at 55 °C for 60 s, and extension with Vent polymerase from New England Biolabs at 73 °C for 45 s. Reactions proceeded for a total of 30 cycles. PCR products of the expected sizes were visualized by agarose gel electrophoresis.

All cloning and expression was carried out in *Escherichia coli* strain X90 (Amann et al., 1983). To facilitate production of single-stranded DNA for sequencing, the Cro.mDG gene was transferred to a vector which contained an M13 origin of replication. A 523 bp *Eco*RI to *Hin*dIII fragment of pUCro.mDG containing the tac promoter and Cro.mDG gene was ligated to the *Eco*RI and *Hin*dII cleaved PUC118 (Vieira & Messing, 1987). The construction of the randomized PCR fragments is outlined in Figure 2. Both pUC118.mDG and the PCR product of the CT and H3 primers were digested with *Bgl*II and *Hin*dIII. Similarly, pUC118.mDG and the

[1] Abbreviations: IPTG, isopropyl β-D-thiogalactoside; EDTA, ethylenediaminetetraacetic acid; BME, β-mercaptoethanol; SDS, sodium dodecyl sulfate; LB, Luria broth; PEI, poly(ethylenimine); CD, circular dichroism spectropolarimetry; PCR, polymerase chain reaction; F_u, fraction unfolded; T_m, temperature at which half of the protein has undergone the unfolding transition; wt, wild type. Primers used in the mutagenesis are designated CT for C-terminal, H3 for helix α3, and RS for reverse sequencing and are described in the text. mDG is used to designate the monomeric Cro mutant K56-[DGEVK]. The residue numbering for the insertion in this protein is as follows: K56 D56a G56b E56c V56d K56e P57 F58... . Urea lysis buffer contains 6 M urea, 100 mM Tris-HCl, pH 8, 1 mM EDTA, 0.5 M NaCl, 0.1% NP-40, and 10 mM BME. KP20 buffer contains 20 mM potassium phosphate, 0.1 mM EDTA, and 1.4 mM BME at pH 7.0.

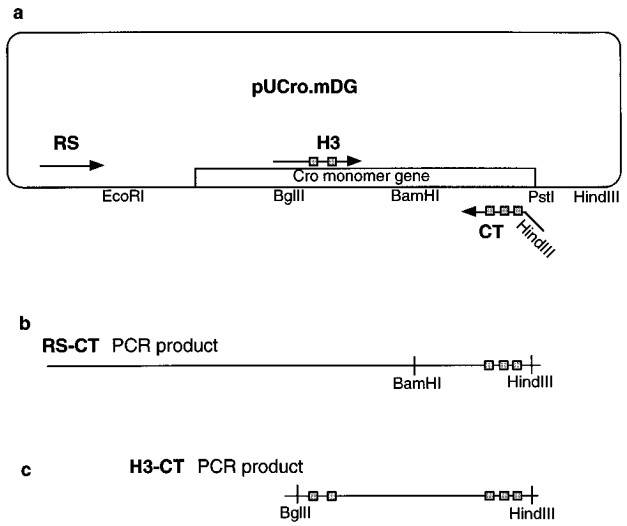

FIGURE 2: Construction of mutant genes. (a) Parental plasmid and PCR primers. H3 introduces changes in codons 29 and 33; CT introduces changes in codons 58, 60, and 61 and truncates the gene prematurely. The RS primer introduces no mutations but allows amplification for the entire Cro gene. (b) The RS−CT PCR product contains randomized codons at 58, 60, and 61 and a stop codon at 62. (c) The H3−CT product contains randomized codons at 29, 33, 58, 60, and 61 and a stop codon at 62.

PCR product of the CT and RS primers were digested with *Hin*dIII and *Bam*HI. Restriction enzymes were inactivated, and ligations were carried out using standard protocols (Maniatis et al., 1982). To help eliminate parental pUC118. mDG plasmids, the ligated products were digested with *Pst*I (the *Pst*I site was eliminated by the CT primer) before being transformed into X90.

Screen and Analysis. Clones were grown overnight in individual wells of 96-well culture plates containing 100 μL of LB broth with 100 μg/mL ampicillin. Each overnight culture was diluted 10-fold into 50 μL of LB with 100 μg/mL ampicillin, again in a 96-well plate, and incubated at 37 °C for 2 h. Protein expression from the plasmid-borne genes was induced by addition of IPTG to a final concentration of 1 mM. Cultures were incubated at 37 °C for another 3 h before the cells were finally mixed with 50 μL of SDS protein loading dye and heated for 5 min at 95 °C. These crude cell lysates were electrophoresed on 20% Tricine gels (Schagger & von Jagow, 1987) and stained with Coomassie Brilliant Blue. mDG protein was run as a standard for size and expression level. Single-strand DNA of selected clones was prepared as described (Vieira & Messing, 1987). Sequencing reactions were carried out with a 21-mer sequencing primer complementary to *lac* Z sequences in the pUC vectors, using the Sequenase II kit (Amersham/USB).

Protein Purification. Proteins were expressed in 2 L cultures of LB broth containing 100 μg of ampicillin/mL at 37 °C with rapid shaking in a 6 L flask. Cultures were induced by the addition of IPTG to 0.5 mM when the absorbance of the culture reached 1.0 at 600 nm. Cells were harvested by centrifugation for 20 min at 3300*g*, resuspended in 25 mL of urea lysis buffer, and sonicated on ice until the residual absorbance due to light scattering at 600 nm was less than 10% of the starting absorbance. A 10% solution of poly(ethylenimine) was added dropwise with stirring to a final concentration of 0.6%. The PEI precipitate and cell debris were removed by centrifugation at 15000*g* for 45 min. Solid ammonium sulfate (0.5 g) was added per mL of the PEI supernatant slowly with stirring. After being stirred for 1 h at 4 °C, the precipitate was collected by centrifugation at 15000*g* for 45 min, resuspended in KP20 buffer plus 6 M urea, and dialyzed overnight against three changes of KP20 buffer (without urea). The dialysate was clarified by centrifugation and applied to a 60 mL Q Sepharose FF column and washed with KP20. The flow-through from this column was applied directly to a 20 mL S Sepharose FF column. Cro monomers eluted at 0.2 M KCl in a 400 mL gradient of 0−1 M KCl in KP20. Cro-containing peaks were concentrated to 3 mL and run on a 120 mL Superdex 70 column in KP20 buffer plus 200 mM KCl. Proteins were >95% pure as judged by gel electrophoresis.

Mass Spectrometry. Proteins dissolved in 50% methanol−water, 2% acetic acid were analyzed by electrospray mass spectrometry on a JEOL-AX 505HA instrument using data accumulation times of 10−30 min, at flow rates of 0.5−1 μL/min. The measured masses agreed to within 3 mass units of the calculated masses for the each of the mutants.

Stability Measurements. CD experiments were carried out on 4 μM protein in KP20 buffer plus 200 mM KCl in a 1 cm cell. The CD signal was monitored at 222 nm while temperature was varied from 10 to 90 °C in 2.5 °C steps with a 2 min equilibration time on an AVIV 62DS Spectrometer equipped with a thermoelectric cell holder. Renaturation of the same protein samples was monitored as they were cooled from 90 to 10 °C following the same protocol. Reversibility was monitored by comparing the ellipticity at 222 nm at 10 °C after a denaturation/renaturation cycle to the initial signal measured under the same conditions. This ratio was always greater than 80%. Linear baselines were extrapolated for the folded and unfolded protein from least squares fits of the low and high temperature points, respectively. The following equation was used to obtain the fraction of protein unfolded (F_u) at a given temperature:

$$F_u = (Y_f - Y)/(Y_f - Y_u)$$

where Y is the CD signal at a particular temperature, Y_f is the extrapolated signal of the folded baseline, and Y_u is the extrapolated signal of the unfolded baseline. T_m is the temperature at which F_u equals 0.5.

Folding free energies were calculated for each of the mutants by the equation

$$\Delta G = -RT \ln\left(\frac{F_u}{1 - F_u}\right)$$

Free energies were calculated for each temperature in the transition zone. Free energies for each mutant at the T_m of Cro K56-[DGEVK] (56.3 °C) were interpolated from least squares fits to the four data points which flank this temperature.

RESULTS

The five residues having the largest relative occupied volumes as determined by the VADAR program (Wishart et al., 1994) in the structure of Cro K56-[DGEVK] (Albright, et al., 1996) were A29, A33, R38, E54, and F58. The solvent accessibility of these residues was calculated as well and ranged from 40% for R38 and 23% for A29 to less than 10% for A33, E54, and F58. R38 and E54 are involved in a salt bridge, which serves to partially bury F58. Codons

Table 1: Summary of the Mutant Screen

	CT primer		H3 + CT primers	
	total	sequenced	total	sequenced
total clones	233	38	392	28
expressing clones	110	37	45	22
expression > mDG	35	21	6	4
expression = mDG	47	12[a]	14	11[b]
expression < mDG	28	4	25	7
nonexpressing clones	123	1	347	6

[a] Includes three parental (Cro.mDG) sequences. [b] Includes six parental sequences.

29, 33, and 58 were randomized to look for alternate packing arrangements.

A total of 625 clones were screened for Cro expression by visualization of crude bacterial lysates by SDS−PAGE. Results of the mutant screen are summarized in Table 1. Clones that had changes only at positions 58, 60, and 61 had a higher frequency of expression (110 out of 233 had some protein expression) than the clones with changes at positions 29 and 33 as well as 58, 60, and 61 (45 out of 392 showed some protein expression) (Table 1). Most of the high expression clones and some moderate and low expression clones as well as a few nonexpressing clones were selected for sequencing. In general, the sequences obtained were statistically consistent with the design of the mutagenic primers. Of the 66 clones sequenced, 57 of them showed the mutant sequence at the expected position while 9 of the clones had the wild-type mDG sequence, probably due to incomplete restriction enzyme digestion in the library construction (Table 2).

Six mutants were selected for purification and stability measurements. We will refer to them by a three-letter designation indicating the identity of residues 29, 33, and 58 as indicated below. The first, K56-[DGEVK] S60K N61K K62term (AAF), was chosen to investigate the effects of C-terminal truncation on stability (Figure 3a). The stabilities of three proteins involving mutations at position 58, K56-[DGEVK] F58L S60K N61term (AAL) , K56-[DGEVK] F58Y S60K N61K K62term (AAY), and K56-[DGEVK] F58W S60K N61K K62term (AAW), are shown in Figure 3b. The stability of two variants containing the stabilizing F58W substitution as well as changes at positions 29 and 33, A29I K56-[DGEVK] F58W S60K N61K K62term (IAW) and A29T A33T K56-[DGEVK] F58W S60K N61R K62term (TTW), are shown in Figure 3c.

Thermal denaturation profiles as monitored by CD show little difference in stability between Cro K56-[DGEVK] (T_m = 56.3 °C) and AAF (T_m = 55.3 °C) (Figure 3a). Among the proteins with wild-type sequences in helix α3, AAW has increased while AAL and AAY have decreased thermal stability (Figure 3b). The protein with the highest stability that we have tested is IAW (Figure 3c). These midpoint temperatures and folding free energies are summarized, along with volume calculations for each of the core residue combinations in Table 3.

DISCUSSION

The first step in designing the mutagenesis scheme was to inspect the structure of the monomer. Figure 1b shows that the separation between the α3 helix and phenylalanine 58 in the monomer structure is larger than is the case in the original Cro dimer structure. The residues specified by the degenerate codons 29, 33, and 58, were generally hydrophobic and covered a range of sizes. Serine and threonine were included because of constraints in the genetic code. In conjunction with the core mutagenesis, the protein was truncated at a series of codons near the beginning of the unstructured C-terminus. Charged amino acids were chosen for the C-terminus to avoid potential problems with proteolysis (Parsell et al., 1990).

Of the 625 clones obtained, 155 showed some degree of protein expression. Proteins with only the C-terminal degeneracy had a higher frequency of protein expression (110 of 233) than proteins which were randomized both in helix α3 and at the C-terminus (45 of 342) (Table 1). This is expected since mutations which are predominately incompatible with function will be more common in populations where more positions are randomized concurrently. The number of mutants analyzed for the C-terminal randomization should

Table 2: Sequences and Expression Levels of Recovered Variants[a]

	residue no.						protein
name[b]	29	33	58	60	61	62	level[c]
mDG	A	A	F	S	N	P	+
	I	V	Y	*			−
	M	A	W	R	*		−
	M	I	C	R	K	*	−
	A	A	*				−
	L	V	W	R	K	*	−
	A	V	W	*			−
	S	V	W	K	K	*	−
	V	I	W	K	K	*	±
	S	V	F	*			±
	A	A	L	*			±
	A	A	C	K	R	*	±
	L	V	L	K	*		±
	I	T	Y	*			±
	L	A	F	K	*		±
	S	M	L	K	K	*	±
	A	A	F	S	N	P	+
	A	A	Y	K	R	*	+
	L	V	W	K	K	*	+
	A	A	F	K	*		+
AAL	A	A	L	K	*		+
	L	T	L	K	K	*	+
	S	T	C	R	K	*	+
	S	V	W	K	*		
	A	A	Y	*[d]			++
	A	A	Y	K	*[d]		++
AAY	A	A	Y	K	K	*[d]	++
	A	A	Y	R	K	*	++
AAW	A	A	W	K	K	*[d]	++
	A	A	W	K	R	*	++
	A	A	W	R	K	*[d]	++
	A	A	W	R	R	*	++
	S	A	W	K	*		++
	A	A	F	*[d]			++
AAF	A	A	F	K	K	*[d]	++
	A	A	F	R	K	*	++
	A	A	K	*			++
IAW	I	A	W	K	K	*	++
TTW	T	T	W	K	R	*	++

[a] Mutants with different residues at positions 29, 33, 58, 60, and 61 along with their expression levels are shown. Asterisks indicate the first stop codon of the mutant gene. [b] Names of the clones described in the text. [c] Clones giving expression levels similar to mDG are indicated with +. Clones giving lower, but still detectable levels are indicated with a ±.. Higher levels are indicated with ++, and no detectable expression is indicated with a (−). [d] Mutant sequences identified in multiple clones.

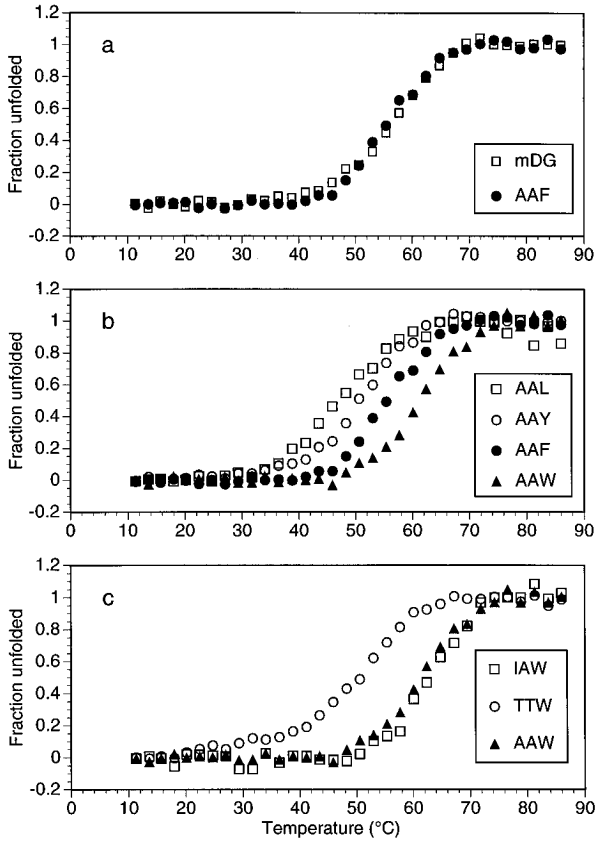

FIGURE 3: Thermal denaturation as followed by CD spectroscopy. (a) Comparison of C-terminal truncation and parental thermal stability. mDG refers to the parental plasmid, K56-[DGEVK]. AAF refers to an identical protein in which the last seven amino acids have been replaced by two Lysines, K56-[DGEVK] S60K N61K K62term. (b) Stabilities of F58 variants. Residues in position 58 are indicated. AAF is described above. AAY refers to Cro mutant K56-[DGEVK] F58Y S60K N61K K62term. AAL refers to Cro mutant K56-[DGEVK] F58L S60K N61term. AAW refers to Cro mutant K56-[DGEVK] F58W S60K N61K K62term. (c) Stabilities of variants at positions 29, 33, and 58. IAW refers to Cro mutant A29I K56-[DGEVK] F58W S60K N61K K62term. TTW refers to Cro mutant A29T A33T K56-[DGEVK] F58W S60K N61R K62term. AAW is described in (b) above. All experiments were performed with 4 μM protein in 20 mM K_2HPO_4, 200 mM KCl, 0.1 mM EDTA, and 1.4 mM BME, pH 7.0, in 1 cm cells using an AVIV 62 DS spectrometer.

statistically yield most possible combinations of C-termini and residue 58 (where the frequency of any given codon is $1/6 \times 1/4 \times 1/4 = 1/96$). That the screen is near saturation in this case is also evidenced by the number of multiple isolates obtained. The number analyzed for the joint H3/CT randomizaton falls short of the number required for complete representation of the randomized library. There are certainly combinations of amino acids that have yet to be screened.

Mutants from Table 2 were chosen for analysis of thermal stability by CD to answer three questions: (1) How is the stability of the protein affected by truncations of the C-terminus? (2) How is the size of residue 58 correlated with stability? (3) How do residues at 29 and 33 affect the thermal stability in the context of W58?

The thermal denaturation behavior of a truncated F58 variant is illustrated in Figure 3a. It is clear that the replacement of the last seven amino acids by a pair of lysines has very little effect at the level of thermal stability of the purified proteins. More work will be required to ensure that the nature and/or the presence of residues at 61 and 62 do not have slight effects on the stability of Cro monomers. It is somewhat surprising that several of the proteins which are marginally less stable than wild type *in vitro* are expressed at higher levels *in vivo* (AAY, TTW). One potential reason for the higher accumulation of these proteins might be that the more highly charged C-terminal residues make the proteins more resistant to proteolysis. This effect has been seen for other small repressor proteins (Parsell et al., 1990).

The long-standing observations of Richards (1977) and others that gaps in protein interiors are rare led us to suspect that the thermal stability of the Cro monomers which had already been increased dramatically over the dimer (at physiological concentrations—by virtue of its unimolecular folding) might be increased further by judicious choice of residues which might better fill the internal space in the core of the protein. A number of the mutants isolated have increased stability, as monitored by protein accumulation *in vivo* or thermal stability *in vitro*. Tryptophan substitutions at position 58 have the most dramatic effects, raising the T_m of the protein by ~5 °C. A 1.3 °C increase in *T*m over that of AAW can be achieved in conjunction with an A29I substitution.

Analysis of the molecular origins of the effects of different amino acid combinations on stability must take into account both the hydrophobicity of the substituted residues and their ability to pack into the available space in the core of the protein. Although a complete analysis awaits structural information for each of the mutant combinations, Table 3 lists some pertinent values for comparison of the volumes and hydrophobicities of the mutants we have studied. The

Table 3: Volume and Energetics of Repacking Mutants

	mutant designation						
	mDG	IAW	AAW	AAF	TTW	AAY	AAL
volumes for Residues 29, 33, and 58							
occupied volume[a] (Å^3)	434.2						
average volume[b] (Å^3)	370.1	481.5	405.4	370.1	464.6	365.7	337.7
excess volume (Å^3)	64.1	−47.3	28.8	64.1	−30.4	68.5	96.5
energetics							
T_m (°C)	56.3	62.6	61.3	55.3	52.8	51.8	46.8
ΔG at 56.3 °C[c] (kcal/mol)	0	0.97	0.74	−0.17	−0.54	−0.67	−1.13
$\Delta\Delta G$ of transfer[d] (kcal/mol)	0	1.95	0.46	0	0.36	−0.83	0.09

[a] The sum of the occupied volumes for residues 29, 33, and 58 was calculated for the reference structure only, as indicated in the text. [b] Atomic radii for volume calculations were those of Shrake and Rupley (1973). [c] Unfolding free energies measured at the T_m of mDG have standard deviations of approximately 0.1 kcal/mol. [d] The hydrophobic transfer free energies were calculated as differences from the sum of the transfer free energies for alanine, alanine, and phenylalanine for the set of three residues at 29, 33, and 58 using the scale of Fauchere and Pliska (1983).

sum of the volumes accessible to A29 (108 Å^3), A33 (103 Å^3), and F58 (223 Å^3) in the monomer structure (Albright et al., 1996) is 434 Å^3. This exceeds the average volume occupied by these three residues by 64 Å^3. As listed in Table 3 the AAW combination should fill the available space more completely while the IAW and TTW combinations might be expected to overfill the available space. Of course, these are very crude approximations which fail to take into account the detailed geometries of the side chains, excess space attributed to other core residues by the Voroinoi polyhedral algorithm, the fact that A29 is partially exposed in the monomer structure, and of course any rearrangements which might take place. One might expect the maximum stabilizing effect to come from combinations of the three most hydrophobic residues which can pack in the available space.

Bearing in mind the above caveats, we can begin to rationalize the relative stabilities of the mutants as follows. IAW might be expected to overfill the core somewhat, but the fact that A29 is partially exposed may allow the isoleucyl side chain to extend into the solvent. AAW has an acceptable volume and ~0.5 kcal worth of hydrophobic stabilization. TTW has too large a volume and more limited hydrophobic benefit. AAY has nearly equivalent packing to AAF, but the hydrophobic penalty of burying the tyrosine hydroxyl approximately equals the stability deficit of this mutant. AAL has an overall hydrophobicity which closely matches that of AAF, but the underpacking of the core is likely responsible for its instability.

The subtle packing adjustments observed in comparing the original "tetramer" crystal lattice of wild-type Cro (Anderson et al., 1981) to the structure of the Cro K56-[DGEVK] monomer may be important in DNA recognition. Recent work (R. A. Albright and B. W. Matthews) on the structure of Cro dimers bound to its operator site has shown an expansion of the core and shift of helix α3 outward from the β-sheet, similar to that seen in the Cro K56-[DGEVK] structure. Since the α3 helix is the primary locus of site-specific contacts between amino acid side chains on the protein and bases in the major groove of DNA, it will be interesting to see whether any of the alternative packing arrangements that we have observed will either enhance binding by locking the structure into its "DNA recognition" conformation or alternatively inhibit binding by making the structure too rigid to adapt to the DNA surface. Experiments are underway to repeat core randomization in the context of the wild-type Cro dimer. In this case there is a convenient selection to identify functional proteins by their ability to repress a conditionally lethal gene in *E. coli* (Mossing et al., 1991).

CONCLUSIONS

We have sought to probe the expansion of the hydrophobic core of an engineered protein observed in the high-resolution crystallographic structure by combinatorial mutagenesis of core residues. Variants with increased stability have been isolated which will be interesting for further studies of the correlation of protein sequence with structure, stability, and DNA recognition. This iterative process, coupling structural inputs to the efficient search of sequence space by combinatorial mutagenesis, is a powerful methodology for protein engineering.

ACKNOWLEDGMENT

We thank Brian Matthews and Dale Tronrud of the University of Oregon for communication of their unpublished Cro structure. David Wishart and Brian Sykes of the University of Alberta kindly provided the VADAR protein structure analysis package. We also thank Graham Lappin for use of the CD instrument and Bill Bogess for the mass spectroscopy measurements.

REFERENCES

Albright, R. A., Mossing, M. C., & Matthews, B. W. (1996) *Biochemistry 35*, 735−742.

Amann, E., Brosius, J., & Ptashne, M. (1983) *Gene 25*, 167−178.

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

Anfinsen, C. B. (1973) *Science 181*, 223.

Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., & Matthews, B. W. (1992) *Science 255*, 178−183.

Fauchere, J.-L., & Pliska, V. (1983) *Eur. J. Med. Chem. 18*, 369−375.

Hubbard, A. J., Bracco, L. P., Eisenbeis, S. J., Gayle, R. B., Beaton, G., & Caruthers, M. H. (1990) *Biochemistry 29*, 9241−9249.

Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mossing, M. C., & Sauer, R. T. (1990) *Science 250*, 1712−1715.

Mossing, M. C., Bowie, J. U., & Sauer, R. T. (1991) *Methods Enzymol. 208*, 604−619.

Parsell, D. A., Silber, K. R., & Sauer, R. T. (1990) *Genes Dev. 4*, 277−286.

Richards, F. M. (1974) *J. Mol. Biol. 82*, 1−14.

Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151−175.

Schagger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368−379.

Shrake, A., & Rupley, J. A. (1973) *J Mol Biol* 79, 351−71.

Sibanda, B. L., Blundell, T. L., & Thornton, J. M. (1989) *J. Mol. Biol. 206*, 759−777.

Vieira, J., & Messing, J. (1987) *Methods Enzymol. 153*, 3−11.

Wishart, D., Willard, L., & Sykes, B. (1994) VADAR, Version 0.9, Department of Biochemistry, University of Alberta, Edmonton, AB, Canada.

BI951959F