

# Linking an Easily Detectable Phenotype to the Folding of a Common Structural Motif

## Selection of Rare Turn Mutations that Prevent the Folding of Rop

Luisa Castagnoli, Costantino Vetriani and Gianni Cesareni†

*Dipartimento di Biologia  
Università di Roma, Tor Vergata, Via Carnevale, 00173 Roma, Italy*

Rop is the simplest and most regular member of a family of proteins characterized by a bundle of four antiparallel helices. Rop is dimeric, each monomer being formed by two helices connected by a sharp bend. In this work we have extensively mutagenized three residues that form the connection between the two  $\alpha$ -helices to ask whether the bend region contains any important folding information. The characterization of a collection of random mutants indicated that this structure is rather insensitive to amino acid substitutions and that most amino acids are tolerated in these positions by the Rop native structure.

In order to identify the rare amino acid sequences that would prevent Rop from folding and/or dimerizing, we exploited the observation that Rop can functionally substitute the dimerization domain of the  $\lambda$  repressor. In fact plasmids expressing a hybrid protein formed by the amino-terminal domain of the  $\lambda$  repressor covalently linked to Rop, confer immunity to  $\lambda$  infection on their hosts. We have shown that this property depends on the ability of the Rop moiety to fold and dimerize. The analysis of 380 Rop mutants containing random amino acid sequences at positions 30, 31 and 32 allowed us to identify three mutant Rop proteins that are defective in dimerization, probably as a consequence of their inability to fold. In these mutants the tripeptides VED, VPD and YPD substitute the wild-type DAD at positions 30, 31 and 32. Other combinations of amino acids are found resulting in levels of immunity that are lower than the wild-type but still sufficient to prevent single plaque formation. This result suggests that a smaller proportion of the corresponding Rop protein reaches a thermodynamic and proteolytically stable dimeric state.

*Keywords:* four helix bundle; gene fusion; loops; mutagenesis; dimer

### 1. Introduction

Prediction of protein conformation from amino acid sequence is a difficult task. At present the only reliable method for predicting a low resolution structure of a peptide depends on the detection of amino acid sequence similarity with a protein whose structure has been determined by direct methods (Pearl & Taylor, 1987; Greer, 1981). In the absence of other information, in the alignment procedure each position is given the same weight, although it is known that different positions have a different degree of tolerance to amino acid substitutions (Bowie *et al.*, 1990). It is therefore desirable to identify in each conformational class those residues that are less tolerant to substitutions in order to be able to give them a prominent weight in the alignment algorithm (Taylor, 1986). In the case of the few proteins for which sequences from many

different species are known, this information can be obtained by aligning the different naturally occurring sequences and observing the degree of conservation at each position (Hampsey *et al.*, 1988; Bashford *et al.*, 1987; Lesk & Chothia, 1980). Alternatively, genetic methods can be applied to cloned genes to introduce specific or random changes along the protein and to compile a collection of amino acid sequences that are compatible with a given conformation (Lim & Sauer, 1991). We have adopted this latter approach to reveal, if any, the sequence determinants of the turn that connects  $\alpha$ -helices in a family of proteins characterized by a bundle of four antiparallel helices.

Turns may be defined as those sites in a protein where the polypeptide chain reverses its direction (Rose *et al.*, 1985). This definition includes tight turns, which can often be assigned to a limited number of classes of rather regular structures, and the longer and less regular 'loops' that connect structural elements of repetitive secondary struc-

† Author to whom all correspondence should be addressed.

ture ( $\alpha$ -helices and  $\beta$ -strands). Turns are rather abundant in globular proteins where they constitute approximately one-third of the molecule. Since they are located on the surface of proteins, they often play important functional roles contributing to the formation of catalytic and binding sites (Rose *et al.*, 1985). Most loop amino acid sequences, however, are poorly conserved in evolutionary related peptides suggesting that, whenever they are not intimately involved in protein function, they can vary widely in sequence and length without affecting protein conformation and function (Leszczynski & Rose, 1986). This conclusion is supported by the observation that loop structures develop late in the process of folding of the protein barnase (Matouschek *et al.*, 1990) and by *in vitro* mutagenesis studies (Starzyk *et al.*, 1987; Castagnoli *et al.*, 1989; Fetrow *et al.*, 1989; Wang & Pielak, 1991). Some other theories, however, sustain that loops either play an active role in protein folding, forming early and pulling together elements of secondary structure (Lewis *et al.*, 1971; Zimmerman & Sheraga, 1977; Dyson *et al.*, 1988), or contribute substantially in stabilizing the equilibrium conformation (Carlacci & Chou, 1990). In a recent study, by utilizing a simple and sensitive genetic test, we have shown that a high proportion of mutant proteins, containing an exapeptide of random amino acid sequence in place of a six residue loop in the H chain of human ferritin, can influence the percentage of chains folding into the correct conformation (Jappelli *et al.*, 1992). However, the interpretation of these results is complicated by the relative complexity of the

monomer conformation and by the multimeric nature of the ferritin 24mer.

In the past few years we have studied the conformational determinants of a simple and rather common structural motif: the bundle of four antiparallel  $\alpha$ -helices. As a paradigm of this class of proteins we have selected its simplest and most regular representative: the small bacterial protein Rop. Rop is dimeric and each monomer folds into two antiparallel helices connected by a sharp bend (Banner *et al.*, 1987). Two monomers assemble with a 2-fold symmetry axis to yield the native dimer (Figure 1, left).

In Rop the bend region is centered on Ala31, which is in the  $\beta$  conformation ( $\phi = -94$ ,  $\psi = 89$ ) and is the only residue whose  $\alpha$ -carbon is outside the idealized cylinders that contain helix 1 and helix 2 (Figure 1, right). However, the regular  $\alpha$  conformation is already broken at Leu29 ( $\phi = -90$ ,  $\psi = -1$ ), while Asp30 ( $\phi = 57$ ,  $\psi = 38$ ) is in  $\alpha_L$  conformation and its N-H group hydrogen bonds to the C-O groups of both Leu26 and Asn27. Asp32 starts the second helix and its side-chain might help in stabilizing it by hydrogen bonding the backbone amino group of Glu33. Thus, Asp30 and Asp32 may be considered as the C-cap and N-cap residues of helix 1 and helix 2, respectively (Richardson & Richardson, 1988).

In the work presented here we have concentrated our attention on the central residue Ala31 and on the two flanking Asp residues, and we have asked whether their side-chains play any essential role in the processes of protein folding and dimerization.

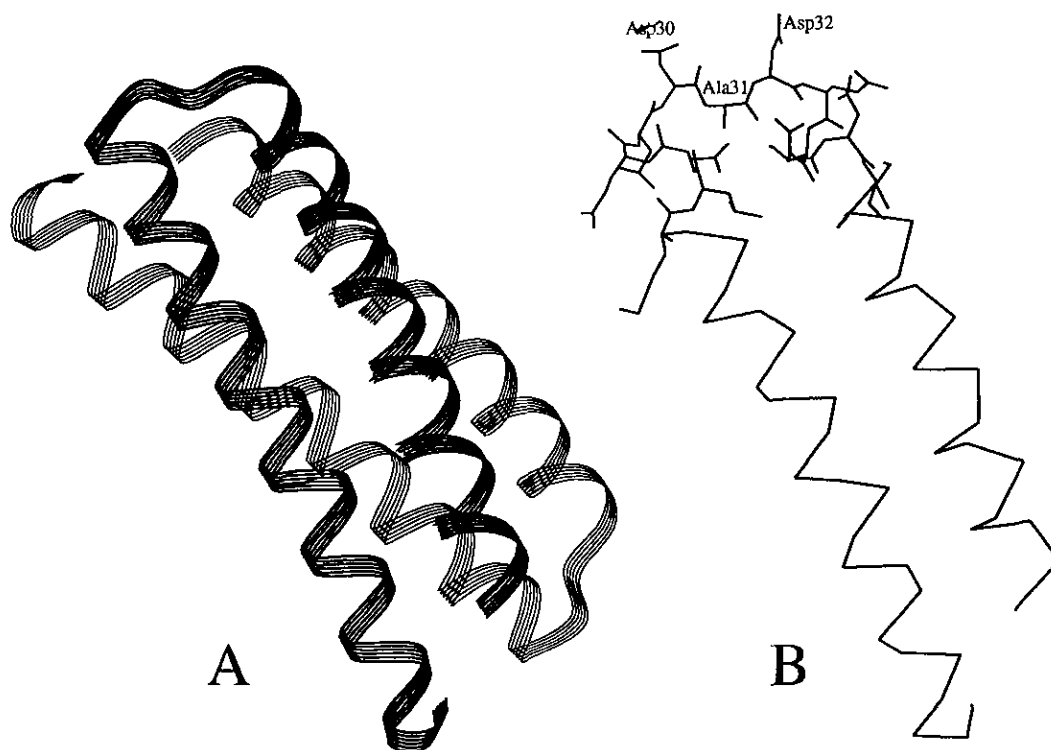


Figure 1. Rop structure. A, Ribbon representation of the Rop dimer. B, Details of the bend region of one Rop subunit.

## 2. Materials and Methods

### (a) Bacterial strains and recombinant DNA methods

*Escherichia coli* 71/18 ( $\Delta$ lac-pro/F<sup>+</sup>[lacI<sup>q</sup> lacZAM15 proAB<sup>+</sup>] supE; Messing *et al.*, 1977) is the host bacterium utilized for most constructions and experiments. When necessary, plasmids carrying amber mutations were transformed into GC76 ( $\Delta$ lac-pro/F<sup>+</sup>[lacI<sup>q</sup>lacZAM15 proAB<sup>+</sup>] sup<sup>0</sup>). 71/72 and GC382 are derivatives of 71/18 and GC76 and contain plasmid pcI857, which carries the temperature-sensitive allele of the  $\lambda$  repressor cI857 and confers resistance to the antibiotic kanamycin (Remaut *et al.*, 1983). Strains containing the pcI857 plasmid were used to host pEX58 derivatives in order to control the  $\lambda$ pL promoter in front of the Rop gene. Rop synthesis is obtained by raising the temperature to 42°C and incubating for 3 h at the same temperature. Alternatively, since pEX58 is a phagemid and can be encapsidated into filamentous phage rods by infection with a helper phage, phagemid supernatants were used to infect, at a multiplicity of 10, exponentially growing male bacteria. Once the DNA is injected into the host and made double-stranded by bacterial enzymes, Rop gene transcription starts from the  $\lambda$ pL promoter since no  $\lambda$  repressor is present. Microbiological techniques, recombinant DNA and DNA sequencing were according to standard protocols (Sambrook *et al.*, 1989).

### (b) Plasmid construction

Rop expression plasmid pEX58 derives from pEX43 (Castagnoli *et al.*, 1989). We first eliminated the *Nde*I site near the fl replication origin by cutting with *Nde*I, filling in the protruding ends with the Klenow subunit of DNA polymerase and ligating at low plasmid concentration. To facilitate cassette mutagenesis we introduced 2 unique restriction enzyme recognition sites (*Nde*I and *Sac*I) flanking the bend region without altering the Rop amino acid sequence. This plasmid was named pEX51. The third step consisted of the construction of pEX52 that contains the fusion of the Rop and  $\alpha$ -peptide coding sequences, separated by an amber codon. Finally, to obtain pEX58 we introduced a TAA nonsense codon in place of the codon corresponding to Asp32 in the wild-type sequence.

PC132 and PC135 are plasmids that direct the synthesis of the cI-Rop fusion protein under the control of the pLac promoter.

We initially assembled the cI-rop fusion gene by inserting the Rop coding sequence between the *Sal*I and the *Bam*HI sites of plasmid pJH391 (Hu *et al.*, 1990). A Rop gene fragment suitable for precise in-frame fusion was amplified in the presence of the *Taq* polymerase and oligonucleotides 170 TTCCAGTCAACGACGTTGTAAAA-CG and R75 TTACACGGTCTGACTCAAATGACCAAAC that hybridize to the 3' and 5' end of the rop gene, respectively. Oligo R75 eliminates the Rop Shine-Dalgarno sequence, changes the initiator codon GTG into an ATG and introduces an in-frame *Sal*I site.

The resulting plasmid PC141 confers  $\lambda$  immunity on the host bacterium. However, since this plasmid contains a second wild-type rop gene whose product could interfere with the  $\lambda$  immunity phenotype by ethero-dimerizing with the  $\lambda$ cI-Rop fusion protein, we moved the fusion gene into a different plasmid background. The *Eco*RI-*Eco*RV DNA restriction fragment of PC141, containing the fusion gene, was inserted between the *Eco*RI and the *Nae*I sites of PC72. PC72 is a pEMBL18 derivative (Dente *et al.*, 1983) that had been deleted of a *Cla*I-*Aat*II fragment containing the sequence encoding

the carboxy terminus of the rop gene. The resulting plasmid PC145 confers  $\lambda$  immunity on the host bacterium. In order to facilitate identification of mutants after mutagenesis experiments, we further modified the  $\lambda$ cI-rop gene fusion by fusing the  $\beta$ -galactosidase  $\alpha$ -peptide gene to the 3' end of the Rop gene. The 2 coding sequences, however, are separated by an amber codon to allow for synthesis of a bipartite (cI-Rop) or tripartite (cI-Rop- $\alpha$ -peptide) fusion protein depending on the suppressor background of the host. This plasmid is called PC132. PC135 differs from PC132 only in the position encoding Asp32 of Rop, substituted by a TAA nonsense codon. supE and sup<sup>+</sup> strains containing this last plasmid are  $\lambda$ -sensitive and form white colonies on X-gal<sup>†</sup> indicator plates. The structures of pEX58 and PC135 are illustrated in Figure 2.

### (c) Mutant construction and selection

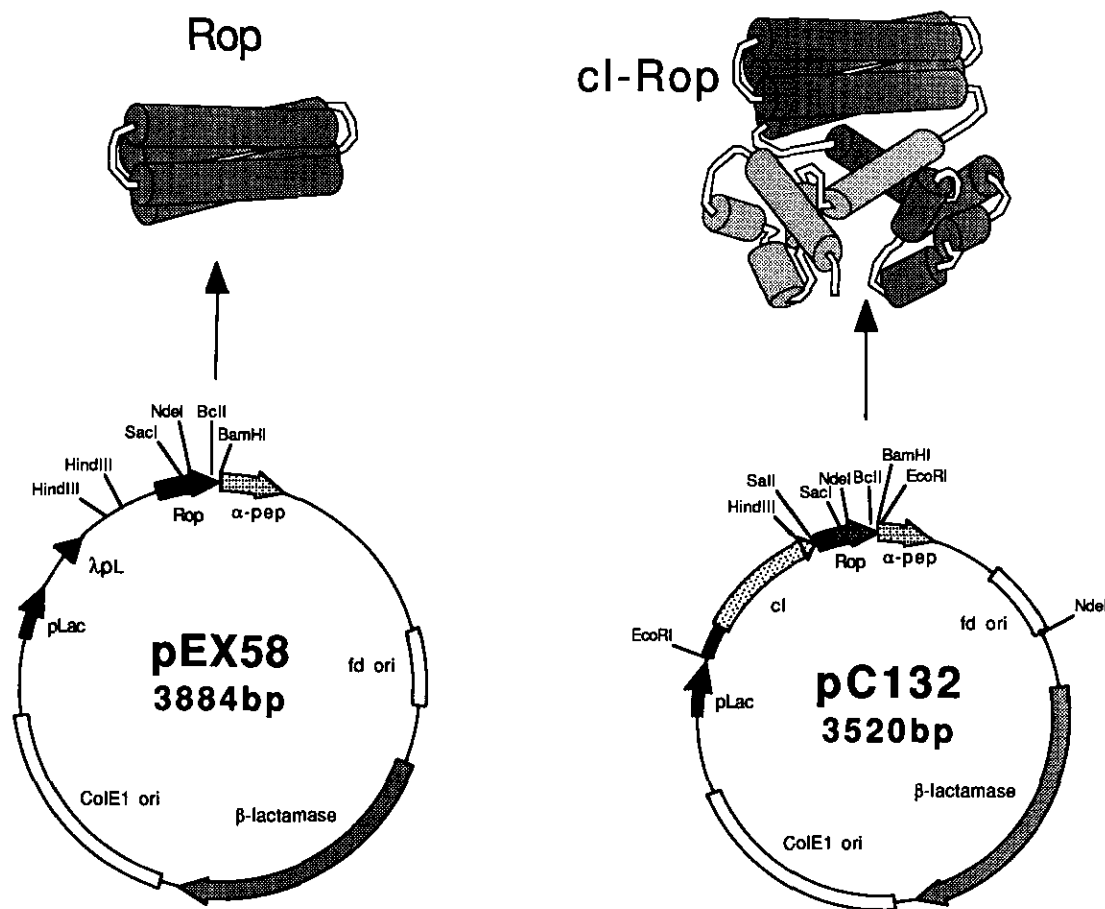
In order to construct a collection of Rop mutants containing different tripeptides in the bend region we used an oligonucleotide of sequence CTGCCTGTTC.ATCCGCGTTC.GAGCTCGTTGAGT (R25). This oligonucleotide is complementary to the sequence encoding the carboxy terminus of helix 1, the bend region (underlined) and the amino terminus of helix 2. During the synthesis, however, the underlined nucleotide precursors were contaminated with 30% of a mixture of the remaining 3 nucleotides. Thus, because of the degeneracy of the genetic code, that differs for Ala and Asp, there is a probability of 61% to obtain, after mutagenesis, a substitution at position D30 and D32 and a probability of 51% to have a substitution at position A31. In practice we observed that the actual percentage of mutants containing single or multiple substitutions depends heavily on the hybridization conditions used in the mutagenesis experiment, high oligonucleotide concentration and long hybridization times favouring the occurrence of mutants with a small number of base substitutions. After oligonucleotide-directed mutagenesis and transformation of 71/18, approximately 10% of the colonies were blue, indicating that the TAA codon at position 32 was substituted by a sense codon. Blue colonies were collected on a plate and tested by replica-plating for the ability to grow on plates containing  $2 \times 10^7$   $\lambda$  KH54 phages carrying a deletion mutation in the cI gene.

## 3. Results

### (a) Random mutations in the Rop bend

As a first approach to identify those amino acid sequences which, once substituted into the Rop bend, would be incompatible with the formation of (or would destabilize) the four helix bundle, we constructed a collection of expression plasmids that direct the synthesis of mutant Rop proteins where the three amino acids Asp30Ala31Asp32 are substituted by different residues. Since we did not expect in our experiments to cover a large fraction of the possible combination of random amino acid sequences (8000) we preferred to statistically bias our collection of mutants to favour mutants with a small number of changes. To enrich for mutants containing only one or two amino acid substitutions we designed a mutagenic oligonucleotide that, in

<sup>†</sup> Abbreviation used: X-gal, 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside.

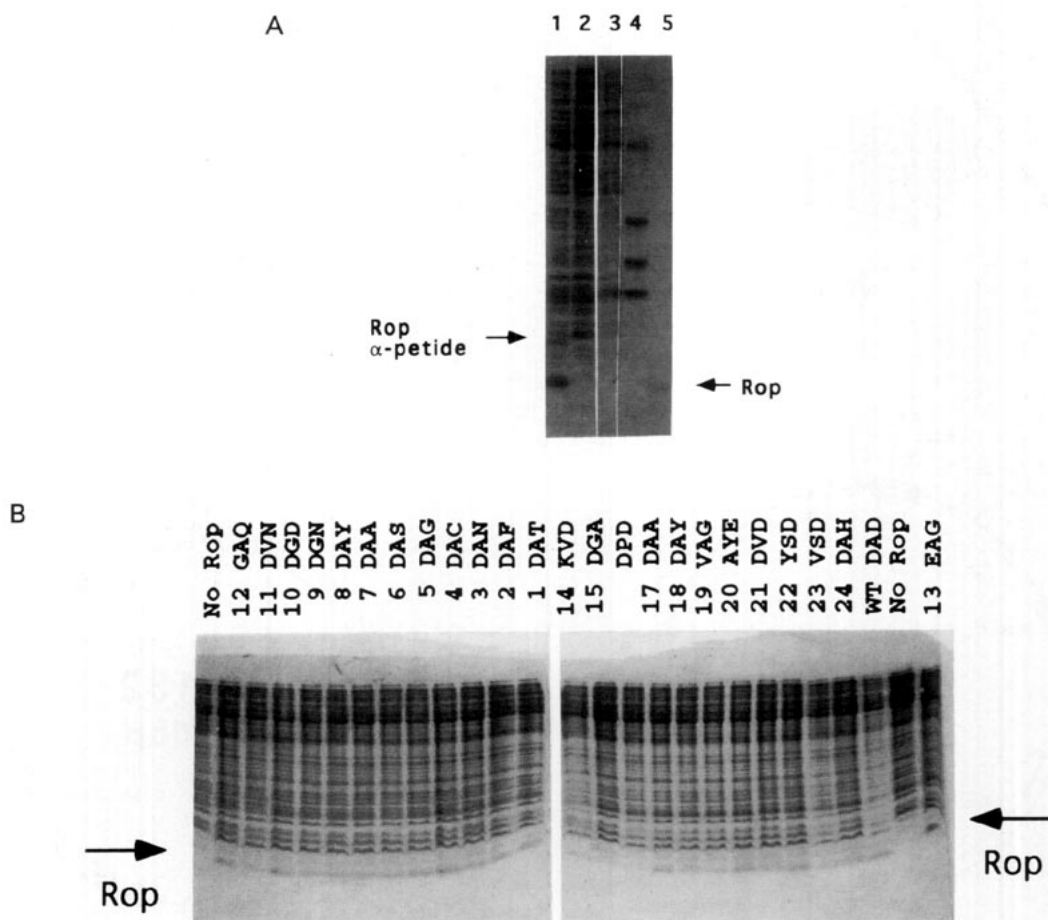


**Figure 2.** Genetic and restriction map of the 2 plasmids that were used in this work. The detail of the constructions are in Materials and Methods. Above the plasmid maps we have drawn a schematic representation of the structure of the proteins synthesized by each plasmid under the control either of the  $\lambda$ pL or of the pLac promoters.

each position of the sequence encoding the tripeptide Asp30Ala31Asp32, contains either the wild-type nucleotide or any of the remaining nucleotides in a ratio of 7:3. The mutagenic oligonucleotide R25 was used to mutagenize the Rop expression plasmid pEX58. As described in more detail in Materials and Methods, this plasmid contains an ochre (TAA) codon at position Asp32 of Rop, a feature that allows one to discriminate between transformants containing a plasmid derived from a non-mutagenized template and clones expressing a mutated Rop chain.

After mutagenesis of the pEX58 template, 32 independent mutant colonies were picked and the sequence of the *rop* gene was determined. In Figure 5A we have reported the deduced amino acid sequence of the bend region of these mutants and of some other mutants obtained by site-directed mutagenesis. In an earlier work (Castagnoli *et al.*, 1989) we observed a strict correlation between the yield of a mutant protein after induction of its expression at 42°C and its thermostability. All the mutant proteins that are recovered in large amounts in the soluble fraction can be shown to fold into a structure very similar to the Rop helix bundle and to have a thermodynamic stability comparable to the

wild-type. By contrast mutant proteins that are either not accumulated after thermal induction at 42°C or are recovered with poor yields are either unable to fold or very thermo-unstable. These conclusions seem rather general and applicable to other proteins (Lim & Sauer, 1991). We have used this criterion to discriminate between the phenotype of the different mutant proteins. Electrophoretic analysis after induction of protein expression (see for instance Figure 3) indicates that all the Rop proteins that have been mutated in the loop region maintain the ability to fold properly and to resist protease degradation. Minor fluctuations in protein yields are not considered in this analysis. The only exception to this conclusion is the mutant Ala31-Pro which was constructed by site-directed mutagenesis and is only produced at 37°C (Figure 3), whereas at 42°C it is degraded by bacterial proteases. We have also checked whether any of the mutant proteins would form inclusion bodies and precipitate with the bacterial debris after low speed centrifugation of a cell extract. None of the *rop* mutants tested accumulates as an insoluble aggregate as demonstrated for some folding-assembly mutants in other systems (Haase-Pettingel & King, 1988; Jappelli *et al.*, 1992).



**Figure 3.** Expression of mutant Rop proteins at 37°C. PAGE analysis of extracts of bacterial cells directing the synthesis of various Rop derivatives. The bands corresponding to the Rop protein are indicated by arrows. Rop expression under the control of the  $\lambda$ pL promoter was induced by infecting 71/18 strain (multiplicity of infection of 10) with a supernatant of phagemid pEX52 or of its *rop* mutant derivatives. The infected bacteria were further incubated for 2 h at 37°C. A, lane 1, 71/18 infected by pEX52 (W.T. Rop); lane 2, 71/18 infected by pC72 (no Rop); lane 3, uninfected 71/18; lane 4, MW markers, 14.3, 18.4, 24, 34.7, 45 KDa, respectively; lane 5, purified Rop. B, Extracts of 71/18 bacterial cells infected with pEX52 derivatives carrying mutated Rop genes. Each mutant is identified by a number and by the amino acid sequence of the bend region in the single letter code.

(b) *Rop can substitute the dimerization domain of the  $\lambda$  repressor*

The results reported in the previous section indicate that the Rop structure is rather insensitive to perturbations caused by amino acid substitutions in the bend and that most tripeptides are compatible with the sharp change of direction of the  $\alpha$ -carbon backbone that connects the two helices. It is possible, however, that at least some of the combinations of three residues are unfavorable to Rop folding either kinetically or thermodynamically. These "unfavorable bends", could be rare and might only be detectable by screening or selection procedures that permit the rapid analysis of a large number of clones for Rop folding and dimerization. The Rop natural function, control of plasmid copy number, cannot be used for this purpose since it is negatively affected by amino acid substitutions in the bend region (Castagnoli *et al.*, 1989; Cesareni *et al.*, 1991). To overcome this problem we took advantage of the observation that the carboxy-

terminal dimerization domain of the  $\lambda$  repressor can be functionally substituted by a protein able to dimerize (Hu *et al.*, 1990). Plasmid PC132 directs the synthesis of a fusion between the amino-terminal DNA binding domain of the  $\lambda$  repressor and the Rop protein under the control of the *lac* promoter. In this plasmid, as in the previously described pEX52 plasmid, the Rop coding sequence is fused *via* an amber codon to the DNA sequence encoding the  $\alpha$ -peptide of the enzyme  $\beta$ -galactosidase. Thus, expression of the  $\lambda$ cI-Rop hybrid protein can be conveniently monitored by measuring  $\beta$ -galactosidase activity in a suppressor strain. PC135 differs from PC132 only in the sequence encoding residue 32 of Rop, where the GAT encoding Asp (PC132) is substituted by a TAA ochre codon (PC135). A strain containing PC135 forms white colonies on lactose indicator plates, since the TAA codon is hardly suppressed by the *supE* allele and the downstream  $\alpha$ -peptide gene is not translated. Furthermore, this strain is sensitive to  $\lambda$  infection since residues 1 to 31 of Rop do not help

the amino-terminal domain of the repressor to dimerize. By contrast the *E. coli* strain containing the plasmid that synthesizes the fusion with the entire Rop molecule (PC132) is insensitive to infection by  $\lambda$ KH54 (carrying a deletion in the *cI* gene) while it is lysed by a virulent strain of  $\lambda$ . These results demonstrate that Rop can functionally substitute the amino-terminal domain of the repressor and provides us with a convenient screening procedure to search for rare Rop folding mutants.

(c) *Linking Rop folding to  $\lambda$  immunity*

To convince ourselves that the immunity phenotype, conferred by the *cI*-Rop fusion protein on the host, depends on the ability of the Rop domain to fold and dimerize, we constructed PC132 derivatives where the wild-type *rop* gene was substituted by three mutant genes whose protein product stability had been characterized by microcalorimetric methods (Weber, 1991). These are mutants constructed with the aim of destabilizing the Rop structure by introducing "hydrophobic holes" (L48V and L48A) or polar side-chains (L41N) in the core of the molecule. Purified L48V Rop mutant ( $t_m = 54^\circ\text{C}$ ) has a thermal stability not very different from the wild-type ( $t_m = 69^\circ\text{C}$ ). By contrast L48A is less resistant to thermal denaturation and unfolds at  $43^\circ\text{C}$ . L41N cannot be over-produced in bacteria, either because it is unable to fold or because its folded conformation is very sensitive to protease degradation. As demonstrated in Figure 4, the thermodynamic properties of the Rop variants correlate with the thermal dependence of

the ability of the corresponding hybrid *cI*-Rop protein to confer  $\lambda$  immunity on the host bacterium. *cI*-RopL48V is indistinguishable from the wild-type at all temperatures tested, while *cI*-RopL41N is not active at any of the temperatures. In accordance with the thermodynamic properties of its Rop domain, *cI*-RopL48A has a temperature-sensitive phenotype, being completely active at  $30^\circ\text{C}$  but gradually losing its activity when the temperature is raised to  $37^\circ\text{C}$  and then to  $42^\circ\text{C}$ .

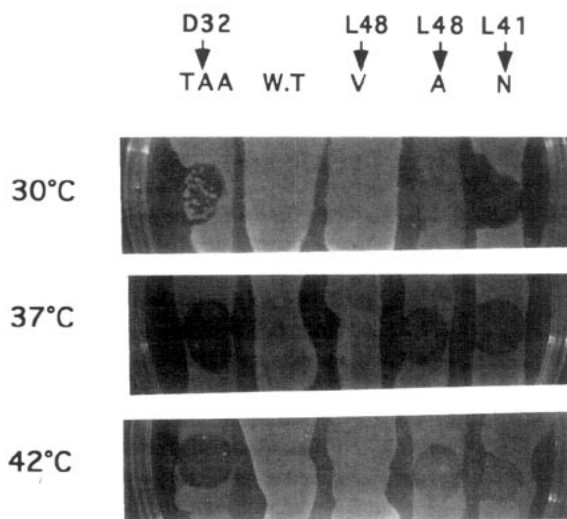
As a further test of the screening procedure we developed, we constructed plasmids expressing *cI*-Rop fusion proteins carrying, in the bend region, some of the mutant peptides that we had previously characterized. Consistent with the conclusions we had already drawn, most of these plasmids were as efficient as the wild-type in conferring immunity to  $\lambda$  infection, the only exception being mutant *cI*-RopDPD that, although active at  $30^\circ\text{C}$  and  $37^\circ\text{C}$ , loses its activity completely at  $42^\circ\text{C}$ . As already discussed RopDPD is the only bend mutant that could not be efficiently expressed by raising the temperature to  $42^\circ\text{C}$ .

(d) *Selection of rare bend mutations that prevent Rop folding*

The results of the experiments discussed in the previous section demonstrate that the  $\lambda$  immunity phenotype of the hybrid *cI*-Rop protein depends on the ability of the Rop domain to fold and dimerize and gave us confidence in using this test to look for rare amino acid combinations which, inserted in the Rop bend, would destabilize the four helix bundle.

To isolate such mutants, we repeated the mutagenesis experiment utilizing oligonucleotide R25 to alter the bend region of plasmid PC135. After transformation and selection on plates containing ampicillin and X-gal, approximately 5% of the colonies turned blue, indicating that the mutagenic oligonucleotide had substituted the three codons in the bend region with sense (or suppressible) codons. The DNA sequence of the *rop* genes in 25 randomly selected blue colonies was determined by the dideoxy method and the corresponding sequence of the Rop bend was deduced (Figure 5B). The observed substitution frequencies in each of the nine positions were 20%, 28%, 26%, 28%, 45%, 32%, 20%, 21%, 9% respectively, to be compared with the theoretical 30% "contamination" in each of the nine positions of the mutagenic oligonucleotide. In the statistical limits of the small sample we observed only minor variations from the expected frequency of each base in the different positions.

A total of 380 blue colonies were replica-plated onto three plates containing  $2 \times 10^7$   $\lambda$  particles and incubated at  $30^\circ\text{C}$ ,  $37^\circ\text{C}$  and  $42^\circ\text{C}$ . Among all the colonies tested, nine grew less vigorously than the wild-type when incubated (at least at one of the three temperatures) in the presence of  $\lambda$ , indicating that the mutated Rop protein was less efficient than wild-type in promoting the dimerization of the DNA binding domain of the  $\lambda$  repressor. When



**Figure 4.** Temperature-sensitive mutants. 71/18 bacteria containing PC132 derivatives differing in the sequence of the *rop* gene were grown to saturation. 50  $\mu\text{l}$  of bacterial culture were mixed with 1 ml of top agar and then streaked on ampicillin plates. 5  $\mu\text{l}$  of suspension of phage  $\lambda$ KH54 at a concentration of  $10^7$  phages/ml were spotted onto each agar streak. The plates were then incubated overnight at different temperatures.

A		B		C	
	30 31 32		30 31 32		30 31 32
Wt	<b>AspAlaAsp</b>	WT	<b>AspAlaAsp</b>		<u>AlaAlaAla</u>
28	Asn	6	Glu		<u>ValValVal</u>
31	Ser	22	Ala		<u>GlvGlvGlv</u>
Rop125	Cys	27	Trp		<u>GluGluGlu</u>
<b>Rop69</b>	<b>Pro</b>	28, 56	Gly		<u>SerSerSer</u>
Rop140	Leu	<b>42, 52, 23</b>	<b>Glu</b>		<u>TyrTyrTyr</u>
10	Gly	44	Ser		<b>Asp Asp</b>
21	Val	26	Asn		<u>Asn Asn</u>
8	Tyr	45	His		LysLys
2	Phe	50	Ser		<u>ProPro</u>
Rop130	Trp	9	GluThr		<u>Thr</u>
Rop129	Glu	20	TyrGly		Arg
7, 17	Ala	51	GluVal		Leu
1	Thr	54	GluPro		<u>His</u>
3	Asn	<b>33</b>	<b>ValPro</b>		Cys
6	Ser	<b>2</b>	<b>TyrPro</b>		Thr
4	Cys	<b>36</b>	<b>SerPro</b>		Gln
5, 29	Gly	<b>15</b>	<b>ValGlu</b>		Phe
111	Gln	11	ArgGly		
18	Val	13	GlyAsn		
24	His	24	GlnGly		
17, 7	Ala	25	GluAsn		
9, 26	GlyAsn	37	LysGln		
11	ValAsn	38	LeuPhe		
15	GlyAla	34, 5	Glu Val		
27	ArgGlu	47	Val Ser		
16	ThrGly	49	Ala Ser		
14	LysVal	8	Glu Pro		
22	TyrSer	<b>14</b>	<b>AlaGluAsn</b>		
23	ValSer	<b>18</b>	<b>ValGluGlu</b>		
13	Glu Gly	<b>7</b>	<b>ValLysGlu</b>		
12	Gly Gln	35	AlaGlyVal		
19	Val Gly				
30	Tyr Gly				
20	AlaTyrGlu				
25	AsnSerAla				

**Figure 5.** *rop* bend mutants. The Figure summarizes the amino acid sequence of the mutants described in this work. Commas separate independent isolate numbers that had the same amino acid sequence. On the left (A) are the mutants that were constructed in the expression vector pEX58 and whose folding and stability was studied by over-expression (see Figure 3). Mutants identified by a number preceded by Rop were obtained by independent site-directed mutagenesis experiments, whereas the others were obtained in a single experiment utilizing an oligo of degenerate sequence as described in Materials and Methods. B, Mutants obtained by mutagenesis of the dimerization probe plasmid PC132 and characterized by testing their immunity to  $\lambda$  infection. The wild-type sequence is drawn at the top in bold characters. Only the amino acids that were changed in the mutagenesis procedure are indicated. Non-boxed sequences represent a random sample from our collection of mutants while the boxed ones were selected among 380 mutants because of their acquired sensitivity to  $\lambda$  infection. Shaded boxes identify the mutants that are unable to dimerize as demonstrated by their inability to restore the function of the amino terminus of the  $\lambda$  repressor. Mutants identified by open boxes confer only partial activity on the corresponding cI-Rop fusion proteins. C, List of the amino acids that were found to be acceptable at each of the 3 bend positions. The wild-type residues are in bold characters, while the residues that can be encoded after single nucleotide substitution are underlined.

poured onto a Petri dish to form a bacterial lawn and spotted with suspensions of varying concentrations of  $\lambda$  only four of them (VED, VPD, DSD and YPD) were able to support the growth of single plaques of lambda. The remaining mutants were lysed only at high  $\lambda$  concentration, indicating that the hybrid cI-Rop molecule retains partial activity. To prove that the observed changes in  $\lambda$  immunity were a consequence of the substitutions in the Rop bend, and not in any other part of the plasmid, we

restored (by substituting the wild-type *Bam*HI-*Sal*I DNA fragment) the wild-type Rop sequence into the plasmids encoding a defective cI-Rop. This operation was sufficient to restore the wild-type phenotype in all but one (DSD) of the plasmids. The defective DSD mutant, which probably contains a second mutation in the cI domain was discarded. This is consistent with the observation that a second independent DSD mutant had a wild-type phenotype.

#### 4. Discussion

In order to shed light on the molecular mechanism underlying protein folding and stability in the small protein Rop we have undertaken two approaches. One consists in a detailed study of a selected number of Rop variants by structural (Banner *et al.*, 1987; Castagnoli *et al.*, 1989; Eberle *et al.*, 1990) and microcalorimetric techniques (Steif *et al.*, 1993) to attempt to reach a quantitative description of the kinetics and thermodynamics of Rop folding in a manner analogous to what has been done successfully in other systems (Dobson, 1993; Fersht & Serrano, 1993). The second approach, as described here, aims at analyzing, by simple qualitative tests, the largest possible collection of sequence variants of the native Rop protein to provide a list of active and inactive sequences. These lists can then be compared to extract information about the sequence features that are important for protein folding and stability. To this end we have constructed a gene fusion whose expression permits us to associate a selectable phenotype to the ability of a Rop variant to form, by dimerizing, a bundle of four  $\alpha$ -helices. The approach is based on the observation by Sauer and collaborators (Hu *et al.*, 1990), that the dimerization domain of the  $\lambda$  repressor can be functionally substituted by other proteins and protein domains that can dimerize. Our experiments, by utilizing proteins whose three-dimensional structure is known (this paper and other unpublished observations), suggest that the method is relatively insensitive to the distance and topology of the fusion site and that it is therefore likely to be generally applicable as a dimerization probe, as long as the peptide linker between the two domains is sufficiently long and flexible.

To add versatility to the screening system we devised a color test that allows us to recognize, from the colony color, bacteria that express a full length Rop domain and to filter out nonsense and frameshift mutants. This is based on the expression of a downstream  $\alpha$ -peptide domain. It is conceivable, however, that this screening could bias our results, since proteolysis-sensitive mutants could also lead to proteolysis of the entire fusion protein including the  $\alpha$ -peptide. In this case some interesting mutant could be discarded as they would show a white colony phenotype. Although we cannot formally exclude this possibility, we have accumulated evidence that indicates that this is not a common event. Three folding defective mutants, that had not been selected by the blue/white test (L48A, L41N and A31P), once transplanted into the PC132 background, form blue colonies. Furthermore from the characterization of a random collection of substitutions at position Ile15, we found that seven out of seven white colonies contain plasmids with a nonsense or a frameshift mutation in the *rop* gene. Twenty independent folding mutants, by contrast, form blue colonies.

In our experiments we found that a large fraction

(> 99%) of our collection of 380 *rop* mutants, containing single, double or triple amino acid substitutions at position 30, 31 and 32, do not prevent Rop folding. In Figure 5C we have tabulated the residues that were found to be acceptable in the two experiments whose results are reported in Figure 5A and B. These findings are in accord with the results of similar experiments on cytochrome *b*-562, recently reported by Brunet *et al.*, (1993). Although a detailed quantitative analysis is likely to reveal subtle differences in the folding process and in the  $\Delta G$  value of these mutant proteins, these results imply that the chemical properties of the wild-type side-chains in the turn do not play an important role in determining the formation of the four helix bundle.

Inspection of the amino acid sequence of the rare mutants that cannot confer immunity on the host point out the negative effect of Val when it replaces Asp at position 30 and Pro replacing Ala at position 31. Branched  $C^\beta$  residues prefer an  $\alpha$  and  $\beta$  conformation strongly over  $3_{10}$  and almost never have a positive  $\Phi$ . This can explain why Val is not liked at position 30, where the native residue Asp has an  $a_L$  conformation ( $\phi = 57, \psi = 38$ ). We would predict that Ile and Thr would also create problems in that position. These residues are not present at position 30 in our mutant collection, since it is not possible to obtain Ile or Thr codons by single nucleotide substitution. Our results are consistent with the observation that in the barnase protein Val is the worse residue at the carboxy-cap position (Serrano *et al.*, 1992). The rigid side-chain of Pro cannot replace Ala at position 31 without substantial adjustment of the  $\alpha$ -carbon backbone, since the  $C^\beta$  of Pro bumps against the backbone O of Leu26.

However, these rationalizations, based on the native backbone conformation, are not completely satisfactory for at least two reasons. The rule "no Val at position 30 or Pro at position 31" is very context-dependent, since mutant proteins are found which tolerate these substitutions (see for instance mutants 23, 19 and 47 as exceptions for Val30 and mutant 54 as an exception for Pro31). This implies that the bend region is relatively malleable and that, in most of our mutants, the  $\alpha$ -carbon backbone is rearranged without disturbing the stabilizing interactions in the hydrophobic core; otherwise we would not be able to explain how mutant 11, for instance, forms a dimer despite the impossibility of fitting the long charged side-chain of Arg in the hydrophobic cavity left by the methyl of Ala31. Furthermore, if we graft on the native backbone of the mutant side-chains, and we minimize the free energy by several cycles of molecular dynamics, we end up with conformations which are all apparently satisfactory and whose calculated free energies do not correlate with the results of the qualitative *in vivo* folding assay.

It is difficult to explain the phenotype of our mutants on the basis of a simple analysis of the stabilizing and destabilizing interactions of the bend side-chains on the conformation of the Rop

variants. An alternative kinetic interpretation, although probably correct, at least in the case of Pro31, is even more difficult to prove by our *in vivo* approach.

Whatever the detailed molecular explanation that is likely to come from quantitative kinetic and thermodynamic studies of some of these mutants, our results indicate that most combinations of amino acid sequences can replace the native residues in the Rop bend, thus underlining the low informational importance of this sequence. By contrast similar experiments in the core region have revealed the primary importance of hydrophobic interactions in determining Rop conformation (M. A. Abril & G. Cesareni, unpublished results). This conclusion is in contrast with the suggestion by Carlacci & Chou (1990) who, on the basis of computer model studies, have proposed that an overwhelming majority of energy favorable for forming a four helix bundle structure is from the interaction between  $\alpha$ -helices and loops.

We thank P. Battaglia who initially pointed out to us the potential of repressor fusion technology for studying dimeric proteins, and J.C. Hu for plasmid and bacterial strains. M. Helmer-Citterich has helped in the preparation of Figure 1. This work was supported by a grant from the EC BRIDGE program and by CNR target project Genetic Engineering and Biotechnology and Bioinstrumentation.

### References

- Banner, D. W., Kokkinidis, M. & Tsernoglou, D. (1987). Structure of the ColE1 Rop protein at 1.7 Å resolution. *J. Mol. Biol.* **196**, 657–675.
- Bashford, D., Chothia, C. & Lesk, A. M. (1987). Determinants of a protein fold. Unique features of the globin amino acid sequences. *J. Mol. Biol.* **196**, 199–216.
- Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990). Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science*, **247**, 1306–1310.
- Brunet, A. P., Huang, E. S., Huffine, M. E., Loeb, J. E., Weltman, R. J. & Hecht, M. H. (1993). The role of turns in the structure of an  $\alpha$ -helical protein. *Nature (London)*, **364**, 355–358.
- Carlacci, L. & Chou, K. C. (1990). Electrostatic interactions between loops and  $\alpha$ -helices in four helix bundle proteins. *Protein Eng.* **4**, 2225–2227.
- Castagnoli, L., Scarpa, M., Kokkinidis, M., Banner, D. W., Tsernoglou, D. & Cesareni, G. (1989). Genetic and structural analysis of the ColE1 Rop (Rom) protein. *EMBO J.* **8**, 621–629.
- Cesareni, G., Helmer-Citterich, M. & Castagnoli, L. (1991). Control of ColE1 plasmid replication by antisense RNA. *Trends Biochem. Sci.* **7**, 230–235.
- Dente, L., Cesareni, G. & Cortese, R. (1983). pEMBL: a new family of single stranded plasmids. *Nucl. Acids Res.* **11**, 1645–1655.
- Dobson, C. M. (1993). Folding and binding. *Curr. Opin. Struct. Biol.* **1993**, 57–59.
- Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. & Wright, P. E. (1988). Folding of immunogenic peptide fragments of protein in water solution. I. Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* **201**, 161–200.
- Eberle, W., Klaus, W., Cesareni, G., Sander, C. & Rösch, P. (1990). Proton nuclear magnetic resonance assignments and secondary structure determination of the ColE1 Rop (Rom) protein. *Biochemistry*, **29**, 7402–7407.
- Fersht, A. & Serrano, L. (1993). Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* **3**, 75–83.
- Fetrow, J. S., Cardillo, T. & Sherman, F. (1989). Deletion and replacement of omega loops in yeast iso-1-cytochrome c. *Proteins*, **6**, 372–381.
- Greer, J. (1981). Comparative model-building of the mammalian serine proteases. *J. Mol. Biol.* **153**, 1027–1042.
- Haase-Pettingel, C. A. & King, J. (1988). Formation of aggregates from a thermolabile *in vivo* folding intermediate in P22 tailspike maturation. A model for inclusion body formation. *J. Biol. Chem.* **263**, 4977–4983.
- Hampsey, M. D., Das, G. & Sherman, F. (1988). Yeast iso-1-cytochrome c: genetic analysis of structural requirements. *FEBS Letters*, **231**, 275–283.
- Hu, J. C., O'Shea, E. K., Kim, P. S. & Sauer, R. T. (1990). Sequence requirements for coiled-coils: analysis with  $\lambda$  repressor-GCN4 leucine zipper fusion. *Science*, **250**, 1400–1403.
- Jappelli, R., Luzzago, A., Tataseo, P., Pernice, I. & Cesareni, G. (1992). Loop mutations can cause a substantial conformational change in the carboxy terminus of the ferritin protein. *J. Mol. Biol.* **227**, 532–543.
- Lesk, A. M. & Chothia, C. (1980). How many different amino acid sequences determined similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **136**, 225–270.
- Leszczynski, J. F. & Rose, G. D. (1986). Loops in globular proteins: a novel category of secondary structure. *Science*, **234**, 849–855.
- Lewis, P. N., Momany, F. A. & Sheraga, H. A. (1971). Folding of polypeptide chains in proteins: a proposed mechanism for folding. *Proc. Nat. Acad. Sci., U.S.A.* **68**, 2293–2297.
- Lim, W. A. & Sauer, R. T. (1991). The role of internal packing interactions in determining the structure and stability of a protein. *J. Mol. Biol.* **219**, 359–376.
- Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M. & Fersht, A. R. (1990). Transient folding intermediates characterized by protein engineering. *Nature (London)*, **346**, 440–445.
- Messing, J., Gronenborn, B., Mueller-Hill, B. & Hofschneider, P. H. (1977). Filamentous coliphage M13 as a cloning vehicle: insertion of a HindIII fragment of the lac regulatory region in M13 replicative form *in vitro*. *Proc. Nat. Acad. Sci., U.S.A.* **74**, 3642–3646.
- Pearl, L. H. & Taylor, W. R. (1987). A structural model for the retroviral proteases. *Nature (London)*, **329**, 351–353.
- Remaut, E., Tsao, H. & Fiers, W. (1983). Improved plasmid vectors with a thermoinducible expression and temperature-regulated runaway replication. *Gene*, **22**, 103–113.
- Richardson, J. & Richardson, D. (1988). Amino acid preferences for specific locations at the end of  $\alpha$ -helices. *Science*, **240**, 1648–1652.
- Rose, G. D., Gierash, L. M. & Smith, J. A. (1985). Turns in peptide and proteins. *Advan. Protein Chem.* **37**, 1–109.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989).

- Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A. R. (1992).  $\alpha$ -Helix stability in proteins. Empirical correlations concerning substitutions of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces. *J. Mol. Biol.* **227**, 544-549.
- Starzyk, R. M., Webster, T. A. & Schimmel, T. (1987). Evidence for dispensable sequence inserted into a nucleotide fold. *Science*, **237**, 1614-1618.
- Steif, C., Weber, P., Hinz, H. J., Floßdorfer, J., Cesareni, G. & Kokkinidis, M. (1993). Subunit interactions provide a significant contribution to the stability of the dimeric four-helical-bundle protein Rop. *Biochemistry*, **32**, 3867-3876.
- Taylor, W. R. (1986). Identification of protein sequence homology by consensus template alignment. *J. Mol. Biol.* **188**, 233-258.
- Wang, X. & Pielak, G. J. (1991). Temperature-sensitive variants of *Saccharomyces cerevisiae* iso-1-cytochrome c produced by random mutagenesis of codon 43 to 54. *J. Mol. Biol.* **221**, 97-105.
- Weber, P. (1991). Ph.D. thesis, University of Regensburg.
- Zimmerman, S. S. & Sheraga, H. A. (1977). Local interactions in bends of proteins. *Proc. Nat. Acad. Sci., U.S.A.* **74**, 4126-4129.

*Edited by A. R. Fersht*

*(Received 3 August 1993; accepted 5 January 1994)*