Characterization of a Helical Protein Designed from First Principles

LYNNE REGAN AND WILLIAM F. DEGRADO

The question of how the primary amino acid sequence of a protein determines its three-dimensional structure is still unanswered. One approach to this problem involves the de novo design of model peptides and proteins that should adopt desired three-dimensional structures. A systematic approach was aimed at the design of a four-helix bundle protein. The gene encoding the designed protein was synthesized and the protein was expressed in *Escherichia coli* and purified to homogeneity. The protein was shown to be monomeric, highly helical, and very stable to denaturation by guanidine hydrochloride (GuHCl). Thus a globular protein has been designed that is capable of adopting a stable, folded structure in aqueous solution.

THE DE NOVO DESIGN OF PROTEINS with predetermined secondary and tertiary structures is an important endeavor that not only tests our understanding of protein structure and folding, but also lays the groundwork for the design of macromolecules with unprecedented structures and properties (*1*). We have recently established an iterative approach aimed at the design of a four-helix bundle protein (Fig. 1) (*2–4*). A helical protein was chosen for the first design attempt to take advantage of the available information concerning the factors stabilizing α-helix formation in aqueous solution (*5*). Also, the presence of α helices can be inferred from the circular dichroism (CD) spectrum of a protein with better certainty than any other secondary structure. Our model protein is an idealized version of the naturally occurring four-helix bundle motif found in proteins such as myohemerythrin and cytochrome c′ (*6*). In the first phase of this approach (Fig. 1A), peptides of 12 to 16 residues in length that cooperatively self-assembled into α-helical tetramers were designed and characterized (*2–4*). In the second phase, a single hairpin loop was inserted between two identical helices to give a peptide that dimerized in aqueous solution (Fig. 1B) (*3*). The free energy associated with dimerization (Fig. 1B) or tetramerization (Fig. 1A) of the designed peptides could be experimentally determined from the concentration dependence of the CD spectra for the peptides. At low concentrations the peptides were found to be monomeric and had low helical contents, whereas at high concentrations they self-associated, adopting helical conformations. The free energies associated with the self-assembly of the random-coil peptides into helical proteins provided an experimental measure of the stability of the folded structures, allowing alternate sequences for the helices and loops to be evaluated (*3*). In this report we describe the properties of the completed protein composed of four identical, designed helices connected by three identical, designed loops (Fig. 1C).

The rationale for the design of the helical and link sequences (Fig. 2A) has been described previously (*2–4*). Briefly, the helical sequence is capable of forming an amphiphilic α-helical structure that can form a tetrameric aggregate with the hydrophobic Leu side chains projecting toward the interior of the tetramer and the hydrophilic residues projecting toward the aqueous exterior. The arrangement of Leu residues is based on the helical packing observed in naturally occurring four-helix bundle proteins (*7*). Glutamic acid and lysine were chosen for the hydrophilic, helical residues, and these were arranged so that they could form favorable ion pairs along one face of the helix. In addition, the helix has a number of negatively charged residues at its NH_2-terminus and positively charged residues at its COOH-terminus to stabilize helix formation by favorably interacting with the helical dipole (*5, 8*). The sequence linking the helices contains a single Pro residue to aid in helix termination, as well as two Arg residues to promote a reversal in the overall peptide chain direction (*9*).

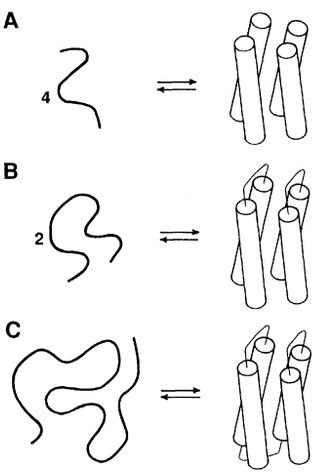

Fig. 1. A schematic illustration of the incremental approach to the design of a four-helix bundle protein.

The gene encoding the full-length protein (Fig. 2B) was assembled from DNA oligonucleotides and inserted behind the *tac* promoter of plasmid pTM201/NS3-3 (*4, 10*). The expression of protein α_4 in *Escherichia coli* could be readily detected by immunoblot (*11*) analysis of crude protein extracts with polyclonal antibodies that were raised against the peptide α_2 (Fig. 2A) (*4*). The protein was purified from an *E. coli* extract with a combination of ion-exchange and size-exclusion chromatography to give a homogeneous preparation (Fig. 3). The NH_2-terminal sequence analysis of the final preparation revealed a single sequence that corresponded to the desired structure over 25 cycles of Edman degradation, with the initiator Met residue retained. Amino acid analysis confirmed the expected composition.

A

Helix -Gly-Glu-Leu-Glu-Glu-Leu-Leu-Lys-Lys-Leu-Lys-Glu-Leu-Leu-Lys-Gly-

Loop -Pro-Arg-Arg-

α_1 Ac-Helix-$CONH_2$

α_2 Ac-Helix-Loop-Helix-$CONH_2$

α_4 Met-Helix-Loop-Helix-Loop-Helix-Loop-Helix-COOH

B

```
Met Gly Glu Leu Glu Glu Leu Leu Lys Lys Leu Lys Glu Leu Leu Lys Gly Pro Arg Arg
CC ATG GGT GAA CTG GAA GAG CTG CTT AAG AAA CTT AAA GAA CTT CTG AAG GGC CCG CGT CGT
   Nco I                    Afl II                          Apa I

    Gly Glu Leu Glu Glu Leu Leu Lys Lys Leu Lys Glu Leu Leu Lys Gly Pro Arg Arg
    GGC GAA CTC GAG GAA CTG CTG AAA AAA CTG AAG GAG CTG CTT AAA GGT CCT AGG CGT
        Xho I                                                   Avr II

    Gly Glu Leu Glu Glu Leu Leu Lys Lys Leu Lys Glu Leu Leu Lys Gly Pro Arg Arg
    GGT GAG CTG GAG GAG CTG CTG AAG AAG CTT AAA GAA TTA CTG AAA GGT CCG CGG CGT
                                  Hind III                      .Sac II

    Gly Glu Leu Glu Glu Leu Leu Lys Lys Leu Lys Glu Leu Leu Lys Gly STOP
    GGC GAG CTG GAA GAA CTG CTT AAG AAA TTG AAA GAA CTT TTG AAG GGC TAG
        Sac I
```

Fig. 2. (**A**) The amino acid sequences of the best helical sequence from stage 1 of the design, the loop used to connect the helices, and the peptides α_1, α_2, and α_4; Ac, acetyl; $CONH_2$, carboxamide. (**B**) The amino acid sequence of the entire four-helix bundle protein, with the gene sequence and the positions of unique restriction sites indicated below. The locations of restriction sites were chosen to facilitate future design manipulations.

E. I. du Pont de Nemours & Company, Central Research & Development Department, Experimental Station, Wilmington, DE 19898.

The results of size-exclusion chromatography and CD together indicate that α_4 adopts a monomeric, compact, globular, α-helical structure in aqueous solution. The apparent molecular mass of the protein, as determined by size-exclusion chromatography with a Sephadex G-50F column calibrated with small, globular, folded proteins (*3*), was in good agreement with the calculated monomeric molecular mass (8.5 versus 8.7 kD). Unfolded or nonglobular conformations, such as a two-stranded coiled-coil of α helices, would have given substantially greater apparent molecular masses (*12*). We observed no evidence for higher molecular mass aggregates in the experiments with protein α_4. The CD spectrum of α_4 (Fig. 4) is indicative of a predominantly α-helical conformation (*13*) and is similar to that for the self-associated forms of α_1 and α_2. The mean residue ellipticity was independent of protein concentration even at the midpoint of the GuHCl-induced denaturation curve, in which any aggregation-induced stabilization of secondary structure would be expected to be most apparent.

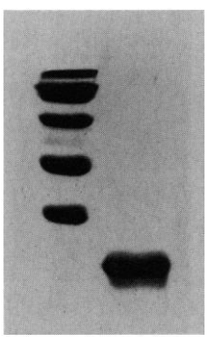

Fig. 3. SDS gel electrophoresis of protein α_4; an 18% SDS polyacrylamide gel stained with Coomassie brilliant blue is illustrated. In the left lane are molecular mass markers of 92.5, 66.2, 45, 31, 21.5, and 14.4 kD; the right lane was overloaded to demonstrate the purity of protein α_4 isolated from *E. coli*.

The protein was purified as follows: 15 liters of Zubay media with 100 μg/ml ampicillin were inoculated with a fresh overnight culture of GW5180 (F′ [*traD36 proAB*$^+$ *lacI*^{q} *Z M15*]/Δ*lac-pro thi strA endA sbcB15 supE recA::Cm*^{r}) containing the plasmid pα_4-1. At an absorbance at 600 nm equal to 0.5, isopropyl-β-D-thiogalactopyranoside was added to a final concentration of 5 m*M*. Ampicillin was added every 2 hours before induction and every 1 hour after induction. The cells were harvested by centrifugation 3 hours after induction. The cell pellets were stored frozen and then thawed on ice before use. The cells were lysed in 500 ml 10 m*M* sodium phosphate *p*H 7.5, 1 m*M* EDTA, 0.1% Triton X-100, and 0.1 m*M* phenylmethylsulfonyl fluoride by disruption in a Waring Blender followed by sonication. The cell debris was removed by centrifugation; the majority of the α_4 protein was present in the supernatant, as assessed by immunoblot analysis (*11*). Polymin P (25 ml of a 10% solution) was slowly added with continual stirring to the supernatant on ice. After stirring for an additional 30 min, the precipitate was removed by centrifugation. The volume of the supernatant was increased to 1 liter by addition of 10 m*M* sodium phosphate, *p*H 7.5, and applied to a column of TSK-Gel Toyopearl CM-650M (Supelco) (5 cm by 13 cm) at a flow rate of 1.5 ml/min. The column was washed with an additional 500 ml of 10 m*M* sodium phosphate, *p*H 7.5, and then eluted with a linear gradient of 0 to 500 m*M* NaCl in 10 m*M* sodium phosphate, *p*H 7.5, over 2 liters. The fractions were analyzed by SDS–polyacrylamide gel electrophoresis (PAGE), staining the gels for total protein and also performing immunoblot analysis, with anti-α_2 serum to identify protein α_4. The α_4 peak eluted at approximately 75 m*M* NaCl. The peak fractions containing α_4 were pooled, lyophilized, resuspended in a minimum volume of distilled water, and applied to a Sephadex G-50F column (1.5 cm by 110 cm) equilibrated in 500 m*M* ammonium acetate. Again, fractions were monitored by SDS-PAGE and immunoblot analysis. Protein α_4 eluted at a position consistent with its monomeric molecular weight. The peak fractions containing protein α_4 were concentrated with a Speed-Vac (Savant), then subjected to NH_2-terminal sequencing and amino acid composition analysis.

Table 1. A comparison of the stability of protein α_4 with that of small natural proteins (*18*). The data were analyzed by plotting ΔG^{obs}, the measured free energy of folding at a given concentration of denaturant, for each protein against the molar concentration of denaturant *C* according to the equation $\Delta G^{obs} = \Delta G^{H_2O} + mC$; ΔG^{H_2O} is the free energy of folding in the absence of denaturant and *m* is the slope. The values of ΔG^{H_2O}, *m*, and $C_{0.5}$ (the concentration of denaturant at the midpoint of the transition) are tabulated. The magnitude of *m* is an indication of the cooperativity of the folding transition. It is important to note that linear extrapolation was used for each protein and that the error associated with the intercept, ΔG^{H_2O}, in each case is small, approximately 20%. Other methods of extrapolation could give values that differ by as much as a factor of 2, but if uniformly applied to this set of proteins the same rank order of stability should be observed.

Protein	ΔG^{H_2O} (kcal mol^{-1})	*C* (*M*)	*m* (kcal mol^{-1} *M*$^{-1}$)
α_4	−22.5	6.3	3.57
Lysozyme	−8.9	4.2	1.88
Myoglobin	−7.6	1.8	4.22
Ribonuclease A	−7.5	3.0	2.5
α-Lactalbumin	−4.2	3.6	1.27

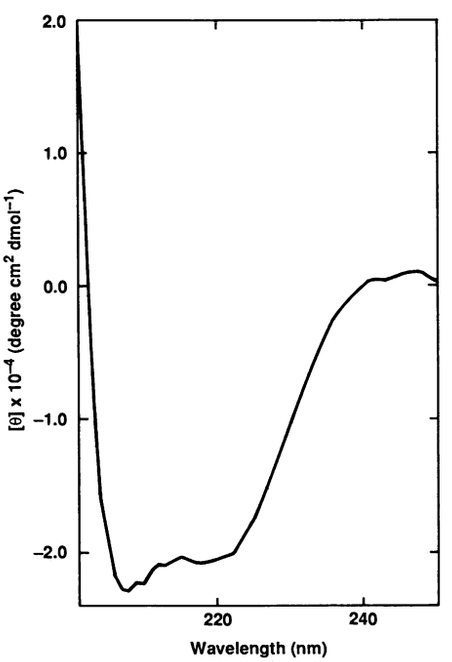

Fig. 4. CD spectrum of protein α_4 at a concentration of 88 μ*M* in 10 m*M* 3-[*N*-morpholino]propanesulfonic acid (MOPS), *p*H 7.0, and 150 m*M* NaCl. The spectrum was recorded in a 0.1-mm path-length cell on a Jovin Yvon Instruments S.A. spectropolarimeter interfaced with an Apple II GS computer.

In the series α_1, α_2, and α_4, the stability toward GuHCl denaturation increases concomitantly with the increase in covalent cross-links between helical monomers. Denaturation curves for α_1, α_2, and α_4 (Fig. 5) were determined by monitoring the ellipticity of the peptide at 222 nm (a local minimum in the CD spectrum of α helices) as a function of the GuHCl concentration. At equivalent peptide concentrations of approximately 5 μ*M*, the midpoints of the denaturation curves occurred at 0.55, 4.5, and 6.5*M* GuHCl for α_1, α_2, and α_4, respectively (*14*). Furthermore, as the number of covalent cross-links was increased, the curves became increasingly cooperative. Factors contributing to the observed increase in stability of α_4 as compared with α_1 and α_2

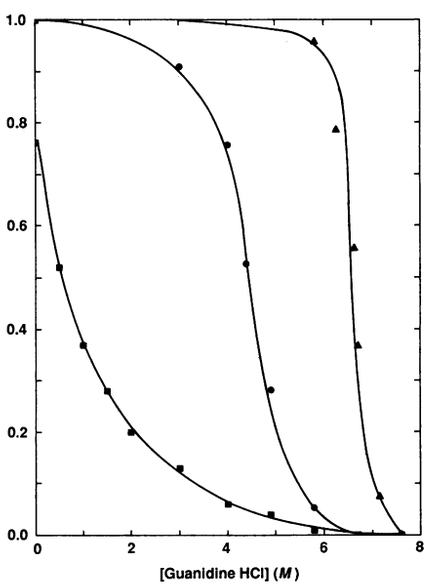

Fig. 5. GuHCl denaturation curves of peptides α_1 (■), α_2 (●), and α_4 (▲) at 6.5, 4.8, and 4.4 μ*M* peptide, respectively. The ellipticity at 222 nm was monitored as a function of GuHCl concentration. To compare the curves for different peptides, we plot *f* (the observed mean residue ellipticity at 222 nm at a given concentration of denaturant, normalized to the observed mean residue ellipticity in the absence of denaturant and at limiting high peptide concentration) versus the concentration of denaturant. In the case of α_1 at the peptide concentration used, the tetramer is partially dissociated and hence does not reach $f = 1.0$ at zero GuHCl concentration.

include both entropic effects and stabilizing interactions between atoms in the link sequence and the rest of the protein. As only some of these free energy components can be quantified at present, it is not yet possible to partition the stabilization energy into its component terms (*15, 16*).

The free energy associated with the transition from the unfolded, random coil state to the folded, helical state for α_4 in the absence of denaturants could be estimated by methods that involve extrapolation to zero GuHCl concentration (*17*); the free energy associated with the folding process was calculated to be -22 kcal mol^{-1} (*18*). Although this value is approximate because of the assumptions invoked in the development of linear free energy models (*17, 19*), it allows comparison of α_4 with other small, natural proteins that have been analyzed in the same way (Table 1) (*20*). α_4 possesses considerable stability when compared with other small proteins and thus provides a stable framework for the incorporation of future design modifications.

It is reasonable to adopt the working assumption that α_4 contains four closely packed α-helical segments. Additional chemical, crystallographic, and spectroscopic investigations are necessary to determine the arrangement of the helices and the positions of the loops in the structure.

REFERENCES AND NOTES

1. W. F. DeGrado, *Adv. Prot. Chem.*, in press; J. S. Richardson and D. C. Richardson, in *Protein Engineering*, D. L. Oxender and C. F. Fox, Eds. (Liss, New York, 1987), pp. 149–163; E. T. Kaiser and F. J. Kézdy, *Science* **223**, 249 (1984); M. Mutter, *Angew. Chem. Int. Ed. Engl.* **24**, 639 (1985); R. Moser *et al.*, *ibid.*, p. 719 (1985); C. O. Pabo, *Nature* **301**, 200 (1983); M. Mutter, K.-H. Altman, T. Z. Vorherr, *Naturforsch* **41b**, 1315 (1986).
2. D. Eisenberg *et al.*, *Proteins Struct. Funct. Genet.* **1**, 16 (1986).
3. S. P. Ho and W. F. DeGrado, *J. Am. Chem. Soc.* **109**, 6751 (1987).
4. W. F. DeGrado, L. J. Regan, S. P. Ho, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
5. K. Shoemaker *et al.*, *Nature* **326**, 563 (1987).
6. P. C. Weber and F. R. Salemme, *ibid.* **287**, 82 (1980); P. Argos, M. G. Rossman, J. E. Johnson, *Biochem. Biophys. Res. Commun.* **75**, 83 (1977); S. Sheriff, W. A. Hendrickson, J. L. Smith, *J. Mol. Biol.* **197**, 273 (1987); B. C. Finzel, P. C. Weber, K. D. Hardman, F. R. Salemme, *ibid.* **186**, 627 (1985).
7. C. Chothia, M. Levitt, D. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4130 (1977); C. Chothia, *Annu. Rev. Biochem.* **53**, 537 (1984).
8. D. E. Blagdon and M. Goodman, *Biopolymers* **14**, 241 (1975).
9. G. D. Rose *et al.*, *Adv. Prot. Chem.* **37**, 1 (1985).
10. K. Knight and R. T. Sauer, unpublished results.
11. W. W. Burnette, *Anal. Biochem.* **112**, 195 (1981).
12. S. Y. M. Lau, A. K. Taneja, R. S. Hodges, *J. Biol. Chem.* **259**, 13253 (1984).
13. N. Greenfield and G. D. Fasman, *Biochemistry* **8**, 4108 (1969).
14. When GuHCl denaturation curves are measured at different concentrations of peptide, the same rank order of stability as shown in Fig. 5 is observed at all physically reasonable peptide concentrations. The precise midpoint of the denaturation curve is, of course, dependent on the concentrations of α_1 and α_2 (*3*).
15. M. I. Page and W. P. Jencks, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1678 (1971).
16. C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry* (Freeman, San Francisco, 1980), vol. 1, pp. 283–285.
17. J. A. Schellman, *Biopolymers* **17**, 1305 (1978).
18. To estimate the stability of protein α_4 in the absence of denaturant, measurements were made in the region near the midpoint of the transition of the GuHCl denaturation curve. From these data, either by linear extrapolation to zero GuHCl or by multiplying the slope by the midpoint (*16*), values of -22.4 kcal mol^{-1} or -22.5 kcal mol^{-1}, respectively, were calculated for the free energy of folding in the absence of denaturant.
19. C. N. Pace, *Biochemistry* **28**, 288 (1979); F. Ahmad and C. C. Bigelow, *Biopolymers* **25**, 1623 (1986).
20. F. Ahmad and C. C. Bigelow, *J. Biol. Chem.* **257**, 12935 (1982); C. N. Pace, *CRC Crit. Rev. Biochem.* 1 (1975).

21. We thank K. Knight and R. T. Sauer for their generous gift of plasmid pTM201/NS3-3; L. Janvier for conducting amino acid analysis; B. Krieger for Edman sequencing and S. Ho for the chemical synthesis of α_4; her preliminary spectroscopic and thermodynamic experiments on this material encouraged us in our work. Also we thank D. Eisenberg, S. Ho, D. Oprian, F. R. Salemme, P. Weber, and Z. Wasserman for helpful advice on various aspects of this work. We dedicate this report to the memory of Professor E. T. Kaiser of Rockefeller University.

29 April 1988; accepted 16 June 1988

Coordinate Hormonal and Synaptic Regulation of Vasopressin Messenger RNA

FRANK BALDINO, JR.,* TERESA M. O'KANE, SANDRA FITZPATRICK-MCELLIGOTT, BETTY WOLFSON

Previous studies have shown that adrenalectomy augments arginine vasopressin (AVP) messenger RNA levels in the adult paraventricular nucleus. It is now demonstrated that unilateral lesions in the lateral septal nucleus enhance the adrenalectomy-induced expression of AVP mRNA. This effect was entirely ipsilateral to the lesion and most prominent in the rostral paraventricular nucleus and related nuclei. Moreover, AVP and AVP mRNA were found to be colocalized with oxytocin in a few neurons. These results indicate that mRNA expression is modulated by synaptic influences and raise the possibility that synaptically mediated selection of neuronal phenotypes is a dynamic feature of the mature central nervous system.

THE EXPRESSION OF AVP MRNA within several subdivisions of the paraventricular nucleus (PVN) is regulated by glucocorticoids (*1–6*). Adrenalectomy induces the expression of AVP mRNA in corticotropin-releasing factor (CRF)–immunoreactive neurons within the PVN (*1, 2*), illustrating the potential of the adult genome to be differentially regulated in response to hormonal challenge. Other physiological influences, however, have also been reported to alter gene expression. The transmitter phenotype produced in developing peripheral ganglionic neurons (*7*), as well as the synthesis of specific muscle proteins (*8*), has been shown to depend, in part, on the type of presynaptic innervation. These findings suggest that synaptic input influences phenotypic expression in target tissues. The purpose of this study was to determine whether specific afferents contribute to the plasticity in AVP expression induced by adrenalectomy. We now report that the adrenalectomy-induced expression of AVP mRNA in the PVN can be enhanced by reducing specific afferent synaptic input.

Adrenalectomized (14-day-old; $n = 25$) and sham-operated ($n = 10$) male Sprague-Dawley rats were subjected to a unilateral electrolytic lesion in one of a number of regions shown by anatomical and physiological studies to project to the PVN (*9–15*). These regions included the bed nucleus of the stria terminalis, several subdivisions of the lateral septal nucleus, and the hippocampus (including the subiculum). The ipsilateral nature of these projections permitted the side contralateral to the lesion to serve as a control in each section. The anatomical locations of all lesions were verified histologically. In some experiments ($n = 6$), dexamethasone pellets (120 μg/day for 7 days) were administered subcutaneously 7 days after adrenalectomy.

In animals with lateral septal lesions, the number of neurons expressing AVP mRNA was increased in the rostral half of the PVN on the side ipsilateral to the lesion (Fig. 1). Most notably, AVP mRNA was detected in a population of neurons that normally do not express AVP or AVP mRNA. These

Medical Products Department, Experimental Station, E. I. du Pont de Nemours & Company, Wilmington, DE 19898.

*To whom correspondence should be sent at Cephalon, Inc., 145 Brandywine Parkway, West Chester, PA 19380.