Thermal Unfolding Studies of a Leucine Zipper Domain and Its Specific DNA Complex: Implications for Scissor's Grip Recognition†

Michael A. Weiss*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, and Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114

Received June 4, 1990; Revised Manuscript Received July 11, 1990

ABSTRACT: A newly recognized class of eukaryotic transcription factors is characterized by a bipartite sequence motif, consisting of a C-terminal dimerization region (the leucine zipper) and an N-terminal basic region (which mediates DNA binding). In studies of isolated leucine zipper peptides, the dimerization region has been characterized as a coiled coil of parallel α-helices. To extend these studies to a functional DNA-binding domain, we describe CD studies of the thermal unfolding and refolding of a 58-residue fragment of GCN4, the yeast homologue of the c-Jun protooncoprotein. This fragment, which contains the complete leucine zipper and basic region, retains the DNA-binding properties of the intact protein. The GCN4 DNA-binding domain exhibits two independent helix–coil unfolding transitions. The major transition (midpoint 65 °C) is due to dissociation of the dimer in accord with previous studies of an isolated leucine zipper. A novel pretransition in the temperature range 0–40 °C is also observed, which reflects partial stabilization of the nascent helix in the basic region. Remarkably, complete folding of the basic region as an α-helix requires specific DNA binding, and the protein–DNA complex exhibits a single cooperative unfolding transition. These results support a major feature of the recently proposed "scissor's grip" model of DNA recognition, in which the basic regions extend from the leucine zipper as bifurcating α-helical arms.

Protein–DNA recognition is mediated by classes of related structural motifs, such as the helix–turn–helix (Pabo & Sauer, 1984) and Zn finger (Klug & Rhodes, 1987). A novel motif has recently been described in a highly conserved class of eukaryotic transcription factors, consisting of a C-terminal dimerization element, the "leucine zipper", and an N-terminal basic region, which mediates DNA binding (Landschultz et al., 1988). Originally described in an avian oncoprotein (v-Jun; Maki et al., 1987) and its yeast homologue GCN4 (Jones & Fink, 1982; Hinnebusch & Fink, 1983; Struhl, 1987a,b, 1988), this bipartite sequence motif (designated bZIP) defines an extended class of mammalian regulatory proteins, including the Jun, Fos, and CREB-related families (Landschulz et al., 1988, 1989). These families coordinate changes in gene regulation in response to extracellular signals.

Physicochemical studies of an isolated leucine zipper peptide indicate that this moiety adopts a coiled-coil structure of parallel α-helices (O'Shea et al., 1989a,b; Oas et al., 1990). Although the structure of an intact leucine zipper protein has not been determined, a detailed model of the DNA-binding domain and its interaction with DNA has recently been proposed (Vinson et al., 1989). In this model (designated the "scissor's grip") the basic regions extend from the leucine zipper (coiled coil) as bifurcating α-helical arms. Support for this model is provided by "zipper swap" experiments (Kouzarides & Ziff, 1989; Sellars & Struhl, 1989) and by a recent study of sequence-specific DNA cleavage by GCN4–Fe·EDTA adduct (Oakley & Dervan, 1990). We focus here on the implications of this model for protein folding: the leucine zipper and basic "arms" are predicted to exhibit independent folding transitions, and the bifurcating arms—lacking higher order interactions—would not be expected to be fully folded in the absence of specific DNA binding.

To test these predictions, we describe CD studies of the thermal unfolding of the DNA-binding domain of GCN4 (Hope & Struhl, 1986, 1987). Remarkably, two distinct types of transitions are observed: a major helix–coil transition (midpoint 65 °C) due to dissociation of the leucine zipper and a novel pretransition at lower temperatures reflecting the partial folding of the basic region as an α-helix. The pretransition is broad and incomplete, as expected for bifurcating arms. In the specific protein–DNA complex the basic region is stabilized as an α-helix, and only a single highly cooperative unfolding transition is observed.

Materials and Methods

Expression and Purification of GCN4 DNA-Binding Domain. The DNA-binding domain of GCN4 (GCN4-p; residues 225–281) was overexpressed in *Escherichia coli* by using the T7 system (Studier & Moffat, 1986; Rosenberg et al., 1987) and was generously provided for these studies by T. E. Ellenberger. The purified protein was >98% pure by SDS–PAGE. The N-terminal sequence contained nonnative residues Met_1-Lys_2. The predicted amino acid composition was verified by composition analysis and N-terminal sequencing. Specific DNA-binding activity was verified by a gel-retardation assay as described in detail elsewhere (Weiss et al., 1990).

Peptide Synthesis and Characterization. The basic-region peptide [GCN4-br, residues 224–249, (N) SSDPAAL-KRARNTEAARRSRARKLQR-$CONH_2$] was synthesized by the solid-phase procedure (Barany & Merrifield, 1979; Stewart & Young, 1984) as a C-terminal amide (benzhydrylamine hydrochloride resin) and purified by reversed-phase HPLC on a C-18 column. Purity was evaluated by analytical HPLC, and composition and sequence analysis of preview (Tregear et al., 1977).

† This work was supported in part by the Markey Charitable Trust and by grants to M.A.W. from the National Institutes of Health, the Ruth Estrin Goldberg Memorial Fund, the Elsa U. Pardee Foundation, and the Massachusetts Chapter of the American Cancer Society. M.A.W. is also supported by the Pfizer Scholars Program for New Faculty and an American Cancer Society Junior Faculty Research Award.

* Address correspondence to this author at Harvard Medical School.

0006-2960/90/0429-8020$02.50/0

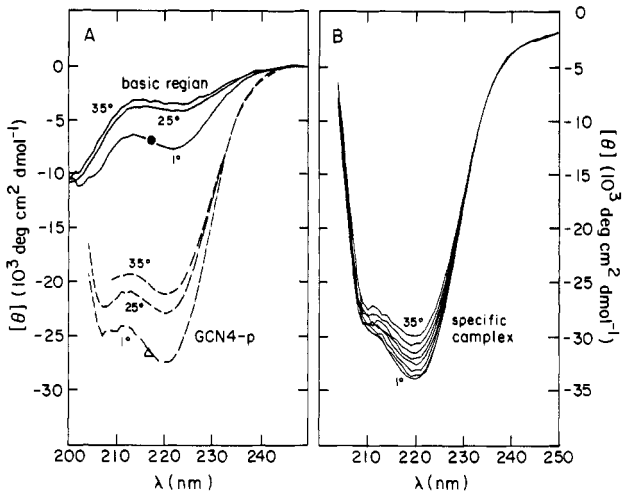

FIGURE 1: (A) Representative far-UV CD spectra of the intact DNA-binding domain (GCN4-p; dashed line and △) and an isolated basic-region peptide (GCN4-br; solid line and ●) at 1, 25, and 35 °C. (B) Representative CD spectra of the specific protein–DNA complex at 0, 5, 10, 15, 20, 25, 30, and 35 °C. Although these spectra contain contributions from both the protein and DNA, the latter contribution (representing an equimolar 16-base-pair duplex) is less than 5% of the former between 210 and 230 nm. The negative initial ellipticity of the complex between 245 and 250 nm reflects the negative contribution of DNA in this region. The concentration of GCN4-p was 145 μM in 50 mM potassium phosphate (pH 7) and 200 mM KCl; the concentration of GCN4-br was 154 μM in the same buffer. The temperature dependence of $[\theta]_{222}$ for each of these species under these conditions is shown in Figure 3.

Aggregation State. This was determined by gel filtration chromatography (Sephadex G-50 fine). At the concentration and conditions of study GCN4-p (58 residues; monomer molecular mass 6.9 kDa) is dimeric; the elution position expected for a dimer was calibrated in reference to the α-subunit of human chorionic gonadotropin (92 residues). GCN4-br is monomeric; the elution position expected for a monomeric basic-region peptide (26 residues; molecular mass 3 kDa) was calibrated in reference to a fragment of parathyroid hormone (residues 1–34).

Oligonucleotide Synthesis. A synthetic GCN4 duplex binding site (5′-GAGATGACGTCATCTC; Sellers et al., 1990) was synthesized by automated phosphoramidite chemistry and provided by PL-Pharmacia, Inc. (Milwaukee, WI). This self-complementary duplex also resembles sites bound by the mammalian ATF/CREB family of proteins (Hai et al., 1988). The strands were dissolved in CD buffer (below) and annealed by slow cooling from 80 °C. Purity (>98%) was evaluated by autoradiography following polyacrylamide gel electrophoresis of an ^{32}P-labeled aliquot.

Circular Dichroism. CD spectra were obtained with an Aviv spectropolarimeter equipped with automated temperature-scanning unit. Samples were dissolved in 200 μM buffer (see figure captions) and placed in a 1-mm path-length cuvette. Protein concentrations in CD samples were calculated following quantitative amino acid analysis. The corresponding UV absorbance spectrum indicates that a 1 mg/mL solution has an absorbance of 0.158/cm at 280 nm (T. E. Ellenberger, personal communication).

Results

The baseline CD spectrum of the dimeric GCN4 DNA-binding domain (GCN4-p) is shown in panel A of Figure 1 and is discussed in detail elsewhere (Weiss et al., 1990). This spectrum exhibits inflections at 208 and 222 nm characteristic of an α-helix. At 25 °C the helix content is approximately 70%, as estimated from the mean residue ellipticity of −22 700

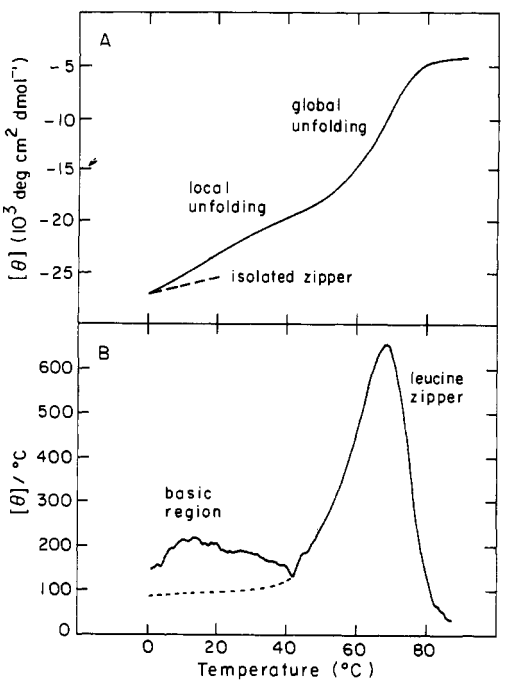

FIGURE 2: (A) Temperature dependence of mean residue ellipticity at 222 nm ($[\theta]_{222}$) of GCN4-p (solid line) and of an isolated GCN4 leucine zipper (dashed line). The latter data are replotted from O'Shea et al. (1989a) and normalized to an initial $[\theta]_{222}$ of 27 000 deg cm² dmol⁻¹. Data points were obtained at intervals of 0.2 °C. (B) First derivative ($d[\theta]/dT$) of the data shown in panel A. A broad pretransition is resolved in the range 0–40 °C, assigned to the basic region, and a cooperative unfolding transition is observed with midpoint 65 °C due to dissociation of the leucine zipper. The midpoint of the global unfolding transition is observed to vary with pH and protein concentration. The derivative was smoothed by averaging the adjacent five points. The protein concentration was 220 μM in 50 mM potassium phosphate (pH 9.5) and 200 mM KCl. The unfolding curve is reversible at protein concentrations less than 50 μM; at higher concentrations the kinetics of refolding are altered by aggregation of the denatured protein (see caption to Figure 4).

at 222 nm ($[\theta]_{222}$), assuming that for a 100% helical peptide $[\theta]_{222}$ is −33 000 at 0 °C and 30 500 at 25 °C (Chen et al., 1974; Woody, 1985). This corresponds to 40–43 residues, which presumably includes the 32–36-residue leucine zipper (Landschulz et al., 1988; O'Shea et al., 1989a). The CD spectrum of the specific protein–DNA complex is shown in panel B. Although this spectrum contains contributions from both the protein and DNA, the DNA contribution (an equimolar 16-base-pair site) is <5% of the protein from 210 to 230 nm. In fact, the deepening of the α-helix-associated minima near 208 and 222 nm is opposite to that expected in the presence of DNA, whose ellipticity is positive at these wavelengths. The helix content of the complex is at least 95%, indicating that the basic region undergoes a coil-to-helix transition upon specific DNA binding (O'Neil et al., 1990; Weiss et al., 1990). These results are in accord with the scissor's grip model proposed by McKnight and colleagues (Vinson et al., 1989).

Thermal Unfolding of the DNA-Binding Domain. The mean residue ellipticity $[\theta]_{222}$ of GCN4-p is shown as a function of temperature in panel A of Figure 2 (solid line). A major unfolding transition is observed with the midpoint near 65 °C, which presumably arises from dissociation of the dimer into unstructured monomers. A similar unfolding transition has previously been observed in a synthetic GCN4 leucine zipper (residues 249–281) and ascribed to cooperative unfolding of a coiled coil (O'Shea et al., 1989a). However, the initial slope of the GCN4-p temperature dependence (0–10

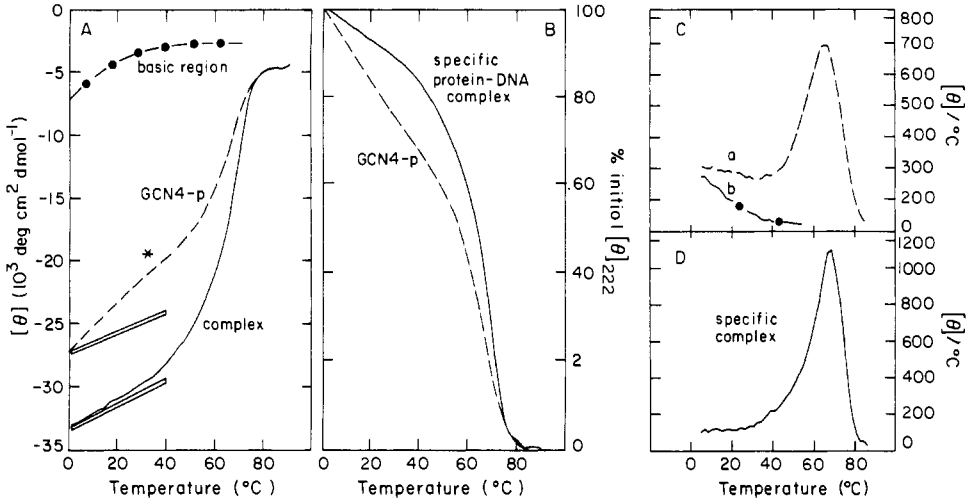

FIGURE 3: (A) Mean residue ellipticity $[\theta]_{222}$ as a function of temperature for the isolated basic region (GCN4-br; -●-), the DNA-binding domain (GCN4-p; dashed line), and the specific protein–DNA complex (solid line). Initial slopes expected on the basis of previous studies of the unfolding of coiled coils (0.3–0.4%/°C) are shown as open bars (O'Shea et al., 1989a; Lehrer et al., 1989). Data points were obtained at intervals of 0.5 °C. (B) Percent change in initial $[\theta]_{222}$ for GCN4-p and the specific complex. (C) First derivative ($d[\theta]_{222}/dT$) of the unfolding transition of GCN4-p (curve a; dashed line) and the isolated basic region (curve b; -●-). Note change in pH from conditions in Figure 1. (D) First derivative of the unfolding of the specific protein–DNA complex. Derivative curves were smoothed as described in the caption to Figure 1.

°C) is significantly greater than that of the isolated GCN4 leucine zipper [dashed line; replotted from data presented by O'Shea et al. (1989a)] and other coiled coils (Lehrer et al., 1989). The anomalous initial slope reflects a pretransition in the intact DNA-binding domain, which may be resolved by taking a first derivative (panel B). Experiments assigning the pretransition to the basic region and the major unfolding transition to the leucine zipper moiety are presented in turn.

A Peptide Model of the Basic Region. A structural mechanism for the GCN4-p pretransition is suggested by inspection of representative CD spectra at 1, 25, and 35 °C (Figure 1A). These spectra demonstrate a gradual increase in helix content to approximately 80% at 1 °C. Because the helix content of the leucine zipper moiety is constant in this temperature range (O'Shea et al., 1989a), these results suggest partial α-helical folding of the basic region at low temperatures.

To test further the hypothesis that the pretransition arises from partial folding of the basic region, we prepared a synthetic basic-region peptide (GCN4-br, residues 224–249–NH_2). The temperature dependence of the mean residue ellipticity $[\theta]_{222}$ is shown in panel A of Figure 3 (dash–dot line); representative CD spectra of GCN4-br are shown at 1, 25, and 35 °C in Figure 1A. Remarkably, this peptide exhibits a partial helix–coil transition at low temperatures that is similar to the pretransition of GCN4-p (dashed line in Figure 3A). At 1 °C the helix content of the basic-region peptide is about 20%. The loss of ellipticity at 222 nm between 1 and 30 °C is nearly identical in the two cases, accounting for the anomalous slope noted in Figure 1. The first derivatives of the GCN4-p and GCN4-br unfolding transitions at pH 7 are shown in Figure 3C (curves a and b, respectively).

^1H NMR Assignment of the Major Transition. The assignment of the major unfolding transition to the leucine zipper moiety of the intact domain is verified by ^1H NMR. GCN4-p contains two aromatic groups, Y265 and H266, which lie in the leucine zipper. Their chemical shifts are shown as a function of temperature in panel A of Figure 4. In the folded state at pH 7 the H266 resonances exhibit a linear change in chemical shift ($\Delta\delta/\Delta T$) with increasing temperature (dashed lines in panels B and C). This is due to a corresponding temperature dependence of the pK of the imidazole ring (and

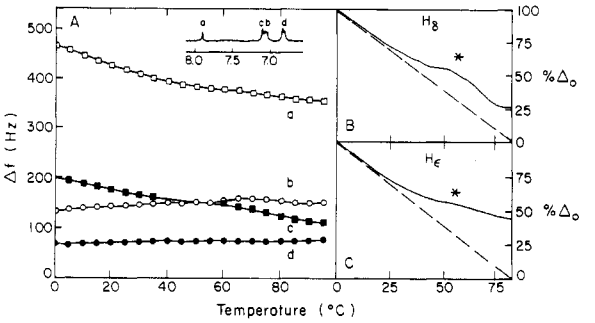

FIGURE 4: (A) ^1H NMR chemical shifts of the aromatic resonances (H266 and Y265) of GCN4-p as a function of temperature and a representative spectrum at 40 °C (inset). The Hϵ and Hδ resonances of H265 are labeled a and b, respectively; the ortho and meta resonances of Y266 are labeled c and d. (B) Percent change in chemical shift in the Hδ resonance of H265 (solid line) normalized to the extrapolated initial slope $\Delta\delta/\Delta T$ (dashed line). An asterisk indicates the unfolding transition. (C) Corresponding normalized data for the Hϵ resonance of H265. The tyrosine resonances do not shift significantly with temperature. Line broadening is observed in the temperature range 60–90 °C, indicating aggregation of the denatured protein under these conditions. Spectra were obtained at 300 MHz. The protein concentration was 1 mM in 50 mM potassium phosphate (pD 7, direct meter reading) and 500 mM KCl in D_2O.

buffer), as previously observed in NMR studies of the *lac* repressor headpiece (Wemmer et al., 1981). Unfolding of the leucine zipper may be monitored as a departure from this linear relationship, as shown in panels B and C. The tyrosine chemical shifts are not significantly affected by unfolding; presumably, their local magnetic environments are similar on the surface of the coiled coil (site 1b) and in the denatured state. Neither aromatic residue exhibits a pretransition in the temperature range 0–40 °C.

Unfolding of the Specific Protein–DNA Complex. Binding of GCN4-p to a synthetic GCN4 DNA-binding site induces helical folding of the basic region (O'Neil et al., 1990; Weiss et al., 1990). The temperature dependence of the mean residue ellipticity $[\theta]_{222}$ of the specific complex is shown in panel A of Figure 3 and in relation to initial value of $[\theta]_{222}$ at 0 °C in panel B (solid lines). Strikingly, the initial slope (0.3%/°C; open bar) is in accord with previous studies of α-helical proteins (Lehrer et al., 1989) and is similar to that of the isolated

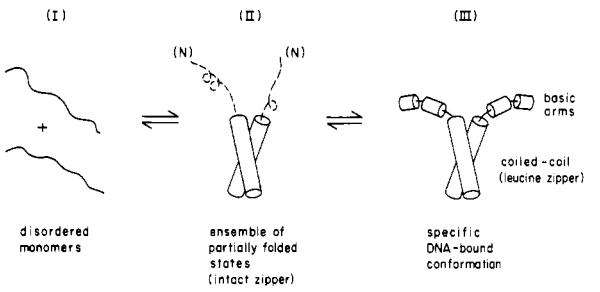

FIGURE 5: Schematic representation of the proposed equilibrium between unfolded monomers (state I), an ensemble of dimers containing an intact zipper but incompletely folded basic arms (state II), and the DNA-bound α-helical conformation that is stably folded (state III). The representation of state III is schematic only and is not meant to suggest details of the DNA-bound state, such as the position of helix breakpoints.

GCN4 leucine zipper (O'Shea et al., 1989a). The first derivative (panel D) exhibits a single unfolding transition (T_m 69 °C); this presumably arises from the cooperative dissociation of the protein–DNA complex into unstructured peptide monomers and single strands of DNA. The absence of a pretransition in the complex implies that the conformation of the basic region is stably folded when bound to a specific DNA site.

Discussion

The leucine zipper class of eukaryotic transcription factors contains a novel bipartite DNA-binding domain (Landschultz et al., 1988; 1989): the C-terminal subdomain (leucine zipper) is proposed to form a dimerization element, and the N-terminal basic region is proposed to contact DNA. The structure and specificity of the dimerization element have been investigated by zipper swap experiments (Kouzarides & Ziff, 1989; Sellars & Struhl, 1989; Agri et al., 1989) and physicochemical studies of synthetic peptides (O'Shea et al., 1989a,b; Oas et al., 1990). The direct involvement of the N-terminal region in DNA recognition is supported by studies of genetically altered leucine zipper proteins (Agri et al., 1989). GCN4, the yeast homologue of the AP-1-related family of mammalian protooncoproteins (Vogt et al., 1987; Struhl, 1988), provides a model system in which the folding properties of this DNA-binding motif may be investigated.

The leucine zipper moiety of GCN4 (residues 249–281) has previously been studied and appears to form a coiled coil of parallel α-helices (O'Shea et al., 1989a; Oas et al., 1990). To extend these studies to a functional DNA-binding domain (Hope & Struhl, 1986), we have studied the thermal unfolding of a fragment of GCN4 (residues 226–281). Remarkably, the GCN4 DNA-binding domain exhibits two independent unfolding transitions; a possible structural mechanism underlying these transitions is illustrated in schematic form in Figure 5. A global unfolding transition is observed (midpoint 65 °C) which is due to dissociation of the dimerization element into unstructured monomers (states I and II in Figure 5); a similar transition is observed in an isolated leucine zipper (O'Shea et al., 1989a).

The basic region exhibits a broad and incomplete pretransition at lower temperatures. This region appears to exist as an equilibrium ensemble of conformers (state II in Figure 5); an α-helical conformation is significantly populated only at temperatures below 25 °C. Comparative studies of a synthetic basic-region peptide (GCN4-br, residues 223–248) indicate that nucleation of helical structure is an intrinsic property of the local sequence and does not require tethering to the leucine zipper. Strikingly, stabilization of the basic region as an α-helix occurs in the specific protein–DNA complex (state III in Figure 5), and only a single highly cooperative unfolding transition is observed (midpoint 69 °C). The similar DNA-binding properties of the isolated DNA-binding domain and native GCN4 (Hope & Struhl, 1986; Hope & Struhl, 1987; Weiss et al., 1990) suggest that these observations reflect intrinsic features of the intact protein.

Although the structure of a leucine zipper protein has not yet been determined, a detailed model has recently been proposed (Vinson et al., 1989). The DNA-binding domain is envisaged as a "Y"-shaped protein, in which the two basic regions form bifurcating α-helical arms extending from the leucine zipper. These arms are proposed to follow the course of the major groove of DNA as a scissor's grip; this aspect of the model has been demonstrated by using a GCN4 adduct as a DNA-cleaving reagent (Oakley & Dervan, 1990). The present results also support certain features of this model. The coiled-coil dimerization element (leucine zipper) is retained in the intact DNA-binding domain and exhibits a global unfolding transition similar to that of an isolated leucine zipper (O'Shea et al., 1989a). The adjoining basic region constitutes an independent folding unit and binds DNA as an α-helix (O'Neil et al., 1990; Weiss et al., 1990). Its incomplete folding in the absence of specific DNA is as expected of bifurcating arms uninvolved in higher order structure. Stabilization of these arms in the protein–DNA complex provides a model for DNA acting as a template for protein folding (Warrant & Kim, 1978; Walters & Kaiser, 1985).

Acknowledgments

We thank S. Leher for use of his CD spectropolarimeter and helpful discussion; T. E. Ellenberger and B. Hahn for GCN4-p; T. E. Ellenberger, Prof. K. Struhl, S. C. Harrison, and J. Habener for helpful discussion and communication of results prior to publication; R. Wobbe for evaluation of the purity of synthetic GCN4 target DNA sites; H. T. Keutmann, S. G. Magil, and K. A. Mason for peptide synthesis and characterization; W. F. DeGrado for communication of results prior to publication; and J. P. Lee for assistance with NMR measurements. The Harvard Medical School NMR Facility was funded by a National Institutes of Health shared instrumentation grant (1 S10 RR04862-01) and an award from the National Health Resources Foundation.

References

Agri, P., Johnson, P. F., & McKnight, S. L. (1989) *Science 246*, 922.

Barany, G., & Merrifield, R. B. (1979) in *The Peptides* (Gross, E., & Meienhofer, J., Eds.) Vol. 2, p 10284, Academic Press, New York, NY.

Hai, T., Liu, F., Allegretto, E. A., Karin, M., & Green, M. R. (1988) *Genes Dev. 2*, 1216–1226.

Hinnebusch, A. G., & Fink, G. R. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5374–5378.

Hope, I. A., & Struhl, K. (1986) *Cell, 46*, 885–894.

Hope, I. A., & Struhl, K. (1987) *EMBO J. 6*, 2781–2784.

Klug, A., & Rhodes, D. (1987) *Trends Biochem. Sci. 12*, 464–467.

Kouzarides, T., & Ziff, E. (1989) *Nature (London) 240*, 568–571.

Landshulz, W. H., Johnson, P. F., & McKnight, S. L. (1988) *Science 240*, 1759–1764.

Landshulz, W. H., Johnson, P. F., & McKnight, S. L. (1989) *Science 243*, 1681–1688.

Maki, Y., Bos, T. J., Davis, C., Starbuck, M., & Vogt, P. K. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2848–2852.

Oakley, M. G., & Dervan, P. B. (1990) *Science 248*, 847-850.
Oas, T. G., McIntosh, L. P., O'Shea, E. K., Dahlquist, F. W., & Kim, P. S. (1990) *Biochemistry 29*, 2891-2894.
O'Neil, K. T., Hoess, R. H., & DeGrado, W. F. (1990) *Science* (in press).
O'Shea, E. K., Rutkowski, R., & Kim, P. S. (1989a) *Science 243*, 538-543.
O'Shea, E. K. Rutkowski, R., Stafford, W. F., III, & Kim, P. S. (1989b) *Science 243*, 1689-1694.
Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem. 53*, 293-321.
Rosenberg, A. H., Lade, B. N., Chui, D.-S., Lin, S.-W., Dunn, J. J., & Studier, F. W. (1987) *Gene 56*, 125-135.
Sellers, J. W., & Struhl, K. (1989) *Nature (London) 340*, 568-571.
Sellers, J. W., Vincent, A. C., & Struhl, K. (1990) *Mol. Cell. Biol.* (in press).
Stewart, J. M., & Young, J. D. (1984) *Solid Phase Peptide Synthesis*, Raven Press, New York, NY.
Struhl, K. (1987a) *Cell 50*, 841-846.
Struhl, K. (1987b) *Cell 49*, 295-297.
Struhl, K. (1988) *Nature (London) 332*, 649-650.
Studier, F. W., & Moffat, B. A. (1986) *J. Mol. Biol. 189*, 113-130.
Tregear, G. W., van Reitschoten, J., Sauer, R. T., Niall, H. D., Keutmann, H. T., & Potts, J. T. (1977) *Biochemistry 16*, 2817-2823.
Vinson, C. R., Sigler, P. B., & McKnight, S. L. (1989) *Science 246*, 911-916.
Vogt, P. K., Bos, T. J., & Doolittle, R. F. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3316-3319.
Walters, L., & Kaiser, E. T. (1985) *J. Am. Chem. Soc. 107*, 6422-6242.
Warrant, R. W., & Kim, S.-H. (1978) *Nature (London) 271*, 130-135.
Weiss, M. A., Ellenberger, T. E., Wobbe, C. R., Lee, J. P., Harrison, S. C., & Struhl, K. (1990) *Nature (London)* (in press).
Wemmer, D., Ribeiro, A. A., Bray, R. P., Wade-Jardetzky, N. G., & Jardetzky, O. (1981) *Biochemistry 20*, 829-836.

A Bifunctional Fusion Protein Containing Fc-Binding Fragment B of Staphylococcal Protein A Amino Terminal to Antidigoxin Single-Chain Fv†

Mei-Sheng Tai,‡ Meredith Mudgett-Hunter,§,‖ Douglas Levinson,‡ Gay-May Wu,‡ Edgar Haber,§,⊥ Hermann Oppermann,‡ and James S. Huston*,‡

Creative BioMolecules, Inc., 35 South Street, Hopkinton, Massachusetts 01748, Cellular and Molecular Research Laboratory, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, and The Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543

Received May 21, 1990; Revised Manuscript Received July 5, 1990

ABSTRACT: A bifunctional molecule was genetically engineered which contained an amino-terminal effector domain that bound immunoglobulin Fc (fragment B of staphylococcal protein A) and a carboxyl-terminal domain that bound digoxin [a single-chain Fv (sFv)]. Effector and sFv binding properties were virtually identical with those of the parent molecules, despite the proximity of the FB to the sFv combining site. This finding is unprecedented since in all molecules of the natural immunoglobulin superfamily, the antigen binding domain is amino terminal to the effector domain. The FB-sFv sequence was encoded in a single synthetic gene and expressed as a 33 106 molecular weight protein in *Escherichia coli*. After purification, renaturation, and affinity isolation, yields of active fusion protein were 110 mg/L of fermented cells (18.5-g cell paste). Bifunctionality was confirmed by the ability of FB-sFv to cross-link IgG to digoxin–bovine serum albumin, as measured by plate assays and by Ouchterlony analysis. Analysis of the expressed fusion protein suggests that the sFv holds promise for the development of multifunctional, targetable single-chain proteins.

The ability to target therapeutic agents with antibodies has been a long-term goal of medical research. The most elegant targetable proteins would consist of the minimum structures needed for selective delivery and effector function. Here we combined the minimal antigen binding site, a single-chain Fv,[1] with an individual Fc binding domain of staphylococcal protein A into a single polypeptide that expresses both digoxin and Fc-binding properties.

In immunoglobulins, separate heavy and light chains contribute the V_H and V_L variable domains that constitute the Fv region (Inbar et al., 1972) and form the antibody combining site. Recently, protein engineering methods have been used to link the V_H and V_L, creating functional single-chain Fv proteins (Huston et al., 1988a; Bird et al., 1988). The resulting

† This research was supported in part by NIH through Small Business Innovation Research Grant CA 39870 and Program Project Grant HL 19259.

* To whom correspondence should be addressed.
‡ Creative Biomolecules, Inc.
§ Massachusetts General Hospital.
‖ Harvard Medical School.
⊥ The Bristol-Myers Squibb Pharmaceutical Research Institute.

[1] Abbreviations: BSA, bovine serum albumin; CDR, complementarity-determining region; EDTA, ethylenediaminetetraacetic acid; Fab, antigen binding fragment derived from IgG by papain cleavage; FB, fragment B, a 58-residue domain of staphylococcal protein A; FB-sFv$^{26\text{-}10}$, fusion protein that comprises FB and 26-10 single-chain Fv; Fc, complement binding region of IgG that associates with the FB domain; FR, framework region; Fv, variable-region fragment consisting of noncovalently associated V_H and V_L domains; GaMFab, goat anti-mouse Fab antibody; Gdn-HCl, guanidine hydrochloride; hIgG, human immunoglobulin G; mIgG, murine immunoglobulin G; PBSA, 0.15 M NaCl + 0.05 M potassium phosphate, pH 7.0, + 0.03% NaN_3; rIgG, rabbit immunoglobulin G; SDS-PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; sFv (single-chain Fv), biosynthetic Fv analogue comprising both variable domains on a single polypeptide chain; V_H, heavy-chain variable region; V_L, light-chain variable region.

0006-2960/90/0429-8024$02.50/0