Covalent Attachment of Arc Repressor Subunits by a Peptide Linker Enhances Affinity for Operator DNA†

Clifford R. Robinson and Robert T. Sauer*

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

Received September 6, 1995; Revised Manuscript Received October 31, 1995⊗

ABSTRACT: By designing a recombinant gene containing tandem copies of the *arc* coding sequence with intervening DNA encoding the linker sequence GGGSGGGTGGGSGGG, the two subunits of the P22 Arc repressor dimer have been covalently linked to form a single-chain protein called Arc-L1-Arc. The 15-residue linker joins the C-terminus of one monomer to the N-terminus of the second, a distance of approximately 45 Å in the Arc-operator cocrystal structure. Arc-L1-Arc is expressed at high levels in *Escherichia coli*, with no evidence of degradation or proteolytic clipping of the linker, and is more active than wild-type Arc in repression assays. The purified Arc-L1-Arc protein has the molecular weight expected for the designed protein and unfolds cooperatively, reversibly, and with no concentration dependence in thermal-denaturation studies. Arc-L1-Arc protects operator DNA in a manner indistinguishable from that of wild-type Arc in DNase I and copper−phenanthroline footprinting studies, but the covalent attachment of the two monomers results in enhanced affinity for operator DNA. Arc-L1-Arc binds operator DNA half-maximally at a concentration of 1.7 pM, compared with the wild-type value of 185 pM, and also binds DNA fragments containing the left or right operator half-sites more tightly than wild type. Because wild-type Arc is monomeric at sub-nanomolar concentrations and must dimerize before binding to the operator, it was anticipated that Arc-L1-Arc would exhibit a lower half-maximal binding concentration. However, even when the change from a monomeric to a dimeric species is taken into account, the affinity of Arc-L1-Arc for operator and half-operator DNA is greater than the wild-type affinity. This tighter binding appears to result from slower dissociation, as Arc-L1-Arc DNA complexes with full or half-site operators dissociate at rates 5−10 times slower than the corresponding Arc−DNA complexes. Hence, the activity of the designed Arc-L1-Arc protein is substantially increased relative to wild-type Arc in a variety of assays.

The design of proteins with improved or novel functions is an important goal with a variety of therapeutic, research, environmental, and industrial applications. One approach is to design completely new proteins which possess the desired structural and functional properties (Bryson et al., 1995). However, substantial advancements in our understanding of the sequence determinants of protein folding and function will be required before this route becomes practical. An alternative and currently more feasible approach is to modify the properties of a natural protein to achieve a desired purpose (Petsko & Sligar, 1994).

Protein subunits in some multimeric complexes can be covalently attached through appropriate gene fusions. For example, in an early study, the α- and β- subunits of the *Escherichia coli* glycyl-tRNA synthetase were fused via short linkers to create a single-chain protein which was fully active (Toth & Schimmel, 1986). Covalent attachment of subunits using linkers and gene fusions has also been used to increase the stability, half-life, renaturability, and, in some cases, activity of dimeric proteins. Examples include the sweet-tasting protein monelin (Kim et al., 1989), CuZn superoxide dismutase (Hallewell et al., 1989), avian retroviral proteases (Bizub et al., 1991), the gene V single-stranded DNA-binding protein of phage fd (Liang et al., 1993), the 434 repressor (Percipalle et al., 1995), and the RNA-binding protein ROP (Predki & Regan, 1995). Single-chain antibodies with improved handing characteristics and novel specificities have also been generated (Bird et al., 1988; Huston et al., 1988; Rumbley et al., 1993; Mallender & Voss, 1994). Gene fusions also provide a method for generating hybrid proteins with novel specificities, activities, or combinations of functions. This approach seems especially well suited to transcription factors, which generally have modular structures as either multidomain proteins or homo-oligomeric proteins (Pabo & Sauer, 1992). Fusion of individual DNA binding domains from a zinc finger protein and a homeodomain protein has been shown to result in a protein capable of binding a novel DNA recognition sequence *in vitro* and *in vivo* (Pomerantz et al., 1995).

In the work reported here, we describe the construction and properties of a single-chain variant of the P22 Arc repressor, a member of the ribbon−helix−helix family of transcription factors (Raumann et al., 1994a). Wild-type Arc is a dimer of identical 53-residue subunits, which controls transcription during lytic growth of bacteriophage P22 (Susskind & Youderian, 1983). In the Arc dimer, residues 8−14 from each monomer pair to form a two-stranded, antiparallel β-sheet, which inserts into the major groove of an operator half-site to make base-specific contacts (Raumann et al., 1994b). Residues N-terminal to the β-sheet form an arm which folds along the DNA backbone, and residues C-terminal to the β-sheet form α-helices which pack against the β-sheet to form the globular protein structure. The Arc complex with the full 21-base-pair operator is assembled by sequential addition of dimers to operator half-sites (Brown et al., 1990; Brown & Sauer, 1993). Cooperative interactions

† Supported by NIH Grant AI-16892 and an NIH/NRSA postdoctoral fellowship to C.R.R.

⊗ Abstract published in *Advance ACS Abstracts,* December 15, 1995.

0006-2960/96/0435-109$12.00/0

between the DNA-bound dimers help to stabilize this tetrameric complex (Brown & Sauer, 1993).

A central feature of the Arc system is that the processes of protein folding, dimerization, and DNA binding are energetically coupled. At high concentrations, Arc is a reasonably stable dimer ($K_d \approx 10^{-8}$ M). However, at the sub-nanomolar concentrations where half-maximal operator binding is observed, Arc dimers dissociate and most molecules exist as unfolded monomers (Bowie & Sauer, 1989a; Brown et al., 1990). Accordingly, the favorable free energy of DNA binding is reduced because some energy must be expended to form active dimers at low protein concentrations. We reasoned that covalent attachment of Arc monomers should yield a single-chain protein which would obviate the requirement for dimerization and thus might exhibit increased DNA binding activity. Indeed, we find that a designed single-chain Arc variant shows substantial improvements in binding to operator DNA *in vitro* and is a more active repressor *in vivo*.

MATERIALS AND METHODS

Molecular Modeling and Linker Design. Modeling was performing using Insight II molecular graphics software (Biosym) and the structure of a DNA-bound Arc dimer from the cocrystal structure (Raumann et al., 1994b). Using the sequence of linkers from single-chain antibodies as a model (Huston et al., 1988), peptide linkers were constructed by adding one residue at a time and adjusting dihedral angles by trial and error to connect the C-terminus of one monomer and the N-terminus of the second monomer for several distinct paths around the outside of the dimer. In no case did the precise orientation or conformation of the linker appear to be critical, although the linker conformations were constrained to have reasonable peptide backbone dihedral angles and to avoid steric clashes (Ramachandran & Sasisekharan, 1968). The shortest path between the N- and C-termini crosses the DNA binding interface and was avoided for this reason.

DNA Fragments and Labeling. DNA oligonucleotides were synthesized using an Applied Biosystems 381A DNA synthesizer. Oligonucleotides used for the construction of the Arc-L1-Arc gene were purified by polyacrylamide gel electrophoresis using standard procedures (Sambrook et al., 1989). Double-stranded cassettes were annealed by mixing equimolar amounts of the complementary oligonucleotides, heating to 90 °C, and cooling slowly. Oligonucleotides used for gel electrophoretic mobility shift assays were synthesized with the trityl group on, purified by reverse phase HPLC, detritylated, and repurified by HPLC. Oligonucleotides used as primers for PCR reactions were purified by reverse phase HPLC.

End-labeling of single-stranded oligonucleotides with γ-^{32}P-labeled ATP and T4 polynucleotide kinase was carried out as described (Brown & Sauer, 1993). For end-labeled oligonucleotides used to construct half-site operator fragments, a 2-fold excess of the complementary strand was added, and the strands were heated to 90 °C and allowed to anneal during a slow-cooling step. The mixture was extracted with phenol−chloroform and passed over a Sephadex G-25 spin column to separate the labeled duplex from unincorporated nucleotides. The double-stranded DNA fragment containing the intact operator has 5′ overhangs and was labeled by end-filling using α-^{32}P-labeled ATP and sequenase v2.0 (United States Biochemical) (Ades & Sauer, 1995). The reaction was heat-treated to inactivate the enzyme, and the labeled duplex was purified as described above.

Plasmid Construction. The gene encoding the Arc-L1-Arc protein was constructed in three steps. Plasmid pLA11A (containing the first copy of the *arc* gene and the first half of the linker) was constructed by cloning a cassette with the sequence:

```
5′agc ttt aag aag gaa ggg cgc att ggc gcg GGC GGC GGC TCC GGA
      aa ttc ttc ctt ccc gcg taa ccg cgc CCG CCG CCG AGG CCT

GGC GGT ACC TAA TAG TAA T
CCG CCA TGG ATT ATC ATT AGC 5′
```

where lower case letters indicate bases from the coding region of the *arc* gene, between the unique *Hin*DIII and *Cla*I sites in plasmid pSA700 (Milla et al., 1993). Plasmid pLA11B (containing the second half of the linker, the second *arc* gene, and the st11 tail (Milla et al., 1993)) was constructed by cloning a cassette with the sequence:

```
5′TC GAG ATG GGT ACC GGT GGC GGC TCC GGA GGC GGC atg aaa gga
       C TAC CCA TGG CCA CCG CCG AGG CCT CCG CCG tac ttt cct

atg agc aaa atg ccg cag ttc aac ctg agg tgg ccg c
tac tcg ttt tac ggc gtc aag ttg gac tcc acc gg 5′
```

between the unique *Xho*I and *Sac*II sites in another copy of plasmid pSA700. Finally, plasmid pLA110 was constructed by subcloning the *Kpn*I−*Cla*I fragment from pLA11B (containing the linker portion and second *arc* gene) between the unique *Kpn*I and *Cla*I sites of pLA11A. The gene encoding Arc-L1-Arc in pLA110 is under transcriptional control of the *tac* promoter. The st11 tail sequence (HHHHHHKNQHE) was included to facilitate purification by nickel-chelate chromatography (Milla et al., 1993). The DNA sequence encoding Arc-L1-Arc in pLA110 was determined using the dideoxy sequencing method (Sanger et al., 1977) to confirm the fidelity of the construction.

Activity Assays in Vivo. The repressor activity of the Arc-L1-Arc protein *in vivo* was assayed in *E. coli* strain UA2F, which contains a recessive allele of a streptomycin resistance gene (Bowie & Sauer, 1989b). Plasmid pSA700 (or the derivative plasmid pLA110) contains a dominant streptomycin sensitivity gene under control of an Arc-repressible promoter. Thus UA2F/pSA700 or UA2F/pLA110 cells expressing functional Arc are resistant to streptomycin. Cells were grown either in liquid culture or on agar plates on LB media containing 150 μg/mL ampicillin and 50 μg/mL streptomycin to test for repression by the Arc-L1-Arc protein. Strain UA2F also contains a chromosomal fusion of the Arc-repressible P_{ant} promoter to the *lacZ* gene (Vershon et al., 1986). UA2F cells containing pSA700 (encoding wild-type Arc), pTA200/Δarc (a plasmid from which the *arc* gene is deleted), or pLA110 (encoding Arc-L1-Arc) were grown in liquid culture on LB media at 37 °C to midexponential phase (OD_{600} = 0.4−0.6) and lysed, and β-galactosidase activity was assayed using the colorimetric substrate *o*-nitrophenyl β-D-galactoside (Miller, 1972; Schildbach et al., 1995). Replicate experiments were performed with growth to an OD_{600} of 0.1, and an additional series of assays was performed with growth at 25 and 30 °C. Values reported in Table 3 represent the average of at least four independent determinations of activity.

Protein Expression and Purification. For analysis of expression, 5 mL cultures of *E. coli* strain X90 harboring plasmid pLA110 were grown at 37 °C in LB broth plus 200

μg/mL ampicillin. At an OD_{600} of 0.8, IPTG was added to a final concentration of 100 μg/mL. A duplicate culture did not receive IPTG. After 2 h of continued growth, cells were harvested by centrifugation from 1.5 mL of culture. The cells were resuspended in 0.2 mL of Laemmli loading buffer (Laemmli, 1970) and boiled for 2 min, and insoluble material was removed with a toothpick. Samples were visualized by Coomassie staining following SDS−PAGE using the Tris−tricine buffer system (Schägger & von Jagow, 1987).

The His_6 sequence in the st11 tail of Arc-L1-Arc allows substantial purification using Ni^{2+}−chelate chromatography as described previously for Arc-st11 (Milla et al., 1993). Briefly, cells were grown for 2 h following IPTG induction, harvested by centrifugation, and lysed by stirring in a buffer containing 6 M guanidinium hydrochloride (GuHCl) at pH 8. After centrifugation, the supernatant was loaded onto a column containing Ni^{2+}-NTA resin (Qiagen). After washing in buffer containing 6 M GuHCl plus 5 mM imidazole, Arc-L1-Arc was eluted from the column with 0.2 M acetic acid plus 6 M GuHCl. The eluate was dialyzed into a buffer containing 10 mM Tris-HCl (pH 7.5) and 0.2 mM EDTA and further purified by ion-exchange chromatography on SP-Sephadex. Samples from each step of the purification were analyzed by SDS−PAGE.

Protein Stability and Spectroscopy. Spectroscopy and analytical ultracentrifugation experiments were performed in a buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM KCl, and 0.2 mM EDTA. The concentration of wild-type Arc was calculated in terms of monomer equivalents using an extinction coefficient at 280 nm of 6756 M^{-1} cm^{-1} (Brown et al., 1990); the concentration of Arc-L1-Arc was calculated using an extinction coefficient twice as large. UV absorbance spectroscopy was performed using a Hewlett Packard 8452A diode array spectrophotometer. Fluorescence spectra were measured using a Perkin Elmer LS-50 luminescence spectrophotometer at a protein concentration of 5 μM. Fluorescence emission was measured from 300 to 400 nm, following excitation at 280 nm. Circular dichroism (CD) spectra at 5 and 20 μM protein concentrations were measured using an AVIV 60DS spectrophotometer. For thermal unfolding experiments, the temperature was increased in 2 °C steps, samples were equilibrated for 60 s, and the ellipticity at 222 nm was measured for 30 s. The denaturation curve was fit to a two-state transition for a monomeric species to determine the apparent T_m (Pace & Lauerents, 1989).

Analytical ultracentrifugation experiments were performed using a Beckman Optima XL-A centrifuge. Arc-L1-Arc samples at 2.7, 27.5, and 110 μM were centrifuged overnight at 15 000 or 25 000 rpm. Absorbance readings were performed at 1-h intervals to determine that the samples had reached equilibrium. Absorbance profiles were measured at either 222, 236, or 280 nm, depending upon the protein concentration, to give absorbance readings between 0.1 and 1.0. Ten scans were averaged and then analyzed according to Laue et al. (1992) to give apparent molecular weights. Based upon amino acid composition, the partial specific volume of Arc-L1-Arc was estimated to be 0.724 using the method of Cohn and Edsall (Cohn & Edsall, 1943; Durchschlang, 1986).

DNA Footprinting. DNA for footprinting experiments was prepared using the polymerase chain reaction (PCR) to amplify a 130-base sequence containing the *arc* operator (Smith & Sauer, 1995). Appropriate end-labeling of the PCR product was achieved by using one unlabeled and one end-labeled primer. The ^{32}P-labeled PCR product was gel purified, and the specific activity was estimated using standard methods. DNase I and copper−phenanthroline footprinting reactions were performed as described (Smith & Sauer, 1995). Gels were exposed using Kodak XAR5 film with intensifying screens at −70 °C.

DNA Binding Assays. Equilibrium binding assays were performed essentially as described (Brown & Sauer, 1993). Arc-L1-Arc or wild-type Arc was incubated with ^{32}P-labeled DNA at room temperature (25 °C) for at least 4 h in a binding buffer containing 10 mM Tris-HCl (pH 7.5), 3 mM $MgCl_2$, 0.1 mM EDTA, 100 mM KCl, 100 μg/mL BSA, and 0.02% Nonidet NP-40. The full-length operator fragment was used at a concentration of 50 fM. Fragments containing the left or right operator half-sites or nonspecific DNA were used at concentrations of 2 pM. Protein concentrations were always at least 10-fold higher than the concentration of the DNA. Glycerol was added to 5% total volume just prior to loading the samples onto 8% polyacrylamide gels in 0.5× TBE buffer. Gels were dried and exposed using Molecular Dynamics phosphoimager screens, and band intensities quantitated using ImageQuant software using the Volume Measurement utility. The fraction of DNA bound was determined as the volume of the slower migrating band (corresponding to bound DNA) divided by the sum of all bands.

Because Arc-L1-Arc is always present in significant excess of the DNA, the concentration of free protein can be closely approximated by using the total protein concentration $[P_t]$. For Arc-L1-Arc binding to an operator half-site, the binding isotherm simplifies to:

$$\Theta \approx \frac{1}{1 + K_d/[P_t]} \tag{1}$$

where Θ is the fraction of bound DNA, and K_d is the equilibrium dissociation constant:

$$K_d = \frac{[P][DNA]}{[P\cdot DNA]} \tag{2}$$

Data were fit to eq 1 by nonlinear least-squares analysis using Kaleidagraph software running on a Macintosh computer. For two Arc-L1-Arc chains binding to the full-length operator, the equilibrium dissociation constant is defined as:

$$K_d = \frac{[P]^2[DNA]}{[P_2\cdot DNA]} \tag{3}$$

and:

$$\Theta \approx \frac{1}{1 + K_d/[P_t]^2} \tag{4}$$

The binding of Arc-L1-Arc to a short nonspecific DNA fragment (NS1) was fit using eq 1. In this case, K_d is a macroscopic binding constant. We assume that the 19 bp NS1 fragment contains nine equivalent and overlapping binding sites of 11 bp each. The microscopic equilibrium dissociation constant for a single nonspecific site was calculated as $9K_d$. Isotherms for wild-type Arc binding to operator, half-operator, and nonspecific DNA were fit as described (Brown et al., 1990; Brown & Sauer, 1993).

Kinetic analysis of Arc-L1-Arc dissociation from operator and half-operator DNA was performed by standard methods (Brown & Sauer, 1993). Arc-L1-Arc was incubated at a

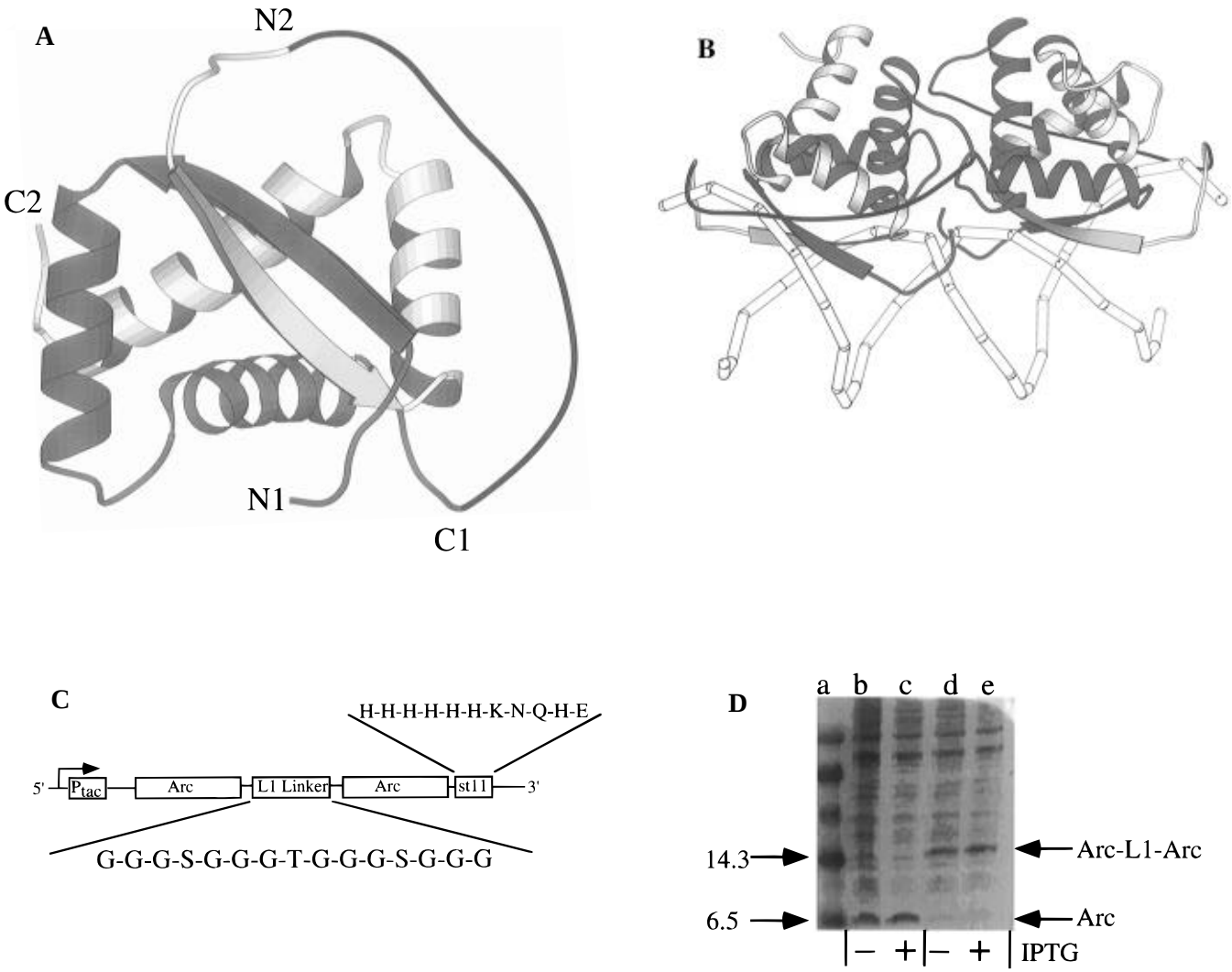

FIGURE 1: (A) Model of the Arc-L1-Arc protein, shown in ribbon trace. The first subunit is colored dark gray; the second subunit is light gray. The L1 linker is colored black, and the conformation shown is one of several possible. The positions of the N- and C-termini in the two subunits of the wild-type Arc dimer are indicated. Coordinates of the DNA-bound form of the Arc dimer were taken from Raumann et al. (1994b). Panels A and B were generated using MOLSCRIPT (Kraulis, 1991). (B) Model of two Arc-L1-Arc molecules bound to operator DNA based on the wild-type cocrystal structure (Raumann et al., 1994b). Coloring is as in panel A. Arc-L1-Arc can be modeled in either of two orientations in each half-site, which are related by a 180° rotation around a local pseudo-twofold axis. The orientations shown are arbitrary. (C) Schematic of the Arc-L1-Arc gene and expression system. Amino acids are lised using the one-letter code. (D) SDS−PAGE of molecular weight standards (a) and cell extracts from strains expressing Arc from pSA700 (b, c) or Arc-L1-Arc from pLA110 (d, e).

concentration of 200 pM with the operator half-site fragments or at a concentration of 20 pM with the whole-operator fragment. For the half sites, the incubation time was 2 h. For the full site, samples were incubated overnight. Dissociation was initiated by a 40-fold dilution into binding buffer containing 5% glycerol, and at appropriate time points samples were loaded immediately onto an 8% polyacrylamide gel running at 300 V. Kinetic data were fit to a single exponential decay using Kaleidagraph. Association rate constants were calculated by dividing the dissociation rate constant by the equilibrium dissociation constant.

RESULTS AND DISCUSSION

Linker Design. Using molecular modeling, we built linker sequences to connect the C- and N-termini of the Arc monomers in a DNA-bound dimer taken from the cocrystal structure (Raumann et al., 1994b). Because the N-terminal arms of Arc are disordered in solution (Breg et al., 1990; Bonvin et al., 1994) but fold against the DNA in the complex, we felt that modeling based upon the cocrystal structure gave the most reasonable chance for a design that would not interfere with the arm-mediated DNA contacts. The shortest path between the C- and N-termini of two subunits (≈35 Å) crosses the DNA binding interface. However, linkers spanning ≈45 Å could be modeled to pass around either side of the dimer (adjacent to the DNA helix), or around the back of the protein (on the opposite side of the protein from the DNA helix). By each of these paths, a peptide linker of 15 residues was sufficient to connect the two monomers. The linker sequence was based upon those used in the constructon of single-chain antigen binding proteins (Huston et al., 1988). The sequence (GGGSGGGTGGGSGGG) was chosen to maximize flexibility and solubility, and to allow the introduction of restriction sites for cloning and gene construction. We refer to this sequence as the L1 linker. Panels A and B of Figure 1 show one of the modeled linker conformations in the free dimer and the protein−DNA complex.

Expression and Purification. A gene encoding the single-chain Arc-L1-Arc protein was constructed in plasmid pLA110 under control of the IPTG-inducible P_{tac} promoter

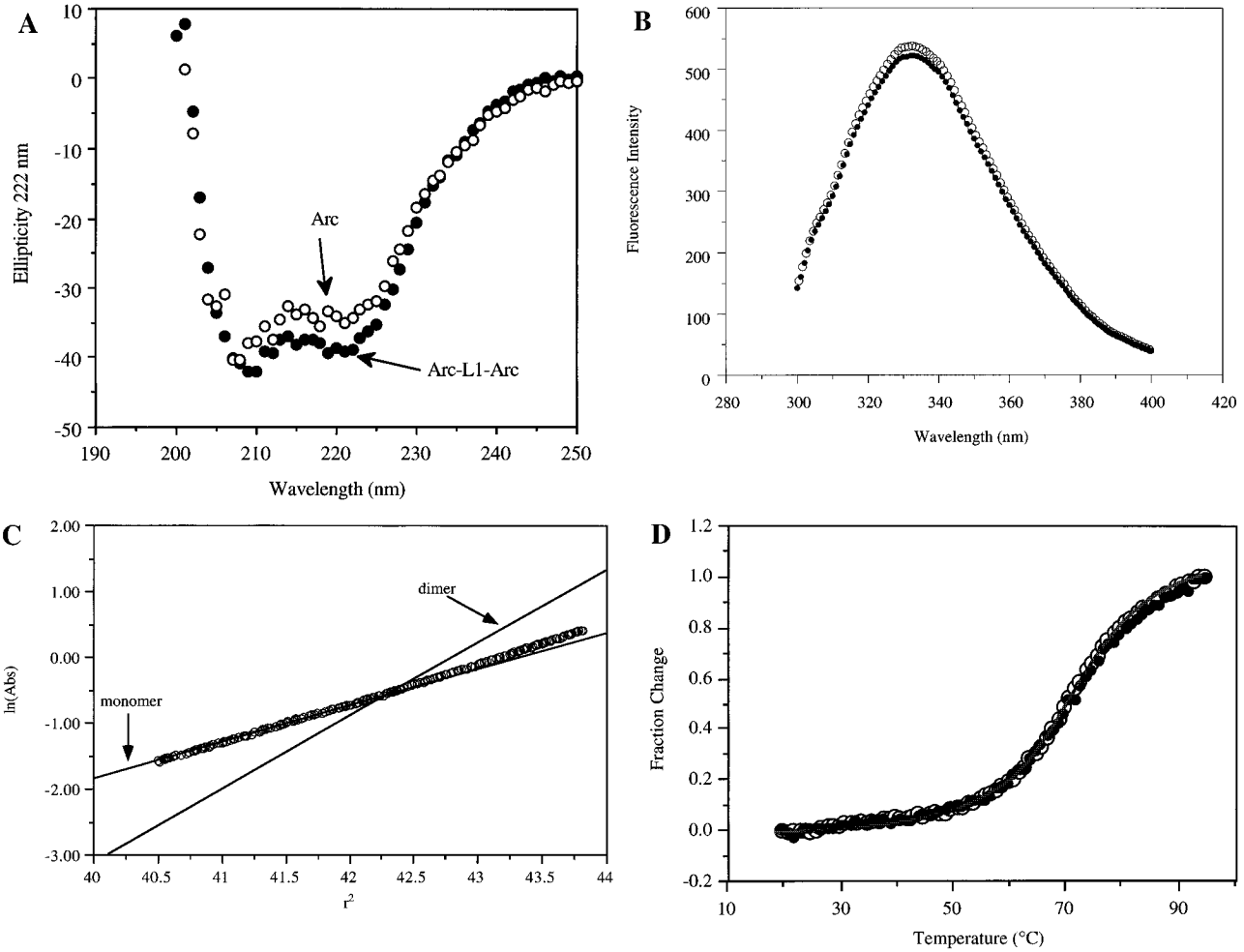

FIGURE 2: (A) Circular dichroism spectra of purified Arc and Arc-L1-Arc. Filled circles, 5 μM Arc-L1-Arc; open circles, 5 μM Arc (total dimer equivalents). Assuming a dissociation constant of 1.8×10^{-8} M (Milla & Sauer, 1994), the concentration of folded Arc dimers at this concentration is 4.8 μM. (B) Fluorescence emission spectra of Arc-L1-Arc and Arc repressor. Solid circles, 5 μM Arc-L1-Arc; open circles, 5 μM Arc (total dimer equivalents). Excitation wavelength was 280 nM. (C) Analytical ultracentrifugation profile of Arc-L1-Arc at 27 μM, 25 000 rpm, 25 °C. Symbols represent experimental data points collected at 236 nm. One solid line indicates the profile expected for a monomeric form of Arc-L1-Arc with molecular mass of 14 840 Da. The other line indicates the profile expected for a dimeric protein with molecular mass of 29 680 Da. Assuming a partial specific volume of 0.724, the experimental data fit best to an apparent molecular mass of 16 430 Da. (D) Thermal unfolding curves for Arc-L1-Arc at concentrations of 5 μM (closed circles) and 20 μM (open circles). CD signal was monitored at 222 nm. The line indicates the best fit of the 20 μM profile with an apparent T_m of 68.6 ($\pm$ 0.8) °C.

(Figure 1C). Upon induction with IPTG, *E. coli* X90/pLA110 cells produce a protein of approximately the molecular weight expected for Arc-L1-Arc, as judged by SDS−PAGE (Figure 1D). Visual inspection of Coomassie-stained gels showed that Arc-L1-Arc was expressed at levels similar to wild-type Arc. The Arc-L1-Arc protein was purified to >95% homogeneity. Smaller molecular weight species, corresponding to proteolytic fragments or monomeric Arc, were not detected during any phase of the overexpression or purification. This possibility had been a concern. If the linker interfered with protein folding in the cell or was too flexible, then degradation of the entire recombinant protein or clipping within the linker would not have been surprising.

Structural Properties, Stability, and Oligomeric Form. The circular dichroism and UV absorbance and fluorescence spectra of the purified Arc-L1-Arc protein are very similar to the wild-type spectra at equivalent subunit concentrations (Figure 2, panels A and B). Thus, Arc-L1-Arc appears to have the same general structural features as the Arc dimer—a mixture of α-helix and β-sheet, with tryptophan residues buried in a hydrophobic environment.

Analytical ultracentrifugation experiments performed at either 15 000 or 25 000 rpm gave distribution profiles for Arc-L1-Arc which fit to an average apparent molecular mass of 16 430 ($\pm$200) Da, equal to 1.11 ($\pm$0.01) times the expected molecular mass of Arc-L1-Arc (Figure 2C). These results were essentially unchanged over a range of protein concentrations from 2.7 to 110 μM, indicating that the predominant species of Arc-L1-Arc is a monomer—there is no evidence for significant "cross-folding" of the Arc subunits to form a dimeric $[\text{Arc-L1-Arc}]_2$ species or for other aggregation or linear polymerization reactions. The slight increase in apparent molecular mass relative to that expected for a monomer could result from small errors in the estimation of the partial specific volume.

Thermal unfolding of the Arc-L1-Arc protein monitored by changes in CD ellipticity at 222 nM shows a cooperative and reversible unfolding transition with a T_m of $\approx$69 °C (Figure 2D). No change in T_m was observed when the concentration was increased from 5 to 20 μM. By contrast, the T_m for dimeric wild-type protein is concentration dependent and increases by about 6 °C over the same concentration range (Bowie & Sauer, 1989a; unpublished data). These results confirm that Arc-L1-Arc forms a monomeric, cooperatively-folded, native-like structure.

Table 1: DNA Binding Parameters[a]

DNA fragment[b]	Arc-L1-Arc		Arc	
	half-maximal binding	K_d[c]	half-maximal binding	K_d[c]
full operator	1.7 pM	$2.8 (\pm 1.6) \times 10^{-24}$ M^2	185 pM	$5.8 (\pm 1.1) \times 10^{-23}$ M^2
left half-site	27 pM	$2.7 (\pm 1.3) \times 10^{-11}$ M	1.3 nM	$3.8 (\pm 0.6) \times 10^{-10}$ M
right half-site	86 pM	$8.6 (\pm 3.3) \times 10^{-11}$ M	1.8 nM	$7.2 (\pm 1.0) \times 10^{-10}$ M
nonspecific DNA	47 nM	$3.9 (\pm 2.1) \times 10^{-7}$ M	230 nM	$1.7 (\pm 0.6) \times 10^{-6}$ M

[a] Binding assays were performed at 25 °C, in buffer containing 10 mM Tris-HCl (pH 7.5), 3 mM EDTA, 100 mM KCl, 100 μg/mL BSA, and 0.02% Nonidet NP-40. [b] See Figure 3 for fragment sequences. [c] K_d values (±SD) for Arc are calculated for the preformed dimer using a dimerization constant of 1.8×10^{-8} M (Milla & Sauer, 1994). K_d values (±SD) for nonspecific DNA are microscopic constants assuming each fragment contains 9 equivalent sites. Half-maximal concentrations for Arc are listed in total dimer equivalents.

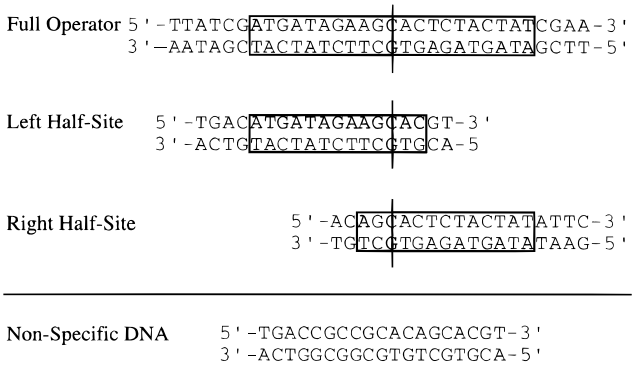

FIGURE 3: Sequences of DNA fragments sites used in binding studies. Operator and half-operator sequences are boxed. The central CG base pair in the full operator is indicated by a vertical line. For orientation, this CG base pair is also marked in the half-sites.

Operator DNA Binding. The sizes of Arc-L1-Arc complexes with different operator DNA fragments (see Figure 3) were compared with wild-type complexes in electrophoretic mobility shift assays. Arc-L1-Arc appears to bind to an operator half-site as a monomer, as it produces a mobility shift identical to that of the wild-type Arc dimer (Figure 4A). Binding of Arc-L1-Arc to full-site operators produces a shift of the same magnitude as the binding of wild-type Arc tetramers, indicating that two Arc-L1-Arc chains bind to the intact operator (Figure 4B). If a single DNA binding domain were formed by cross-folding of Arc domains from two different Arc-L1-Arc chains, then significantly larger shifts would have been expected for the half-site operator and possibly for the whole-site operator. DNaseI and copper−phenanthroline DNA footprinting experiments using Arc-L1-Arc gave wild-type patterns of cleavage protection and enhancements, suggesting that Arc-L1-Arc binds to the operator with the same overall orientation as wild-type Arc (data not shown).

Binding isotherms of Arc-L1-Arc for operator and non-operator DNA fragments, whose sequences are shown in Figure 3, were determined using electrophoretic mobility shift assays, and equilibrium dissociation constants were determined by fitting these data using eqs 1 and 4. Several points are worth noting: (i) Arc-L1-Arc half-maximally binds operator DNA at significantly lower concentrations than wild-type Arc (Figure 4A,B, Table 1). This is also true for binding to half-operator sites and nonspecific DNA. (ii) The binding of Arc-L1-Arc to the half-site operators is first-order in protein concentration, whereas binding to the whole operator is second-order in protein concentration (Figure 4B). Because analytical ultracentrifugation and thermal denaturation experiments show that Arc-L1-Arc is predominantly monomeric at concentrations much higher than those used for the DNA binding experiments, the observed reaction orders support the model that a single Arc-L1-Arc molecule

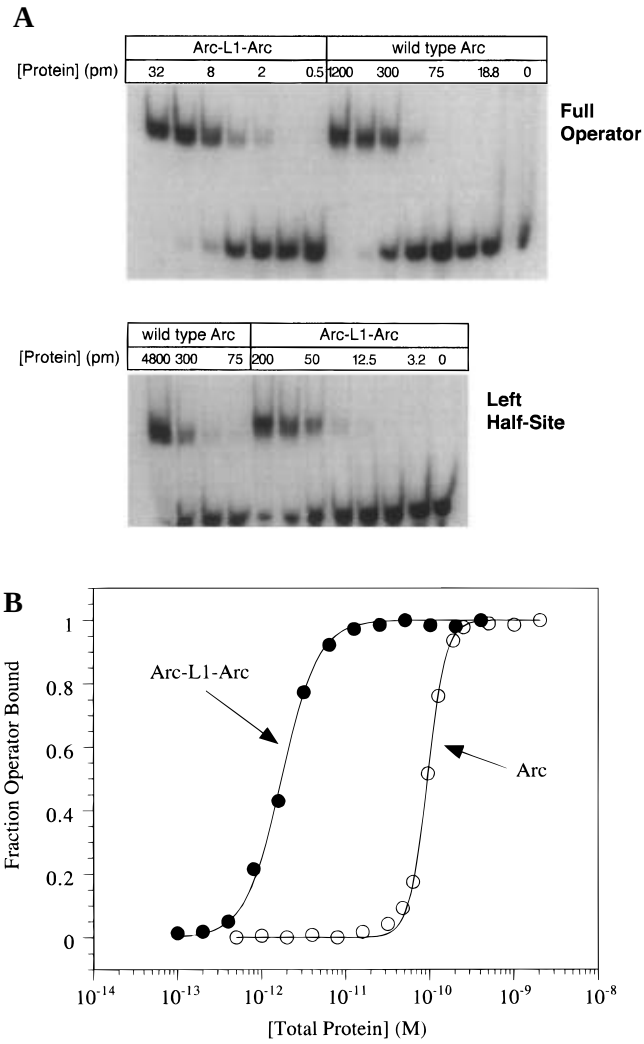

FIGURE 4: (A) Autoradiograms of DNA electrophoretic mobility shift assays. Top panel: Arc-L1-Arc and Arc binding to full operator. Unless indicated, protein concentrations were increased in 2-fold increments from right to left. Bottom panel: Arc and Arc-L1-Arc binding to the left half-site operator. In both gels, the right lane is a no-protein control. (B) Operator-binding isotherms. The symbols indicate the experimental data points. The lines represent the best theoretical fits. To facilitate comparison of the two proteins, the concentration is plotted in monomer equivalents for the single-chain Arc-L1-Arc and dimer equivalents for Arc. For Arc-L1-Arc (closed circles), the data were fit to eq 4, giving a $K_d = 2.8 \times 10^{-24}$ M^2. For wild-type Arc (open circles), the data were fit to eq 7 of Brown and Sauer (1993) using a dimerization constant K_u of 1.8×10^{-8} M, giving a $K_d = 5.8 \times 10^{-23}$ M^2. In both experiments, the DNA concentration was 50 fM.

binds to a half-site and that two molecules bind to the whole operator. (iii) It was anticipated that Arc-L1-Arc would bind operator DNA at lower concentrations than wild-type Arc, because the wild-type protein must dimerize before binding. However, even when this effect is eliminated by using the

Table 2: Rate Constants for Operator Binding Reactions[a]

	Arc-L1-Arc		
DNA fragment	half-life	k_{off}	k_{on}
left half-site	28 s	0.02 s^{-1}	7.4×10^8 M^{-1} s^{-1}
full operator	304 min	3.8×10^{-5} s^{-1}	1.4×10^{19} M^{-2} s^{-1}
	Arc		
DNA fragment	half-life	k_{off}	k_{on}
left half-site	2 s	0.35 s^{-1}	1.1×10^9 M^{-1} s^{-1}
full operator	72 min	1.6×10^{-4} s^{-1}	5.5×10^{18} M^{-2} s^{-1}

[a] Temperature and buffer conditions are listed in Table 1.

dimer dissociation constant to calculate the affinity of preformed wild-type dimers for operator DNA, Arc-L1-Arc still has approximately 10-fold higher affinity for the operator and half-operator sites, and about 5-fold higher affinity for nonspecific DNA (Table 1).

The kinetics of the dissociation reactions were studied using mobility shift assays to monitor operator or half-operator occupancy following dilution. The experimentally determined dissociation rate constants and the calculated association rate constants are listed in Table 2. The complexes of Arc-L1-Arc with whole-operator or half-operator fragments dissociate 5- to 10-fold more slowly than the corresponding wild-type complexes. These increases in the kinetic stability of Arc-L1-Arc·DNA complexes are consistent with the increased affinities measured in equilibrium binding assays. As observed previously for wild-type Arc, the association rates for Arc-L1-Arc binding to operator or half-operator fragments are close to the diffusion limit (Table 2).

The approximate 10-fold increase in the affinities and half-lives of Arc-L1-Arc complexes compared to wild-type Arc complexes has several possible origins. The energies required (1.5−3 kcal/mol) could be obtained by the formation of a few new hydrogen bonds (≈1−2 kcal/mol each) or van der Waals contacts (≈0.2−0.3 kcal/mol per atom) between atoms in the linker and the sugar−phosphate backbone or nucleotide bases in the DNA site (Creighton, 1984). Alternatively, the linker might help to position residues in the N-terminal arms to make better contacts with the DNA. These N-terminal residues are disordered in wild-type Arc when it is not bound to DNA (Breg et al., 1990; Bonvin et al., 1994). Simply by restricting the number of conformations accessible to the arms in a disordered state, the linker might reduce the entropic penalty that accompanies folding of the N-terminal arm during formation of the protein−DNA complex (Spolar & Record, 1994). Because the linker also slows the rate at which complexes dissociate, unfolding of the wild-type arm would need to play a role in the rate-determining step of complex dissociation for the entropic model to explain both the equilibrium and kinetic results. Finally, small changes in the structure of the overall protein−DNA complex could result in new or strengthened interactions involving regions of Arc other than the arms. We note, however, that if such changes do occur, they do not alter the chemical or enzymatic footprint of Arc-L1-Arc bound to the operator.

Cooperativity. Using the equilibrium binding constants measured for the whole-site and half-site operators, the free-energy cycle shown in Figure 5 was constructed to calculate the magnitude of cooperative interactions between adjacently bound Arc-L1-Arc molecules on the intact operator. When

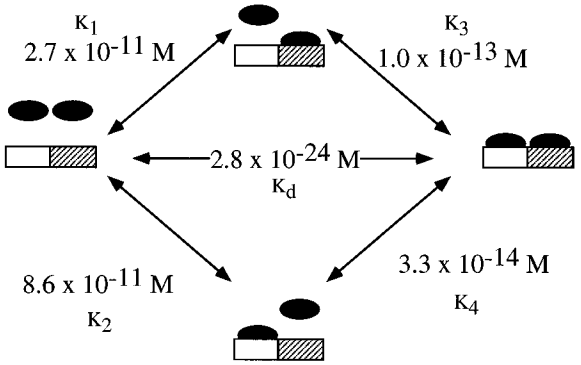

FIGURE 5: Thermodynamic cycle for binding of two Arc-L1-Arc proteins (black ovals) to the full operator (rectangles). K_1, K_2, and K_d are the experimentally determined equilibrium dissociation constants for Arc-L1-Arc binding to right half-site, left half-site, and full-length operator sequences, respectively. K_3 and K_4 are the calculated equilibrium dissociation constants for binding of the second Arc-L1-Arc molecule. The cooperative enhancement for binding of the second protein is calculated as K_1/K_4 or K_2/K_3.

a single Arc-L1-Arc protein is bound to the operator, the second Arc-L1-Arc protein binds approximately 830-fold more strongly corresponding to cooperative interactions of 3.9 kcal/mol. A similar estimate of cooperative stabilization (530-fold; 3.7 kcal/mol) can be calculated by comparing the half-lives of Arc-L1-Arc complexes with operator and half-operator DNA (Table 2). These values are slightly lower than the 5.1 kcal/mol of cooperative enhancement measured for interactions between wild-type Arc dimers under the same solution conditions (Brown & Sauer, 1993). This modest decrease in cooperativity does not result in detectable populations of singly occupied half-sites in binding experiments using the full-length operator (Figure 4), but may indicate that the L1 linker does influence formation of the cooperative binding interface. The linker can be modeled to pass close to this interface.

Arc-L1-Arc Repressor Activity in Vivo. As discussed above, the Arc-L1-Arc protein is expressed at levels similar to wild type in the cell. Can the single-chain protein also act as an efficient repressor of transcription? To test this, we employed two different assays for intracellular activity. In the first, Arc or an Arc variant must repress expression of a protein that gives rise to a dominant streptomycin-sensitive phenotype to allow the cell to grow in the presence of streptomycin (Bowie & Sauer, 1989b). In the presence of 50 μg/mL streptomycin, *E. coli* UA2F cells harboring pLA110 grew at rates indistinguishable from UA2F cell harboring pSA700 which encodes wild-type Arc. This result indicates that the Arc-L1-Arc protein is functional as a repressor *in vivo*. To provide a more quantitative assessment of repressor activity, we assayed the ability of Arc-L1-Arc to repress a chromosomal fusion of the P_{ant} promoter to the *lac*Z gene. As shown in Table 3, 2−3 times lower levels of β-galactosidase are expressed in cells containing the Arc-L1-Arc construct than in otherwise isogenic cells containing wild-type Arc. This result suggests that Arc-L1-Arc has higher repressor activity than wild-type Arc. Although we have not ruled out that somewhat higher levels of Arc-L1-Arc are produced in the cell under the noninducing conditions used for the repression assays, the observed increase in Arc-L1-Arc activity *in vivo* is most easily explained by the increased operator-binding activity of this protein.

Table 3: Repression of the P_{ant}−β-Galactosidase Fusion in Strain UA2F[a]

temp (°C)	pLA110 (Arc-L1-Arc)	pSA700 (Arc-st11)	pTA200/Δarc (no Arc)
37	320 ± 40	780 ± 60	1650 ± 70
30	440 ± 30	1170 ± 80	2410 ±110
25	790 ± 80	2150 ± 180	4820 ± 250

[a] β-Galactosidase activities expressed in Miller units (Miller, 1972).

Summary. We have designed a single-chain variant of the P22 Arc repressor in which the C-terminus of one subunit is covalently joined to the N-terminus of the second subunit by a 15-residue peptide linker. The resulting protein, Arc-L1-Arc, is expressed at comparable levels but is more active than the wild-type Arc dimer in the cell. The purified Arc-L1-Arc protein is a stable, soluble, cooperatively-folded protein which binds operator DNA more tightly than wild type. As a result, Arc-L1-Arc meets or exceeds all the design criteria. This work represents the first instance of which we are aware in which both enhanced activity *in vivo* and enhanced DNA binding affinity *in vitro* have been achieved via covalent linkage of the subunits of a transcription factor.

The L1 peptide appears to be one of the longest successfully designed linkers—only single-chain antibody linkers are of equivalent length (Huston et al., 1991)—suggesting that it may be feasible to covalently join protein subunits even when the appropriate C- and N-termini are not close in space. Although modeling studies suggest that the linker in Arc-L1-Arc must transverse a significant distance across the protein surface, the linker does not interfere with protein folding or activity and is not cleaved by proteases in the cell.

Single-chain Arc proteins should provide the opportunity to address several questions concerning molecular recognition in protein−protein and protein−DNA complexes. By enabling amino acids in the two Arc domains to be varied independently, the Arc-L1-Arc protein provides a system in which to dissect the roles of specific interactions in the binding of a symmetric dimer to an asymmetric DNA sequence. Additionally, single-chain Arc proteins should serve as models to test the effects of linker length and sequence upon stability, folding, and activity, and to evaluate the extent to which these parameters can be optimized in other systems.

ACKNOWLEDGMENT

We thank Dr. Cathy Rumbley for helpful discussions and suggestions and Professors Tania Baker, Carl Pabo, and Paul Schimmel for the use of equipment.

REFERENCES

Ades, S. E., & Sauer, R. T. (1995) *Biochemistry 34*, 14601−14608.
Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S.-M., Lee, T., Pope, S. H., Riordan, G. S., & Whitlow, M. (1988) *Science 242*, 423−426.
Bizub, D., Weber, I. T., Cameron, C. E., Leis, J. P., & Skalka, A. M. (1991) *J. Biol. Chem. 266*, 4951−4958.
Bonvin, A. M. J. J., Vis, H., Breg, J. N., Burgering, M. J. M., Boelens, R., & Kaptein, R. (1994) *J. Mol. Biol. 236*, 328−341.
Bowie, J. U., & Sauer, R. T. (1989a) *Biochemistry 28*, 7139−7143.
Bowie, J. U., & Sauer, R. T. (1989b) *J. Biol. Chem. 264*, 7596−7602.
Breg, J. N., van Opheusden, J. H. J., Burgering, M. J. M., Roelens, R., & Kaptein, R. (1990) *Nature 346*, 586−589.
Brown, B. M., & Sauer, R. T. (1993) *Biochemistry 32*, 1354−1363.
Brown, B. M., Bowie, J. U., & Sauer, R. T. (1990) *Biochemistry 29*, 11189−11195.
Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T., & DeGrado, W. F. (1995) *Science*, *270*, 935−941.
Cohn, E. J., & Edsall, J. T. (1943) *Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions*, Rheinhold, New York.
Creighton, T. E. (1984) *Proteins*, pp 136−137, W. H. Freeman and Co., New York.
Durchschlang, H. (1986) (Hinz, H.-J., Ed.) Chapter 3, p 45, Springer-Verlag, New York.
Hallewell, R. A., Laria, I., Tabrizi, A., Carlin, C., Getzoff, E. D., Tainer, J. A., Cousens, L. S., & Mullenbach, G. T. (1989) *J. Biol. Chem. 264*, 5260−5268.
Huston, J. S., Mudgett-Hunter, M., Tai, M.-S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R., & Opperman, H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5879−5883.
Huston, J. S., Mudgett-Hunter, M., Tai, M.-S., McCartney, J., Warren, F., Haber, E., & Opperman, H. (1991) *Methods Enzymol. 203*, 46−89.
Kim, S.-H., Kang, C.-H., Kim, R., Cho, J. M., Lee, Y.-B., & Lee, T.-K. (1989) *Protein Eng. 2*, 571−575.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Laemmli, U. K. (1970) *Nature 227*, 680−685.
Laue, T. M., Shah, B. D., Ridgeway, T. M., & Pelletier, S. L. (1992) in *Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S., Rowe, A., & Horton, J., Eds.) pp 90−125, Royal Society of Chemistry, Cambridge, U.K.
Liang, H., Sandberg, W. S., & Terwillinger, T. C. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7010−7014.
Mallender, W. D., & Voss, E. W., Jr. (1994) *J. Biol. Chem. 269*, 199−206.
Milla, M. E., & Sauer, R. T. (1994) *Biochemistry 33*, 1125−1133.
Milla, M. E., Brown, B. M., & Sauer, R. T. (1993) *Protein Sci. 2*, 2198−2205.
Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Pabo, C. O., & Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.
Pace, C. N., & Lauerents, D. V. (1989) *Biochemistry 28*, 2520−2525.
Percipalle, P., Simoncsits, A., Zakhariev, S., Guarnaccia, C., Sánchez, R., & Pongor, S. (1995) *EMBO J. 14*, 3200−3205.
Petsko, G., & Sligar, S. G. (1994) *Curr. Opin. Struct. Biol. 4*, 591−592.
Pomerantz, J. L., Sharp, P. A., & Pabo, C. O. (1995) *Science 267*, 93−96.
Predki, P. F., & Regan, L. (1995) *Biochemistry 34*, 9834−9839.
Ramachandran, G. N., & Sasisekharan, V. (1968) *Adv. Protein Chem. 23*, 283−437.
Raumann, B. E., Brown, B. M., & Sauer, R. T. (1994a) *Curr. Opin. Struct. Biol. 4*, 36−43.
Raumann, B. E., Rould, M. A., Pabo, C. O., & Sauer, R. T. (1994b) *Nature 367*, 754−757.
Rumbley, C. A., Denzin, L. K., Yantz, L., Tetin, S. Y., & Voss, E. W., Jr. (1993) *J. Biol. Chem. 268*, 13667−13674.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5469.
Schägger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368−379.
Schildbach, J. F., Milla, M. E., Jeffrey, P. D., Rauman, B. E., & Sauer, R. T. (1995) *Biochemistry 34*, 1405−1412.
Smith, T. L., & Sauer, R. T. (1995) *J. Mol. Biol. 249*, 729−742.
Spolar, R. S., & Record, M. T. J. (1994) *Science 263*, 777−784.
Susskind, M., & Youderian, P. (1983) in *Lambda II* (Hendriz, R. W., Roberts, J. W., Stahl, F. W., & Wiesberg, R. A., Eds.) pp 347−364, Cold Spring Harbor Press, Cold Spring Harbor.
Toth, M. J., & Schimmel, P. (1986) *J. Biol. Chem. 261*, 6643−6646.
Vershon, A. K., Bowie, J. U., Karplus, T. M., & Sauer, R. T. (1986) *Proteins Struct. Funct. Genet. 1*, 302−311.

BI9521194