Arc Repressor Is Tetrameric When Bound to Operator DNA†

Bronwen M. Brown, James U. Bowie,‡ and Robert T. Sauer*
Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
Received April 30, 1990; Revised Manuscript Received August 14, 1990

ABSTRACT: The Arc repressor of bacteriophage P22 is a member of a family of DNA-binding proteins that use N-terminal residues in a β-sheet conformation for operator recognition. Here, Arc is shown to bind to its operator site as a tetramer. When mixtures of Arc (53 residues) and an active variant of Arc (78 residues) are used in gel retardation experiments, five discrete protein–DNA complexes are observed. This result is as expected for operators bearing heterotetramers containing 4:0, 3:1, 2:2, 1:3, and 0:4 ratios of the two proteins. Direct measurements of binding stoichiometry support the conclusion that Arc binds to a single 21-base-pair operator site as a tetramer. The Arc–operator binding reaction is highly cooperative (Hill constant = 3.5) and involves at least two coupled equilibria. In the first reaction, two unfolded monomers interact to form a folded dimer (Bowie & Sauer, 1989a). Rapid dilution experiments indicate that the Arc dimer is the kinetically significant DNA-binding species and allow an estimate of the equilibrium dissociation constant for dimerization [$K_1 = 5\ (\pm 3) \times 10^{-9}$ M]. The rate of association of Arc–operator complexes shows the expected second-order dependence on the concentration of free Arc dimers, with $k_2 = 2.8\ (\pm 0.7) \times 10^{18}$ $M^{-2}\ s^{-1}$. The dissociation of Arc–operator complexes is a first-order process with $k_{-2} = 1.6\ (\pm 0.6) \times 10^{-4}$ s^{-1}. The ratio of these kinetic constants [$K_2 = 5.7\ (\pm 2.3) \times 10^{-23}$ M^2] provides an estimate for the equilibrium constant for dissociation of the DNA-bound tetramer to two free Arc dimers and the operator. An independent determination of this complex equilibrium constant [$K_2 = 7.8\ (\pm 4.8) \times 10^{-23}$ M^2] was obtained from equilibrium binding experiments.

The Arc repressor of bacteriophage P22 is a sequence-specific DNA-binding protein, which uses N-terminal residues in an extended β-conformation to bind to operator DNA (Vershon et al., 1986, 1987a; Breg et al., 1989, 1990; Knight & Sauer, 1989; Zagorski et al., 1989). Both the 21-base-pair *arc* operator sequence and the pattern of Arc–DNA contacts show approximate 2-fold symmetry, suggesting that the 53-residue protein binds to its operator as an oligomer, with different subunits making contacts with each operator half-site (Vershon et al., 1987a, 1989). At micromolar concentrations and above, Arc is dimeric in solution (Vershon et al., 1985). At nanomolar concentrations, Arc dissociates in a concerted reaction to unfolded monomers (Bowie & Sauer, 1989a).

In this paper, we show that Arc is tetrameric when bound to a single operator DNA site. Using mixtures of wild-type Arc and an active variant that contains 25 additional residues, we observe five distinct operator complexes in gel electrophoresis experiments. This result suggests that Arc binds its operator as a tetramer, a conclusion supported by direct measurements of the stoichiometry of binding. We also find that the equilibrium binding reaction is highly cooperative, with a Hill constant of 3.5. Kinetic experiments indicate that Arc is a mixture of monomers and dimers at the concentrations where operator binding is observed and show that dimers are the kinetically significant species in forming the DNA-bound tetramer.

Materials and Methods

Buffers. Loading buffer contains 50% glycerol, 0.02% bromophenol blue, and 0.02% xylene cyanol. Binding buffer I contains 10 mM Tris (pH 6.5), 3 mM $MgCl_2$, 0.1 mM EDTA, 200 mM KCl, 100 μg/mL bovine serum albumin (BSA), and 0.02% Nonidet P-40 (NP40). Binding buffer II contains 10 mM Tris (pH 7.5), 3 mM $MgCl_2$, 0.1 mM EDTA, 100 mM KCl, 100 μg/mL BSA, and 0.02% NP40. The BSA and NP40 are important components of both binding buffers as they prevent Arc from sticking to glass and plastic surfaces at low protein concentrations. An 8 M stock solution of GuHCl was obtained from Pierce Chemical Co. All other buffer components were reagent grade. Water was distilled and deionized with a Millipore MilliQ system.

Protein Purification. Wild-type Arc was expressed and purified as described by Vershon et al. (1986). The concentration of Arc, in monomer equivalents, was determined by using an extinction coefficient at 280 nm of 6756 M^{-1} cm^{-1}. This value was calculated as the sum of the extinction coefficients of tryptophan and tyrosine (Fasman, 1975), each of which occurs once in the Arc monomer. The extinction coefficient was not corrected for differences between the absorbance of native Arc and denatured Arc, but this difference is less than 3%. The Arc-*lt*1 variant was purified as described by Bowie and Sauer (1989b). The *lt*1 mutation is a frame shift that changes the C-terminal residue of Arc from Ala to Arg and adds 25 extra amino acids to the protein (Bowie & Sauer, 1989b). As the additional residues in Arc-*lt*1 do not include Tyr or Trp, the extinction coefficient of this protein was also assumed to be 6756 M^{-1} cm^{-1}. Unless noted, all protein concentrations are calculated in terms of mole equivalents of the Arc monomer per liter.

DNA Fragments. A 27-base-pair operator was synthesized for use in many of the binding studies. DNA was synthesized with an Applied Biosystems 381A DNA synthesizer and purified by FPLC chromatography using an anion-exchange Mono-Q column. The duplex operator was generated by mixing complementary oligonucleotides in equimolar amounts, heating to 65 °C, and cooling to room temperature. This operator fragment contains the 21-base wild-type *arc* operator sequence (in bold, with the central base marked) with a few flanking nucleotides:

†Supported by NIH Grant AI-16892.
‡Present address: Department of Chemistry and Biochemistry, UCLA, 405 Hilgard Ave., Los Angeles, CA 90024.

0006-2960/90/0429-11189$02.50/0

```
                    |
5′-ATCGATGATAGAAGCACTCTACTATCG-3′
3′-TAGCTACTATCTTCGTGAGATGATAGC-5′
                    |
```

A variant of this fragment, missing four bases at each 3′-end, was also synthesized to allow convenient end-labeling. For some binding studies, a 50-base-pair *Eco*RI/*Hin*dIII restriction fragment containing the AO110 *arc* operator was purified from plasmid pAO110 and used (Vershon et al., 1989). The AO110 operator is symmetric around the central base pair and has an affinity for Arc within 2-fold of the wild-type operator (Vershon et al., 1989). For assays of nonspecific DNA binding, three different DNA fragments were used. These include the 750-base-pair *Eco*RI/*Pst*I fragment from pAO110, a synthetic 27-base-pair fragment corresponding to the λ repressor operator, and a 50-base-pair restriction fragment bearing the Mnt operator. With all three fragments, half-maximal DNA binding was observed at Arc concentrations from 0.2 to 0.8 μM.

DNA was end-labeled by filling in the 5′-overhangs with α-^{32}P-labeled nucleoside triphosphates and the Klenow fragment of DNA polymerase I (Vershon et al., 1987b). The specific activity of end-labeled DNA was determined by measuring the radioactivity incorporated into a known quantity of DNA according to standard procedures (Maniatis et al., 1977).

Gel Retardation Assays. Gel retardation assays and associated methods were performed essentially as described by Vershon et al. (1987a,b). For equilibrium and kinetic assays using labeled operators, DNA was present at concentrations from 4 to 12 pM. Arc protein was incubated with labeled DNA in either microtiter plates or polypropylene microcentrifuge tubes for a given period of time; a $^1/_{10}$ to $^1/_{30}$ volume of loading buffer was added to each sample, and aliquots containing 500–800 cpm were loaded onto a nondenaturing 0.5× TBE 7% polyacrylamide gel [acrylamide:bisacrylamide (29:1)]. Electrophoresis was performed at 250 V and continued until the bromophenol blue dye reached the bottom of the gel. Gels were exposed to Kodak XAR5 film using intensifying screens at −70 °C. Films were scanned with an LKB Bromma 2202 ultroscan laser densitometer equipped with a Hewlett-Packard 3390A integrator. Three different scans were performed on each lane, and the values were averaged. The fraction of operator bound (θ) was calculated as the area of the bound fraction divided by the sum of the areas of the bound and free fractions. Within each experiment, the sum of the bound and free areas was generally constant to within 20%, and similar values of fractional saturation were obtained by the method described above, from the decrease in the intensity of the free band, or from the increase in intensity of the bound band.

For gel assays of binding stoichiometry, unlabeled operator DNA was used at tenth micromolar concentrations. In this case, samples were electrophoresed as described and gels were stained for 15 min in 0.5 μg/mL ethidium bromide. The DNA was visualized by using a Fotodyne shortwave ultraviolet source and photographed with Polaroid 665 positive/negative film; the negative was analyzed by densitometry as described for the autoradiographs. The band corresponding to the protein–DNA complex did not fluoresce as intensely as the free DNA band, presumably because bound Arc inhibits ethidium bromide intercalation and/or partially quenches the fluorescence. The relative increase in the intensity of the bound band did, however, correspond to the decrease in intensity of the free band. For these data, θ was calculated as the intensity of the bound band at a given Arc concentration divided by the intensity of this band at saturating concentrations of Arc.

Mixed Oligomer Studies. Binding buffer II was used for DNA and protein solutions in initial mixed oligomer experiments. Binding buffer I was used in later experiments because it was found to improve the resolution of the five retarded species in the gel retardation assay. Arc and Arc-*lt*1 were diluted to 50 μg/mL, mixed together in different ratios and incubated at room temperature for 15 min, and then serially diluted as desired. Control experiments showed that coincubating the proteins at room temperature was as effective a mixing protocol as denaturing both proteins by heating to 65 °C for 2 min and then allowing the mixture to renature after cooling. End-labeled DNA was diluted to 10^5 cpm/mL, and 30 μL of the DNA solution was mixed with 30 μL of the protein mixture and incubated at room temperature for 1 h. A 2-μL aliquot of loading buffer was added to each sample, and samples were electrophoresed as described above.

Binding Studies. Binding buffer II was used for all DNA and protein solutions, and experiments were performed at room temperature (18–22 °C). In all cases, binding buffer II was preequilibrated at room temperature prior to use. Equilibrium binding studies were performed with wild-type protein, which was serially diluted in buffer II in microcentrifuge tubes and then allowed to dissociate for 15 min. One-tenth volume of the labeled 27-base-pair operator was added to give an operator concentration of ≈5 pM, and the mixture was incubated for 2 h in microtiter plates. Binding was assayed by the gel retardation assay as described above.

Stoichiometric binding of Arc to operator DNA was studied with protein and DNA concentrations of 0.25 μM or greater. Increasing amounts of Arc, diluted in binding buffer II, were added to a constant amount of the 27-base-pair operator and incubated for 2 h. The samples were electrophoresed and visualized with ethidium bromide as described above.

Dissociation Rates. In previous studies, Vershon et al. (1987a) showed that Arc dissociated from the wild-type operator contained on a 240-base-pair DNA restriction fragment with a half-life of ≈80 min, under conditions similar to those used here. We repeated these experiments using the 50-base-pair AO110 operator fragment to test whether the length of the DNA fragment significantly affected the half-life. Arc at a concentration of 20 nM in binding buffer II was mixed with $^1/_{10}$ volume of labeled operator DNA and allowed to equilibrate at room temperature. Dissociation reactions were begun by adding 20-μL aliquots of the protein–DNA mixture to 20 μL of a solution containing 2 μM unlabeled operator DNA. The reactions were started at various times to allow all samples to be loaded for the gel retardation assay at the same time. When Arc was mixed with the unlabeled DNA prior to addition of the labeled DNA, no binding to the labeled DNA was observed. The rate constant, k_{-2}, for dissociation of the Arc–AO110 operator complex was found to be 1.6 (±0.6) × 10^{-4} s^{-1}, corresponding to a half-life of approximately 72 min.

Association Rates. For measurements of association rates, binding buffer II, preequilibrated at room temperature, was used for DNA and protein dilutions, and the resulting solutions were kept at room temperature. Association reactions were initiated by adding the AO110 operator fragment with gentle mixing to a tube containing wild-type Arc. The volume of the added DNA was only $^1/_{10}$ the final volume to minimize effects due to further dilution and dissociation of Arc dimers. The operator concentration in the reaction mixture was ≈5 pM. Portions were removed at various times and added to an equal

volume of quench solution containing unlabeled operator DNA at a concentration of 80 nM. Control experiments showed that the unlabeled operator prevented any further association of Arc with the labeled operator. Samples were loaded onto gels for assay within 5 min of addition of the quench solution.

Association rate experiments in the presence of GuHCl were performed as described above with the exception of the dilution protocols. In one experiment, wild-type Arc at a concentration of 200 nM in 4 M GuHCl was diluted 100-fold into binding buffer II, and labeled operator DNA was added within 10–15 s. In a second experiment, wild-type Arc at a concentration of 200 nM in binding buffer II was diluted 100-fold into binding buffer II containing 40 mM GuHCl, and again the incubation with DNA was begun within 10–15 s. In a third experiment, Arc was diluted to a concentration of 2 nM in binding buffer II plus 40 mM GuHCl and allowed to equilibrate for 10–15 min before addition of the labeled operator DNA.

Calculations. If the binding reaction of Arc to its operator is written as

$$4U + O \leftrightharpoons 2A_2 + O \leftrightharpoons A_4O \quad (1)$$

where U is the unfolded Arc monomer, O is operator DNA, A_2 is the folded Arc dimer, and A_4O is the bound complex, then the relevant equilibrium and conservation expressions are

$$K_1 = [U]^2/[A_2] \quad (2)$$

$$K_2 = [A_2]^2[O]/[A_4O] \quad (3)$$

$$K_1^2K_2 = [U]^4[O]/[A_4O] \quad (4)$$

$$[A_{total}] = [U] + 2[A_2] + 4[A_4O] \approx [U] + 2[A_2] \quad (5)$$

Free tetramers of Arc are not included in eq 5 because they are not observed in solution (Vershon et al., 1985). Moreover, because Arc is always present in significant excess over total operator DNA in our experiments, the concentration of bound Arc can be ignored (the validity of this simplification was confirmed by showing that 3-fold changes in operator concentration did not affect the observed binding curves). Substituting $[U]^2/K_1$ for $[A_2]$ in eq 5 and solving the resulting quadratic expression gives

$$[U] = (K_1/4)(-1 + \sqrt{1 + 8[A_{total}]/K_1}) \quad (6)$$

In DNA-binding experiments, the fraction of operator bound is given by

$$\theta = \frac{[A_4O]}{[A_4O] + [O]} = \frac{1}{1 + K_1^2K_2/[U]^4} \quad (7)$$

Substituting for [U] in this expression from eq 6 allows θ to be calculated if K_1, K_2, and $[A_{total}]$ are known.

In most association rate experiments, data were used for only the first few minutes of binding. Because the protein–operator complex has a half-life of greater than 1 h under these conditions, the operator dissociation reaction can be ignored. Moreover, because Arc is present in excess over labeled operator, the total Arc concentration and the concentration of Arc dimers should remain effectively constant during the experiment. Under these conditions, the forward rate expression can be simplified to

$$\frac{d[A_4O]}{dt} = -\frac{d[O]}{dt} = k_2[A_2]^2[O] = k_{app}[O] \quad (8)$$

and the expected fraction of operator bound at time t will be given by

$$\theta = 1 - e^{-k_{app}t} \quad (9)$$

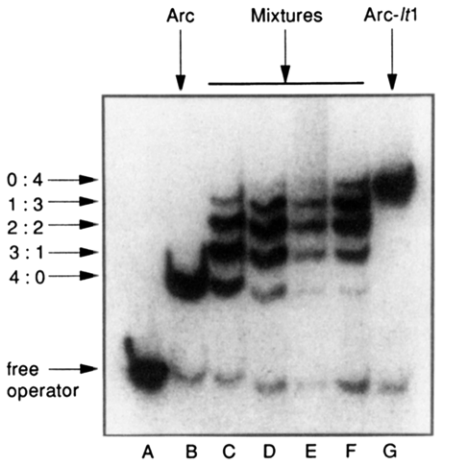

FIGURE 1: Gel retardation assays of operator DNA binding by Arc, Arc-*lt*1, and mixtures of the two proteins. Lane A shows operator DNA with no added protein. Lane B shows operator DNA incubated with wild-type Arc alone. Lanes C–F show operator DNA incubated with mixtures of Arc and Arc-*lt*1 at 2:1, 1:1, 2:3, and 1:1 molar ratios, respectively. Lane G shows Arc-*lt*1 alone incubated with operator DNA. The total protein concentration was 4 nM in each assay.

Results

A variant of Arc, called Arc-*lt*1, contains 25 additional C-terminal residues as a consequence of a frame-shift mutation (Bowie & Sauer, 1989b). This variant is fully active in vivo, and the additional residues do not affect the structure, stability, or operator affinity of Arc as measured by biochemical experiments in vitro. In gel retardation assays, operator complexes containing the 78-residue Arc-*lt*1 variant are retarded to a greater extent than operator complexes containing the 53-residue wild-type Arc protein (Bowie & Sauer, 1989b; cf. lanes B and G, Figure 1).

Binding of Mixed Oligomers to Operator DNA. To probe Arc's oligomeric form when bound to operator DNA, Arc and Arc-*lt*1 were mixed and used in gel retardation assays of operator binding. As shown in lanes C–F of Figure 1, these mixtures give rise to five differentially retarded protein–operator complexes. The slowest and fastest migrating complexes comigrate with the bands corresponding to operator complexes of Arc-*lt*1 and Arc, respectively. The remaining three bands have intermediate mobilities, suggesting that the corresponding protein–DNA complexes have different numbers of Arc and Arc-*lt*1 subunits. The simplest interpretation of these data is that Arc is tetrameric when bound to its operator and that the three intermediate bands are heterotetramers containing 3:1, 2:2, and 1:3 ratios of Arc to Arc-*lt*1 subunits, respectively. These protein–DNA complexes are specific, as Arc only begins to bind to nonoperator DNA fragments from 27 to 750 base pairs in length at concentrations 100–200-fold higher than those used in Figure 1 (data not shown).

To determine the stoichiometry of binding directly, assays were performed at high protein and DNA concentrations, where the equilibrium strongly favors formation of the protein–DNA complex. Figure 2 shows the result of one such experiment. The ratio of total Arc to bound operator can be calculated for each point before saturation of the operator is reached. The average value of the stoichiometry from two independent experiments was 3.8 (±0.8), supporting the idea that the five species seen in the mixed oligomer studies do represent tetramers of different Arc and Arc-*lt*1 composition.

In principle, a single Arc oligomer could bind more than one molecule of operator DNA. To test for this possibility, we performed separate and combined binding experiments with

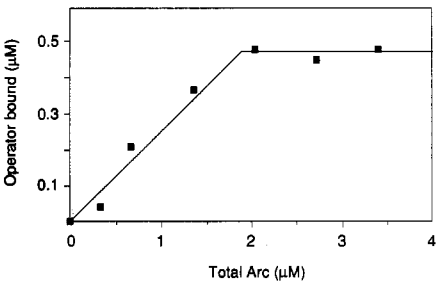

FIGURE 2: Stoichiometric binding of Arc to operator DNA. The 27-base-pair *arc* operator fragment was present at a concentration of 0.46 μM.

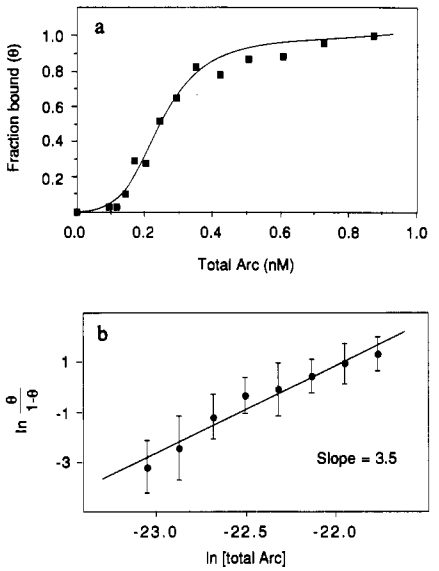

FIGURE 3: (a) Equilibrium binding of Arc to the 27-base-pair *arc* operator fragment. Half-maximal binding occurs at a total Arc concentration of 0.25 nM. The curve was calculated from eqs 6 and 7, with $K_1 = 5 \times 10^{-9}$ M and $K_2 = 1.1 \times 10^{-22}$ M^2. (b) Hill plot of equilibrium binding data from (a) and two independent experiments. The data points represent average values (±1 SD).

a 27-base-pair and a 50-base-pair operator fragment. The complexes of these operator fragments with Arc were clearly resolved in gel retardation experiments, but only the expected complexes and no complexes of intermediate mobility were observed when the two operators were present in the same binding experiment (data not shown). We conclude that the Arc tetramer is able to bind strongly to only one operator molecule at a time.

Equilibrium Binding. The binding of Arc to operator DNA is likely to involve several coupled equilibria. Although Arc is tetrameric when bound to operator, tetramers of Arc have not been observed in solution (Vershon et al., 1985). Moreover, the Arc dimer is known to dissociate to unfolded monomers in a concerted reaction at low concentrations (Bowie & Sauer, 1989a). The simplest scheme for operator binding that involves the known oligomeric species of Arc is

$$4U + O \overset{K_1}{\rightleftharpoons} 2A_2 + O \overset{K_2}{\rightleftharpoons} A_4O$$

We can test whether this model provides a good description of the actual binding reaction by studying how operator binding depends upon Arc concentration.

Figure 3a shows a typical binding experiment for equilibrium binding of Arc to operator DNA. Half-maximal operator binding occurs at an Arc concentration of 0.25 nM in the experiment shown in Figure 3a. In four independent experiments, the average Arc concentration required for half-maximal operator binding was found to be 0.22 (±0.04) nM. This value is about 10-fold lower than that previously reported (Vershon et al., 1987a, 1989), a difference that can be attributed to the use of NP40 in our assay buffer. The presence of this detergent prevents inactivation of Arc by surface adsorption at low protein concentrations. The equilibrium binding reaction is also highly cooperative, with binding increasing from 10% to 90% over a narrow range of Arc concentration. The Hill coefficient, a measure of the cooperativity of the reaction, is 3.5 (see Figure 3b). The model of eq 1 would predict a Hill constant of 4 if monomers were the only populated species at the 0.1–1 nM concentrations where operator binding is observed. A value of 3.5 could indicate that Arc is a mixture of monomers and dimers at these concentrations.

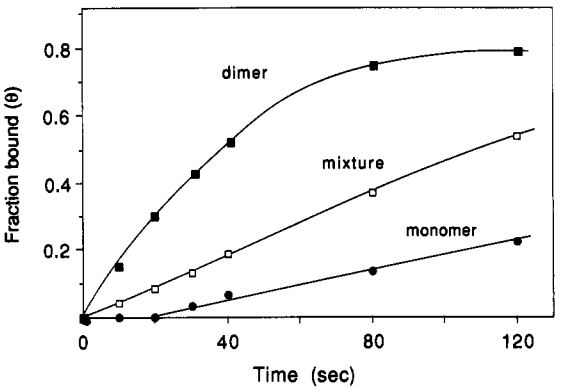

FIGURE 4: Rates of operator binding following different dilution protocols. (■) Rapid dilution of Arc from 200 nM; DNA added within 15 s. (□) Dilution of Arc from 200 nM; DNA added 15 min after the dilution. (●) Rapid dilution of Arc from 200 nM in 4 M GuHCl; DNA added within 15 s. The final Arc concentration in each experiment was 2 nM, and the final buffer composition was binding buffer II plus 40 mM GuHCl.

Kinetics of Arc–Operator Association. The model predicts that a solution of Arc dimers should show faster DNA binding than a solution containing an equilibrium mixture of monomers and dimers. This latter solution, in turn, should show faster binding than one consisting solely of monomers. To test this prediction, we performed binding experiments in which the protein solution was treated in three different ways to obtain the desired solution species. To obtain a solution containing mostly dimers, we began with Arc at a concentration of 200 nM, where it is predominantly dimeric (Vershon et al., 1985), and then rapidly diluted the protein 100-fold and added operator DNA within 10–15 s. To obtain a solution consisting initially of monomers, we denatured Arc in 4 M GuHCl (Bowie & Sauer, 1989a) and then rapidly diluted the protein 100-fold and added operator DNA within 10–15 s. To obtain an equilibrium population of monomers and dimers, we allowed Arc to equilibrate at a concentration of 2 nM and then added operator DNA. In all three cases, the final buffer composition following dilution was the same (binding buffer II plus 40 mM GuHCl) and the final Arc concentration was 2 nM. As shown in Figure 4, the DNA-binding reaction is fastest for the Arc sample that should be largely dimeric, slower for the sample that should contain an equilibrium mixture of monomers and dimers, and slowest for the sample that should be predominantly monomeric. The final level of binding obtained after long incubations of all three samples was identical (data not shown). These data provide support for the idea that Arc exists as an equilibrium mixture of monomers and dimers in the nanomolar concentration range.

If dimers of Arc are the only kinetically significant binding species (as in eq 1), then the rate at which protein–DNA

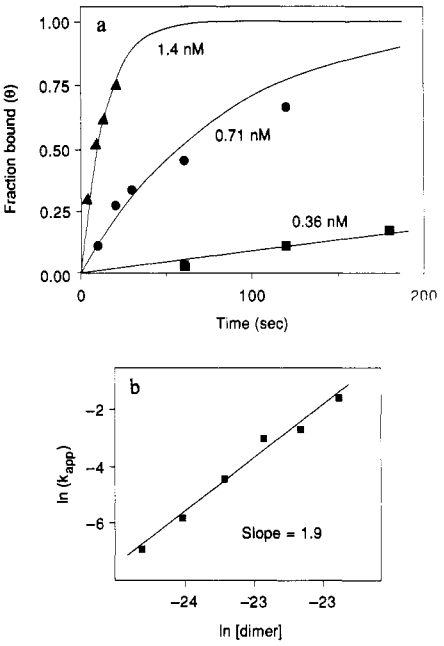

FIGURE 5: (a) Kinetics of operator binding at different total concentrations of Arc. The curves were calculated from eq 9 with values of k_{app} of 0.001 s^{-1} (0.36 nM), 0.012 s^{-1} (0.71 nM), and 0.068 s^{-1} (1.4 nM). (b) Dependence of the apparent association rate on the concentration of the free Arc dimer. The line is a least-squares fit of the data.

complexes form should depend on the square of the free dimer concentration (eq 8). Under conditions similar to those used in our binding experiments (20 °C, pH 7.5, 100 mM KCl), an estimate of the dimerization constant, K_1 = 5 nM, has been obtained from extrapolation of GuHCl denaturation experiments to zero denaturant concentration [see Figure 4 in Bowie and Sauer (1989a)]. Using this value to calculate the dimer concentration in the equilibrated mixture and assuming that the rapid dilution experiment generates a population of 100% dimers, we would expect an 8.4-fold difference in rate in the two experiments. If the rapid dilution experiment contained only 50% dimers, for example, because some of the dimers dissociate during the dilution protocol, then the rate difference would be 2.1-fold. Preliminary experiments indicate that the half-life of the Arc dimer is on the order of 15–60 s, and thus some dissociation would be expected. In the experiment shown in Figure 4, the initial rate of association for the "dimer" is 3.6 times faster than that observed for the equilibrium mixture of monomers and dimers. Therefore, these data are within the range predicted for a 5 nM dissociation constant and support the idea that dimers are the kinetically significant binding species. Using the observed rate difference to calculate the equilibrium constant for dimerization, we get a value of K_1 = 1.7 nM assuming 100% dimers and a value of 8.4 nM assuming 50% dimers in the rapid dilution experiment. We assume then that $K_1 = 5\ (\pm 3) \times 10^{-9}$ M and use the 5 nM value for the calculations described below.

To further test the basic model of eq 1, we measured the rate of operator association at different Arc concentrations. Figure 5a shows several of these experiments. Figure 5b shows that the apparent association rates obtained by these experiments exhibit a near second-order dependence upon the free concentration of Arc dimers (calculated assuming K_1 = 5 nM). This is expected from the model, since

$$k_{app} = k_2[A_2]^2$$

(see eq 8) where k_2 is the third-order rate constant for formation of the Arc–operator complex from two Arc dimers. An average value of $k_2 = 2.8\ (\pm 0.7) \times 10^{18}\ M^{-2}\ s^{-1}$ was calculated from the experiments shown in Figure 5a and three additional experiments performed at total Arc concentrations of 0.5, 1, and 2 nM. The rate constant for dissociation of the Arc–operator complex is $k_{-2} = 1.6\ (\pm 0.6) \times 10^{-4}\ s^{-1}$ (see Materials and Methods). From these values, we can calculate the equilibrium constant, $K_2 = 5.7\ (\pm 2.3) \times 10^{-23}\ M^2$, for formation of the Arc tetramer–operator complex from the operator and two Arc dimers. An independent estimate of $K_2 = 7.8\ (\pm 4.8) \times 10^{-23}\ M^2$ can be obtained from the equilibrium binding data by using eqs 6 and 7. The values of K_2 calculated from the kinetic data and the equilibrium data agree within experimental error, indicating that a single set of constants can reasonably fit both the equilibrium and kinetic data. This consistency provides additional evidence that the model of eq 1 is a good description of the Arc–operator binding reaction.

Discussion

Several lines of evidence indicate that Arc repressor binds to its operator as a tetramer. Mixtures of Arc and Arc-*lt*1 give rise to five distinct protein–operator DNA complexes in gel retardation experiments. This result is expected if Arc binds DNA as a tetramer, with the five bands representing an independent assortment of Arc and Arc-*lt*1 monomers. Although the same result might be expected if Arc bound as an octamer with the five bands representing mixtures of Arc and Arc-*lt*1 dimers, this possibility can be ruled out since binding experiments at high concentrations indicate that no more than four Arc monomers are needed to bind a single operator molecule. The mixed oligomer method for determining the binding state of a DNA-binding protein was first used by Hope and Struhl (1987) to show that GCN4, a yeast transcriptional activator, binds DNA as a dimer. Sorger and Nelson (1989) have recently used the same approach to demonstrate that the yeast heat-shock transcription factor is trimeric when bound to DNA.

Why are tetramers of Arc needed to bind the 21-base-pair operator site? On sizing columns, Arc migrates at an apparent molecular mass of 13 kDa, close to the 12.4-kDa value expected for a globular dimer (Vershon et al., 1985). A globular dimer of this size would be expected to have a diameter of ≈31 Å. Vershon et al. (1987a) have shown that Arc makes contacts with its operator site that span 21 base pairs or approximately 68 Å of B-form DNA. Hence, in the absence of an extraordinary conformational change, a single Arc dimer could only contact about half of the operator, and two dimers would be required to make all of the observed DNA contacts. In fact, recent model-building experiments based upon an NMR structure of the Arc dimer (Breg et al., 1990) indicate that two Arc dimers are needed to contact the full operator site.

There are two obvious pathways by which the complex of the operator and the Arc tetramer could be assembled. In principle, the tetramer could form in solution and then bind to the operator:

$$2A_2 + O \rightleftharpoons A_4 + O \rightleftharpoons A_4O \qquad (10)$$

Although tetramers are not stable in solution, a transiently formed species might be trapped by operator binding. The second possibility is that dimers bind to half-sites sequentially:

$$2A_2 + O \rightleftharpoons A_2O + A_2 \rightleftharpoons A_4O \qquad (11)$$

In this case, the binding of the first dimer would be weak, and cooperative interactions between the DNA-bound dimers would be needed to stabilize the complex. Under our standard

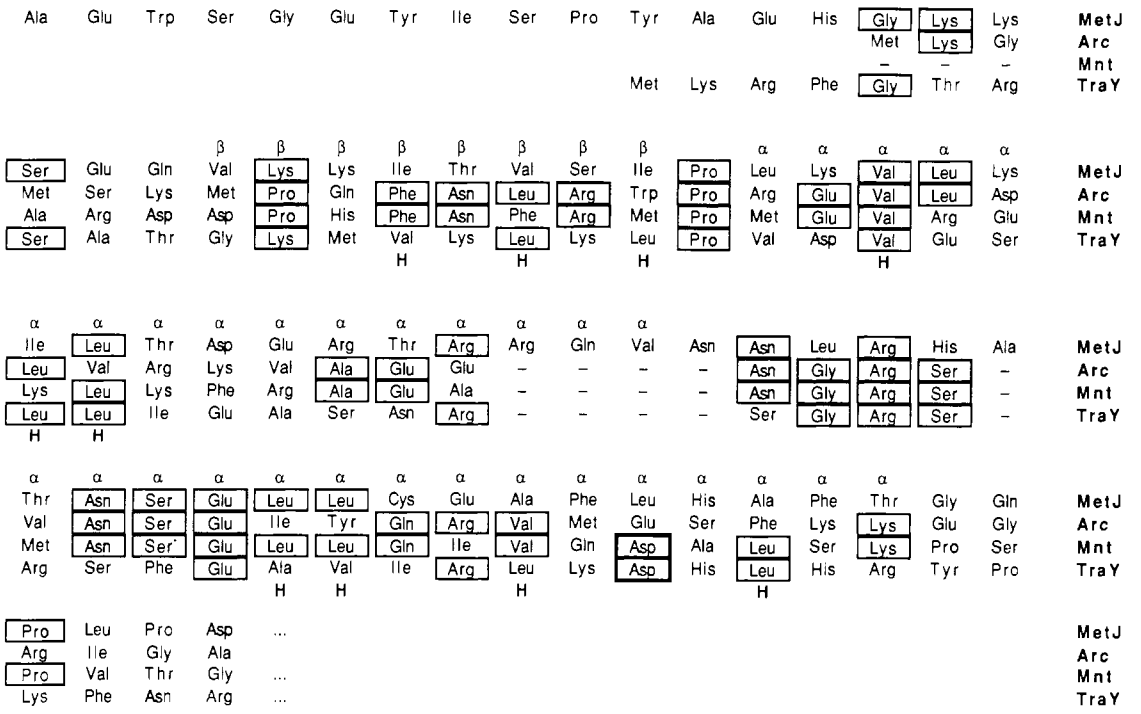

FIGURE 6: Sequence alignment of the MetJ, Arc, Mnt, and TraY proteins. Regions of β-structure (β) and α-helix (α) in MetJ (Rafferty et al., 1989) are marked. In Arc, residues 8–14 are in an extended β-conformation, while residues 16–29 and 35–49 are α-helical (Breg et al., 1989; Zagorski et al., 1989). Boxed residues indicate sequence identities. H indicates conserved hydrophobic positions. Sequences begin with their natural N-termini. The MetJ, Mnt, and TraY proteins have additional C-terminal sequences that are not included. Original sequence references and some pairwise alignments can be found in Sauer et al. (1983), Bowie and Sauer (1990), Fowler et al. (1983), Inamoto et al. (1988), Lahue and Matson (1990), and Breg et al. (1990).

gel shift conditions, we observe only free operator or operators with bound tetramers. This does not rule out the possibility of the A_2O intermediate but does indicate that it must be poorly populated and unstable relative to the O and A_4O species.

By either of the models discussed above, tetrameric contacts between Arc dimers play an important role in operator binding. Vershon et al. (1986) have identified a number of Arc mutations that cause large decreases in operator binding affinity. A subset of these mutations was found not to alter the stability or structure of the Arc dimer and was therefore thought to affect side chains that mediated direct contacts with the DNA. While this interpretation is reasonable, our results suggest that mutations of this type could also affect side chains that mediate contacts between DNA-bound dimers.

Arc is closely related to the Mnt repressor of phage P22 (Sauer et al., 1983) and more distantly related to the TraY proteins of F and related episomes (Bowie & Sauer, 1990) and the *E. coli* MetJ repressor (Breg et al., 1990), a protein whose crystal structure is known (Rafferty et al., 1989). Figure 6 shows an alignment of the sequences of these proteins. Mnt, like Arc, binds operator DNA as a tetramer (Vershon et al., 1987b). However, unlike Arc, Mnt forms a stable tetramer in solution and binds to its operator in a noncooperative fashion (Vershon et al., 1985, 1987a). Knight and Sauer (1989) have shown that a hybrid protein containing 9 N-terminal residues of Arc and 76 C-terminal residues of Mnt binds specifically to the *arc* operator, and in fact binds more tightly to the *arc* operator than does wild-type Arc. The specific binding of the hybrid to the *arc* operator indicates that the tetrameric Mnt core can appropriately position the Arc DNA-binding residues for specific binding and supports the idea that the overall structures of the DNA-bound Arc and Mnt tetramers must be similar. The tighter binding of the hybrid to the operator is consistent with the idea that some of the DNA-binding energy of Arc is lost because the active binding species is not stable at the concentrations where operator binding occurs.

ACKNOWLEDGMENTS

We thank Robin Kelley, Mike Mossing, John Reidhaar-Olson, and Kevin Shoemaker for advice, help, and comments on the manuscript.

Registry No. *arc* operator, 130246-70-7.

REFERENCES

Bowie, J. U., & Sauer, R. T. (1989a) *Biochemistry 28*, 7139–7143.

Bowie, J. U., & Sauer, R. T. (1989b) *J. Biol. Chem. 264*, 7596–7602.

Bowie, J. U., & Sauer, R. T. (1990) *J. Mol. Biol. 211*, 5–6.

Breg, J. N., Boelens, R., George, A. V. E., & Kaptein, R. (1989) *Biochemistry 28*, 9826–9833.

Breg, J. N., van Opheusden, J. H. J., Burgering, M. J. M., Boelens, R., & Kaptein, R. (1990) *Nature 346*, 586–589.

Fasman, G. D. (1975) *Handbook of Biochemistry and Molecular Biology, Vol. I, Proteins*, CRC Press, Cleveland, OH.

Fowler, T., Taylor, L., & Thompson, R. (1983) *Gene 26*, 79–89.

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

Inamoto, S., Yoshioka, Y., & Ohtsubo, E. (1988) *J. Bacteriol. 170*, 2749–2757.

Knight, K. L., & Sauer, R. T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 797–801.

Lahue, E. E., & Matson, S. W. (1990) *J. Bacteriol. 172*, 1385–1391.

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

Laboratory Press, Cold Spring Harbor, NY.
Rafferty, J. B., Somers, W. S., Saint-Girons, I., & Phillips, S. E. V. (1989) *Nature 341*, 705–710.
Saint-Girons, I., Duchange, N., Cohen, G. N., & Zakin, M. M. (1984) *J. Biol. Chem. 259*, 14282–14285.
Sauer, R. T., Krovatin, W., DeAnda, J., Youderian, P., & Susskind, M. (1983) *J. Mol. Biol. 168*, 699–713.
Sorger, P. K., & Nelson, H. C. M. (1989) *Cell 59*, 807–813.
Vershon, A. K., Youderian, P., Susskind, M. M., & Sauer, R. T. (1985) *J. Biol. Chem. 260*, 12124–12129.
Vershon, A. K., Bowie, J. U., Karplus, T. M., & Sauer, R. T. (1986) *Proteins: Struct., Funct., Genet. 1*, 302–311.
Vershon, A. K., Liao, S.-M., McClure, W. R., & Sauer, R. T. (1987a) *J. Mol. Biol. 195*, 323–331.
Vershon, A. K., Liao, S.-M., McClure, W. R., & Sauer, R. T. (1987b) *J. Mol. Biol. 195*, 311–322.
Vershon, A. K., Kelley, R. D., & Sauer, R. T. (1989) *J. Biol. Chem. 264*, 3267–3273.
Zagorski, M. G., Bowie, J. U., Vershon, A. K., Sauer, R. T., & Patel, D. J. (1989) *Biochemistry 28*, 9813–9825.

Nucleotide Sequence of the Gene for the b Subunit of Human Factor XIII[†,‡]

Ralph E. Bottenus, Akitada Ichinose, and Earl W. Davie*
Department of Biochemistry, University of Washington, Seattle, Washington 98195
Received May 22, 1990; Revised Manuscript Received August 10, 1990

ABSTRACT: Factor XIII (M_r 320 000) is a blood coagulation factor that stabilizes and strengthens the fibrin clot. It circulates in blood as a tetramer composed of two a subunits (M_r 75 000 each) and two b subunits (M_r 80 000 each). The b subunit consists of 641 amino acids and includes 10 tandem repeats of 60 amino acids known as GP-I structures, short consensus repeats (SCR), or sushi domains. In the present study, the human gene for the b subunit has been isolated from three different genomic libraries prepared in λ phage. Fifteen independent phage with inserts coding for the entire gene were isolated and characterized by restriction mapping, Southern blotting, and DNA sequencing. The gene was found to be 28 kilobases in length and consisted of 12 exons (I–XII) separated by 11 intervening sequences. The leader sequence was encoded by exon I, while the carbonyl-terminal region of the protein was encoded by exon XII. Exons II–XI each coded for a single sushi domain, suggesting that the gene evolved through exon shuffling and duplication. The 12 exons in the gene ranged in size from 64 to 222 base pairs, while the introns ranged in size from 87 to 9970 nucleotides and made up 92% of the gene. The introns contained four Alu repetitive sequences, one each in introns A, E, I, and J. A fifth Alu repeat was present in the flanking 3′ end of the gene. Two partial *Kpn*I repeats were also found in the introns, including one in intron I and one in intron J. The *Kpn*I repeat in intron J was 89% homologous to a sequence of approximately 2200 nucleotides flanking the gene coding for human β globin and approximately 3800 nucleotides from the L1 insertion present in the gene for human factor VIII. Intron H also contained an "O" family repeat, while two potential regions for Z-DNA were identified within introns G and J. One nucleotide change was found in the coding region of the gene when its sequence was compared to that of the cDNA. This difference, however, did not result in a change in the amino acid sequence of the protein.

Factor XIII is a zymogen that circulates in blood as a tetramer (a_2b_2) and is converted to an active transglutaminase by thrombin in the presence of calcium and fibrin [for reviews, see Lorand et al. (1980) and Folk (1983)]. The two pairs of a and b subunits separate during the activation reaction, giving rise to two activated peptides and factor XIIIa, an enzyme composed of two a′ subunits. Factor XIIIa catalyzes the formation of ε-(γ-glutamyl)lysine bonds between fibrin monomers as well as a number of other plasma proteins. The function of the b subunits is not clear, but it is likely that they protect and stabilize the plasma zymogen prior to activation (Folk & Finlayson, 1977). It may also play a role in the regulation of the contact activation pathway (Halkier & Magnusson, 1988).

Plasma levels of factor XIII have been determined by Yorifuji et al. (1988), who reported that all of the a subunits are in the tetrameric form, which is present in plasma at a concentration of 11 μg/mL. An excess of the b subunits also circulates in plasma in a monomeric form at a concentration of 10 μg/mL. Electron microscopy has shown that the tetrameric form of factor XIII appears as two globular a subunits surrounded by two flexible, rod-shaped b domains (Carrell et al., 1989).

The primary structure of both the a and b subunits has been determined by a combination of cDNA cloning and amino acid sequence analysis in our laboratory (Ichinose et al., 1986a,b) and those of others (Grundmann et al., 1986; Takahashi et al., 1986). The gene for the a subunit is over 160 kb in length (Ichinose & Davie, 1988) and has been localized to chromosome 6 at p24–25 (Board et al., 1988). The b subunit is 641 amino acids in length and is made up of 10 repeating domains (approximately 60 amino acids each) that have been called GP-I domains (Davie et al., 1986), short consensus repeats, or sushi domains (Ichinose et al., 1990). The term GP-I

[†]This work was supported in part by a research grant (HL 16919 to E.W.D.) and a National Research Service Award (HL 07843 to R.E.B.) from the National Institutes of Health.
[‡]The nucleotide sequence in this paper has been submitted to GenBank under Accession Number J05294.
*Author to whom correspondence should be addressed.

0006-2960/90/0429-11195$02.50/0