Mechanism of DNA Binding to the Gene 5 Protein of Bacteriophage fd†

Gary D. Brayer‡ and Alexander McPherson*

ABSTRACT: A model for the bimolecular complex arising from the interaction of single-stranded DNA with the gene 5 DNA binding protein (G5BP) of bacteriophage fd is proposed on the basis of difference Fourier analyses and the correlation between structural and physicochemical data. The essential DNA binding element is the G5BP dimer which provides two antiparallel DNA binding channels, each constructed from amino acid contributions of both monomers within the pair. These channels display identical bonding environments but opposite polarities as a consequence of the inherent dyad symmetry within each dimer unit. The two channels are separated by 30 Å, and each is 10 Å wide and 40 Å long. We propose that DNA binding is a consequence of two general sets of interactions. Aromatic side chains of G5BP stack upon nucleic acid bases, and the DNA phosphate backbone is bound by a series of appropriately positioned lysyl and arginyl side chains. The DNA conformation is fully extended, and each binding channel can accommodate up to five nucleotides. Only minor conformational changes in the native G5BP structure are required to optimize the binding of DNA. The G5BP–DNA complexation model presented here serves to explain some of the mechanistic features associated with the role this protein plays in the formation of a nucleoprotein helix during the bacteriophage fd life cycle.

The interaction of proteins with nucleic acids plays a central role in genetic expression and regulation. An understanding of this process at the atomic level must ultimately be achieved in order to describe the structural and functional basis for the underlying mechanisms. Direct visualization of such interactions can, at this time, only be obtained by single-crystal X-ray diffraction analyses of the proteins and nucleic acids involved and of complexes between these macromolecules. From such studies, it is possible to deduce, or visualize directly, the atomic structures and contacts responsible for the forces determining complex formation or maintenance. One system that provides us such an opportunity is the crystalline form of the gene 5 DNA binding protein (G5BP) from bacteriophage fd.

The properties of the gene 5 DNA binding protein have been extensively investigated in solution by a number of laboratories using a broad range of physical and chemical techniques [for reviews, see Denhardt (1975), Ray (1977), or Coleman & Oakley (1980)]. The amino acid sequence has been determined (Cuyper et al., 1974; Nakashima et al., 1974) and is comprised of 87 amino acids in a single polypeptide chain of 9700 molecular weight. No disulfide bridges are present. This relatively small protein exists predominantly as a dimer in solution, unless nucleic acid material is present (Pretorius et al., 1975; Cavalieri et al., 1976). Attention has been focused on the interactions between G5BP and various nucleic acids and nucleic acid fragments. Several investigations have identified a number of amino acid residues essential to the binding of nucleic acids and have provided insight into the conformational state of DNA in such complexes. Chemical modification studies have shown three of five tyrosines to be involved in DNA binding (Pretorius et al., 1975; Anderson et al., 1975), as well as one phenylalanine (Coleman et al., 1976; Garssen et al., 1980; Lica & Ray, 1977). Lysine and arginine residues have also been shown to participate in liganding DNA (Anderson et al., 1975; Alma-Zeestraten, 1982).

The stoichiometry of G5BP–DNA binding has also been investigated. The number of nucleotides bound per protein monomer has been variously reported to be five (Pratt et al., 1974; Gray, C. W., et al., 1982; Cavalieri et al., 1976; Torbet et al., 1981), four (Cavalieri et al., 1976; Day, 1973; Alberts et al., 1972; Oey & Knippers, 1972), and three (Anderson et al., 1975; Gray, D. M., et al., 1982). The binding affinity of G5BP for bacteriophage fd DNA has been estimated to be of the order of 10^9 M (Coleman & Oakley, 1980). This protein has a high affinity for single-stranded DNA of any base sequence of composition. Bound DNA appears to be completely extended and unstacked regardless of its original conformational state (Alberts et al., 1972; Day, 1973). A substantial cooperative effect is associated with the complexation of longer polynucleotides. That is, once a single G5BP protein dimer is bound to a DNA strand, binding of the next dimer unit immediately adjacent to the first is enhanced an estimated 60-fold (Cavalieri et al., 1976; Alberts et al., 1972).

Although its primary physiological role is to switch on synthesis of single-stranded DNA daughter virions and then stabilize as well as protect these strands, G5BP is also found to be a potent helix-destabilizing protein. That is, the T_m's of a variety of double-stranded DNAs are lowered by approximately 40 °C in the presence of G5BP, irrespective of nucleotide composition (Alberts et al., 1972). The melting or unwinding activity arises from a strong preference for binding single-stranded DNA over its double-stranded counterpart. In this way, G5BP disturbs the natural equilibrium between native and single-stranded regions of duplex DNA, leading to the unwinding of the nucleic acid. Thus, this protein belongs to the class of helix-destabilizing proteins which includes the calf thymus unwinding protein (Herrick & Alberts, 1976), the gene 32 protein of T_4 (Alberts et al., 1968), and bovine pancreatic ribonuclease (Felsenfeld et al., 1963).

To provide a foundation for interpretation of the accumulated physicochemical data as well as further insight into the nature of protein–nucleic acid interactions, the high-resolution structure refinement of the gene 5 protein has been completed (Brayer & McPherson, 1983). The molecular structure has been refined to a standard residual of $R = 0.217$ by using restrained parameter least-squares techniques at 2.3-Å resolution. The 87 amino acids of this protein were found to be primarily in the β-conformation, organized as a three-stranded

† From the Department of Biochemistry, University of California, Riverside, California 92521. *Received June 28, 1983.* This research was supported by Grant GH-27838-04 from the U.S. Public Health Service and by the Medical Research Council of Canada.

‡ Present address: Department of Biochemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5.

0006-2960/84/0423-0340$01.50/0

sheet and a two-stranded ribbon with a broad connecting loop. Two G5BP monomers are tightly interlocked about an intermolecular dyad axis to form a compact dimer unit. Dimer linkage is manifested primarily by the formation of a large intermolecular, six-stranded, antiparallel β barrel. The dimer is characterized by two symmetry-related antiparallel clefts that traverse its surface in a plane essentially perpendicular to the dyad axis. Resident within these clefts are the side chains of those residues implicated in nucleic acid complexation by other techniques. Also protruding from each G5BP monomer and adjacent to each binding cleft is an elongated wing of polypeptide chain which we believe to be intimately involved in nucleic acid binding.

Presented here is a description of that region of the G5BP molecule that we feel evidence indicates is directly in contact with nucleic acid, and also presented is an analysis of the atomic contacts likely to be formed. In addition to the physical and chemical data, we have relied on the high-resolution refined structural analysis of G5BP and the results of several difference Fourier experiments involving the diffusion of oligonucleotides into native G5BP crystals.

Experimental Procedures

Three approaches were taken toward elucidating the details of DNA binding to G5BP. Two of these attempted the direct visualization of bound DNA through crystallographic techniques. The first relied on the introduction of various deoxyoligonucleotides directly into the mother liquor of native G5BP crystals of the type originally used in the structure determination (McPherson et al., 1976). Variables examined included oligonucleotide length, sequence, concentration, and contact time. Diffraction data from four of the most promising experiments [$d(pA)_4$, d(CTTC), $d(CG)_3$, d(CCG)] were collected to 2.8-Å resolution and processed as described by Brayer & McPherson (1983). Difference Fourier methods were employed to localize bound nucleotides by using the phases calculated at the end of native G5BP structure refinement.

A second approach toward visualizing the bound DNA was cocrystallizing deoxyoligonucleotides with G5BP. Unfortunately, all such attempts have resulted either in native G5BP crystals or in crystal forms having 12 G5BP molecules per asymmetric unit (McPherson et al., 1982). Each of these multiple-copy G5BP crystals represents a difficult crystallographic problem. Although a structure determination using molecular replacement is in progress on one such form, even with the knowledge of the G5BP dimer structure, considerable additional time and effort will be required before binding information from this source becomes available.

In view of the problems associated with direct crystallographic visualization of the G5BP–nucleic acid complexes, we have also attempted to model this interaction. The direction and constraints for this approach were grounded in the extensive physicochemical literature that has accumulated concerning complexation, and the structural basis was provided by the refined coordinates for native G5BP.

Modeling involved the use of a Kendrew model (2 cm/Å) constructed according to the refined coordinates of G5BP obtained by restrained-parameter least-squares techniques (Hendrickson & Konnert, 1980; Brayer & McPherson, 1983). In addition, a portion of the dyad-related G5BP monomer of the dimer unit was constructed. This requirement arose because amino acid side groups from both G5BP molecules of the dimer are present in each of the two DNA binding clefts traversing the dimer face. A model of five covalently linked DNA bases was also assembled from Kendrew parts and then

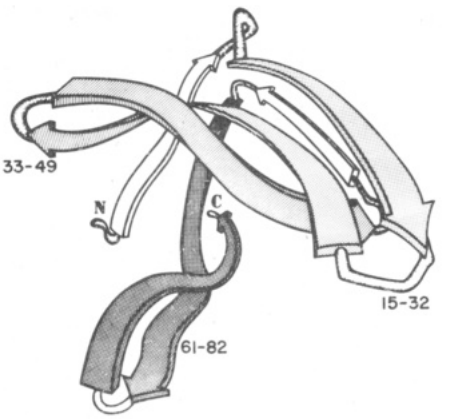

FIGURE 1: Stylized drawing of the polypeptide chain backbone of G5BP illustrating the three β loops, the dotted strands being the DNA binding loop (residues 15–32), the striped loop being the complex loop (residues 33–49), and the checkered loop being the dyad loop (residues 61–82). The formation of a centrally located three-stranded β sheet by the DNA binding and complex loops is evident. The amino and carboxyl termini are positioned at the rear in this view.

systematically fitted into one of the G5BP binding clefts.

Following nucleic acid positioning, atomic coordinates were measured by using the plumb-line method. These coordinates were then geometrically idealized (starting root mean square from ideality, 0.20 Å) by using a program provided by Dr. R. Wing of University of California at Riverside. Subsequent comparison of idealized coordinates against measured values indicated no significant readjustments had occurred during this procedure. Similarly, adjustments in polypeptide chain conformation to accommodate the positioning of bound DNA and to optimize favorable interactions were modeled and measured. Stereochemical idealization of measured polypeptide coordinates was then accomplished by using the method of Hendrickson & Konnert (1980). Inspection of the idealized nucleic acid and protein models indicated that these remained spatially compatible.

The resultant protein–nucleic acid complex was displayed and examined on a Tektronics 4010 graphics terminal and plotted through the use of a CalComp plotter. Additional examinations were conducted on an Evans and Sutherland graphics system (University of California at San Diego) and with the color graphics system at the National Institutes of Health.

Results and Discussion

Structure of the G5BP Dimer. The three-dimensional structure analysis of G5BP by crystallographic methods (Brayer & McPherson, 1983) established the polypeptide chain conformation of this protein. The least-squares refined atomic coordinates resulting from that study represent the starting point of the present analysis of interactions likely to occur in G5BP–nucleic acid complexes.

The G5BP molecule is essentially composed of three major antiparallel β loops which project outward from a common hydrophobic core (see Figure 1). These have been designated as the "DNA binding loop" (residues 15–32), the "complex loop" (residues 33–49), and the "dyad loop" (residues 61–82). Two of these β loops (DNA binding and complex loops) are arranged such that a three-stranded antiparallel β sheet is formed across the mid-portion of the molecule. This sheet, in conjunction with a short eight-residue strand toward the back of the molecule, forms a partial β barrel that encloses the hydrophobic core of G5BP. If the molecule is positioned such that the three β-loop extremities point collectively upward

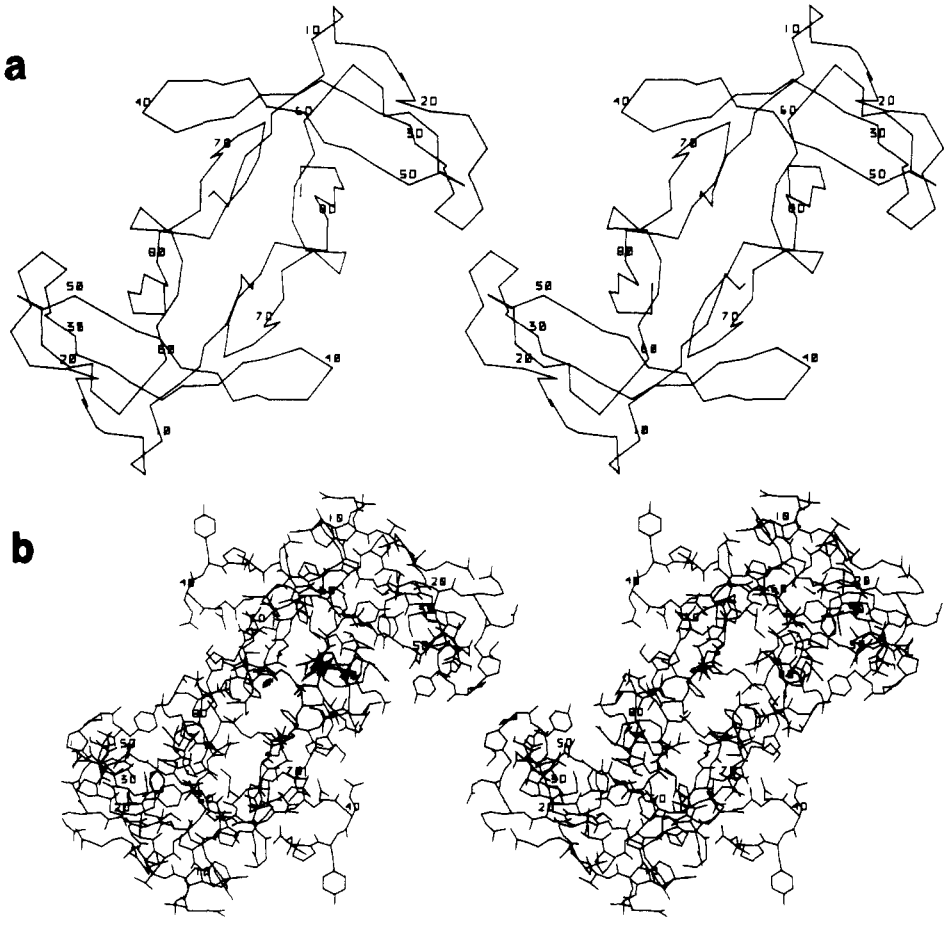

FIGURE 2: Stereo drawing of (a) α-carbon atoms only and (b) all the non-hydrogen atoms of the G5BP dimer structure. The orientation of these drawings is similar to that of Figure 1. The 2-fold axis relating G5BP monomers is into the plane of this illustration, and its location is evident from a comparison of structural elements common to each monomer. In both drawings, the α-carbon of every fifth amino acid is labeled according to its position in the amino acid sequence (Cuyper et al., 1974).

to form a tripod arrangement, the N and C termini of the polypeptide chain are located at the extreme back, in close juxtaposition.

The major species of G5BP found in solution is the dimer unit (Pretorius et al., 1975; Rached & Pohl, 1974; Cavalieri et al., 1976; O'Connor & Coleman, 1983), which does not self-associate into larger complexes unless single-stranded nucleic acid is present. In the native crystal, two G5BP molecules are closely associated about a perfect dyad axis which is coincident with a 2-fold symmetry element of the crystallographic unit cell. A high degree of intermolecular bonding and complementarity between dyad-related subunits is evident (see Figure 2), resulting in the creation of a new structural element. It consists of a compact and well-delineated six-stranded β cage or barrel, formed from the two extended dyad β loops and the amino-terminal strands of two monomers. This quaternary structure is apparent in the α-carbon drawing of Figure 2a and the drawing of all nonhydrogen atoms of the G5BP dimer in Figure 2b. Of interest is the array of hydrophobic side chains that coalesce within this structure so that the hydrophobic cores of the G5BP monomers become contiguous across the entire G5BP dimer. The dimer β barrel is further fortified by an extensive surface bonding network involving a number of different kinds of side- and main-chain interactions (Brayer & McPherson, 1983).

It is clear that dimer formation substantially alters the surface of G5BP from that exposed by the isolated monomer structure. The outward appearance of one portion of the polypeptide chain, however, remains unperturbed by dimer association, and this encompasses residues 20–30 which form the extremity of the DNA binding loop. A high degree of mobility or disorder was found for these residues in the structure analysis of the protein. This, we feel, is reasonable given the manner in which the loop projects from the surface of the G5BP dimer. In the crystalline state, this loop is fixed only by virtue of its contact with another independent molecule in the crystal, one unit cell away.

The DNA binding loop and the manner in which it dramatically projects away from the bulk of the molecule are evident in Figure 3, which shows the profile of a dimer unit. In this view, the dimer 2-fold axis is aligned vertically and in the plane of the illustration rather than perpendicular to it as in Figure 2. The last interaction of this loop with the remainder of the molecule that might secure its position is a rather long salt bridge between Arg-80 and Glu-30. Aside from this tie, however, there is little to restrain the mobility of this extended β loop, which incidently carries Tyr-26 on its tip, a residue strongly implicated in direct interaction with bound DNA (Anderson et al., 1975; O'Connor & Coleman, 1983).

Also seen in Figure 3 is a smaller projection of atoms provided by the dyad loops of the monomers (downward in this view) which occurs at the 2-fold interface and is thus centrally located between the two extended β loops. This projection, in conjunction with the extended β loops, forms two shallow channels which are oriented almost perpendicular to the plane of Figure 3. These depressions appear to be the binding sites for single-stranded DNA. Virtually all of the amino acid side chains implicated in DNA binding by other physicochemical means are located here.

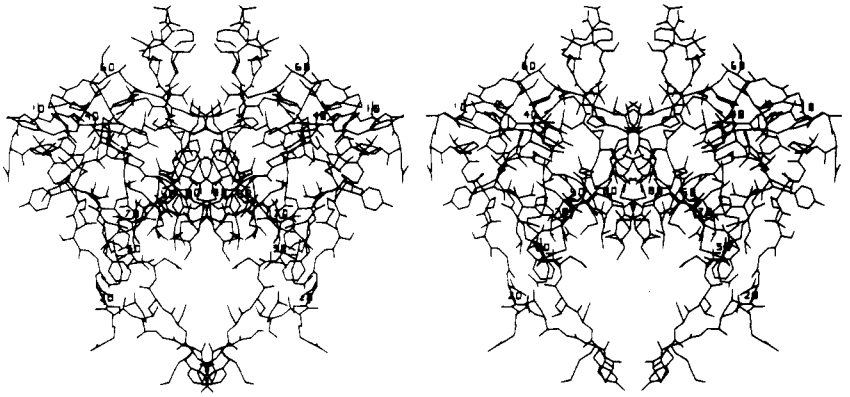

FIGURE 3: Stereo drawing of all the non-hydrogen atoms of the G5BP dimer with the view perpendicular to that of Figure 2. The 2-fold axis of the dimer unit runs in the plane of the illustration as is evident from comparison of the two extended β loops. These extended β loops consist of residues 20–30 and comprise the outer extremity of the DNA binding clefts. The protrusion of atoms at the monomer–monomer interface results from the juncture of dyad β loops. The two channels formed between the extended DNA binding loops and the juncture of dyad loops run approximately perpendicular to the 2-fold dimer axis and are the binding sites for single-stranded DNA.

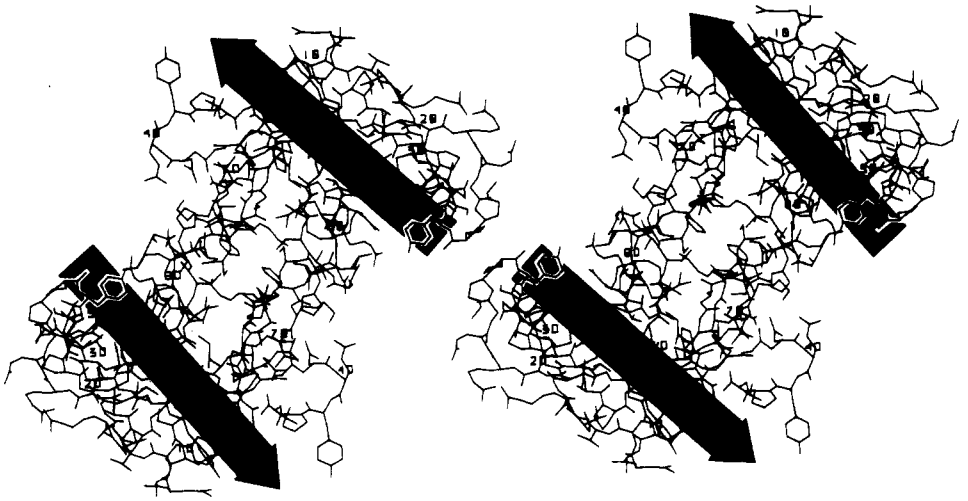

FIGURE 4: Stereo drawing of the G5BP dimer in which the positions of the two DNA binding channels have been schematically illustrated by black arrows. The view presented here is that of Figure 2. Note that each binding channel has an identical chemical composition as a result of the dimer 2-fold axis. The polarity or sense of molecular contacts in each is reversed from the other by the operation.

Each DNA binding channel is approximately 10 Å in width and 35 Å in length, consistent with what would be necessary to accommodate a fully extended nucleic acid chain of four or five nucleotides in length. Figure 4 shows schematically the approximate positions of these channels and their relationship to other features of the dimer unit. The view presented is the same as that of Figure 2a,b. Essentially, these channels traverse a path on either side of the molecular dyad and are defined by the extended DNA binding loops and the ridge of atoms formed at the dyad interface (also see Figure 3). The environment within each binding cleft is identical as a result of the dimer 2-fold axis, but the polarity of molecular contacts in each is reversed from the other by the same operation.

Nucleic Acid Binding Data from Difference Fourier Methods. Difference Fourier maps of four putative nucleic acid complexes formed by diffusing short [$d(pA)_4$, d(pCTTC), $d(pCG)_3$, d(pCCG)] oligonucleotides into native G5BP crystals were examined. In each case, the resultant difference electron density maps were disappointing in that they could not be interpreted on the basis of the oligomer structure diffused into each crystal. However, it was of some support to the DNA binding model proposed below that the major positive electron density peaks of the maps did appear in two locations within each DNA binding channel. These occurred at the ends of the two arrows in Figure 4. Specifically, the probable binding occurs near Tyr-34 and Tyr-26, that is, near the upper and lower ends, respectively, of the binding clefts. Both of these residues have been implicated as playing essential roles in G5BP–DNA complexation by other physicochemical techniques (Anderson et al., 1975; Hilbers et al., 1978; Garssen et al., 1980; Alma et al., 1981; O'Connor & Coleman, 1983). In both cases, the peaks were significantly above the general noise level of the difference maps and were common to almost all of the difference Fourier experiments. There were, however, other peaks in the map that approached these in magnitude, but they were in general not common to several difference maps.

Examination of crystal packing down the y axis of the native crystal indicates the probable cause for our inability to bind deoxyoligomers to G5BP in an ordered way. Two G5BP dimers as they are positioned along this axis are shown in Figure 5a,b. In this illustration, the crystallographic y axis is coincident with the 2-fold axis of the G5BP dimers. The view presented is rotated by 90° to that of Figure 3. Comparison with Figure 4 shows that the DNA binding channels in Figure 5 run almost directly into the plane of the illustration. These are positioned just inside the two outer extended β loops and separated by the ridge of atoms at the molecular 2-fold axis. In Figure 5, the adjacent dimer molecule packed along the crystallographic y axis is positioned so that its amino and carboxyl termini fit directly into the two DNA binding channels of the original molecule. Thus, the greater part of both binding channels is occluded, making it physically impossible to bind contiguous strands of oligonucleotides across

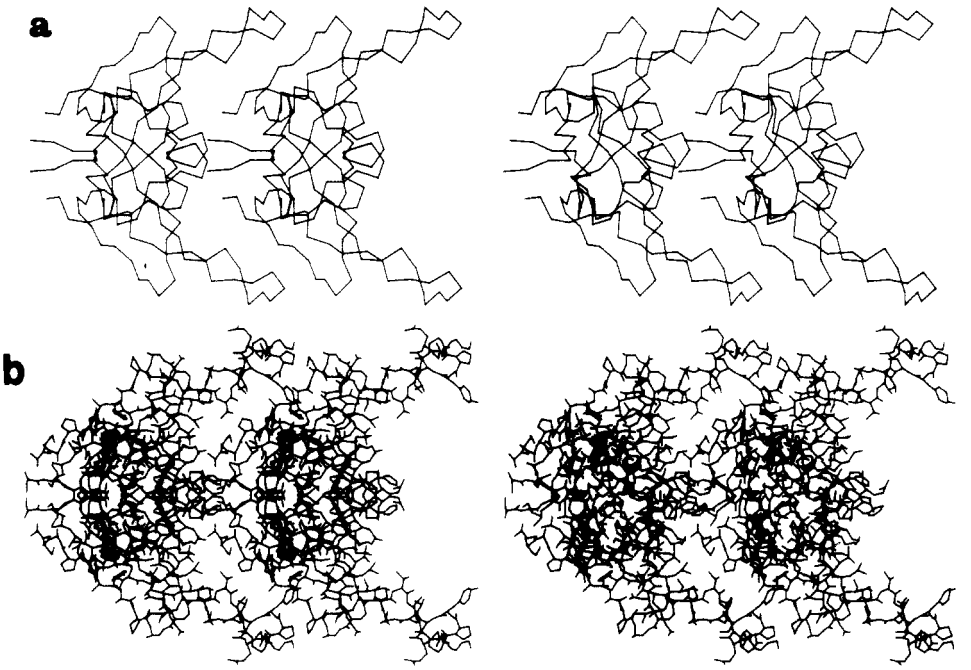

FIGURE 5: Drawing of (a) only α-carbon backbone atoms and (b) all non-hydrogen atoms of two G5BP dimers as packed along the y axis in the crystal. The crystallographic 2-fold symmetry element is coincident with the intermolecular dyad bisecting each dimer unit. The primary lattice interactions along the y direction involve the amino and carboxyl termini of one dimer being positioned directly in the DNA binding channels of the adjacent G5BP dimer. The salt bridges and hydrogen-bonding network thus formed apparently preclude the specific binding of oligonucleotides in native G5BP crystals. Binding is possible at the extremities of each binding channel as indicated by difference Fourier electron density maps.

these regions. Further inspection shows that the termini of the translationally adjacent dimer form salt bridges to charged residues in the binding channels that otherwise might bind the phosphate backbone of DNA. Hydrogen bonding between other residues further restricts access to these channels. This leaves only portions of the outer extremities of each binding channel accessible, and it is here that there is evidence of nucleotide binding in our difference Fouriers.

It is our feeling that the difference electron density at either end of the DNA binding clefts in fact represents disordered oligonucleotides that are weakly or nonspecifically bound so that the resultant difference electron density is diffuse. We believe that in spite of the fact that the difference electron density is not observed to inhabit the entire binding cleft, due to lattice interactions, it does at least provide a rudimentary definition of the end points of a bound strand. Since the DNA strand is linear, it marks the course that the nucleic acid must take.

It is also of note in Figure 5 that the two extended DNA binding loops (residues 15–32) of the G5BP dimers are in contact with translationally adjacent molecules in native G5BP crystals. Given their tenuous ties to the main body of G5BP (Brayer & McPherson, 1983), this, no doubt, influences or determines their conformation in the crystalline state. In solution, unfettered by lattice constraints, it is likely that at least the outer portion of this loop would have highly dynamic properties and assume other forms.

Complex Modeling: Constraints of Physical and Structural Studies. Given the extensive degree of monomer–monomer overlap and the apparent rigidity of the intermolecular hydrophobic barrel, it seems unlikely that DNA binding would substantially alter the dimer protein conformation. This is supported by the absence of evidence for conformational change upon complex formation as monitored by circular dichroism–optical rotatory dispersion (CD–ORD) spectroscopy, NMR, and chemical modification techniques (Day, 1973; Hilbers et al., 1978; Coleman et al., 1976). Probable exceptions to this, however, are the extended wing or DNA binding loop (residues 20–30) and the tip of the complex loop (residues 39–43), both of which are unconstrained by dimer interactions (Figures 1, 2, and 4). This does not rule out movements of interacting side chains or minor readjustments of short segments of the polypeptide chain. It does, however, support our contention that a DNA chain can be fitted to the crystallographically determined G5BP structure without distortion or substantial change in conformation, and with a reasonable expectation that the interactions suggested do, in fact, occur.

It might be anticipated simply from our knowledge of the binding sites of enzymes which interact with extended linear substrates that the DNA binding site of G5BP would be a groove, cleft, or extended depression in the surface of the protein. As described above, in the G5BP dimer there are such well-defined channels capable of simultaneously binding two single strands of DNA. For the gene 5 protein, however, there are additional constraints on the characteristics of the DNA binding clefts. The gene 5 protein exists as a dimer, and each subunit binds to opposite strands of duplex DNA or the opposing sides of a covalently closed single-stranded circular DNA. The polarity in space of the two DNA binding sites, therefore, must be opposite, binding a DNA strand running 3′ to 5′ on one monomer and 5′ to 3′ on the other. This reversal of binding sense is accomplished by the intermolecular 2-fold axis of symmetry and is, in fact, the only means by which this can be achieved and still maintain the two binding sites otherwise identical. This requires, however, that the binding paths for the DNA lie in a plane more or less perpendicular to the direction of the intermolecular 2-fold axis; otherwise, they could not bind contiguously along DNA single strands without producing severe distortions at the interfaces between dimers. This further implies that the two dimer DNA binding clefts must be essentially parallel across the dimer face and not cross over into the region of the 2-fold related binding cleft.

We would further expect that the dyad-related, and

therefore parallel, binding sites would separate the two single strands of DNA by a distance greater than that found in the DNA double helix, since the protein is known to produce complete physical separation of the two polynucleotide chains when complexation occurs. At the same time, it cannot separate the two strands, and therefore, the binding sites cannot in general be further apart than the maximal distance between strands anticipated from electron micrographs of the helical gene 5 DNA complex, i.e., not >45 Å. The gene 5 protein binding site must also be relatively straight and without marked deviations from linearity since the polynucleotide chain has been shown to bind with little or no distortion from its natural extended conformation (Day, 1973; Anderson et al., 1975; Coleman et al., 1976). This is apparent also from the gross structure of the gene 5 DNA helical complex as characterized by electron microscopy (Gray, C. W., et al., 1982).

Given the extended nature of bound DNA, one would expect nucleotide binding sites along the length of each binding channel at approximately 7.0-Å intervals. Further, since each channel could bind as many as five deoxynucleotides (Pratt et al., 1974; Gray, C. W., et al., 1982; Cavalieri et al., 1976; Torbet et al., 1981), the expected length of such a cleft would be of the order of 35 Å. Each binding channel must be accessible to a DNA strand over its entire length since the nucleic acid bound physiologically by G5BP (bacteriophage fd DNA) is continuous and could not thread its way through a constricted opening.

The considerations discussed above were made at the level of the gross structure of the protein and the overall conformation of bound DNA. Reference to Figure 4 shows that at this level the binding channels of the G5BP dimer satisfy all the necessary requirements to allow complexation. Both, for example, are nearly linear and are freely accessible to solvent. The two are separated by 30 Å (center to center), and therefore, the two bound DNA chains would not interact with each other. Each channel is sufficiently wide (10 Å) and long (35 Å) to accommodate extended DNA of up to five bases. Also, the two channels are essentially parallel and lie in a plane nearly perpendicular to the dyad, allowing contiguous DNA to span the entire face of the G5BP dimer. Finally, there are no other grooves or channels on the gene 5 dimer other than the ones we propose which could accommodate linear DNA, as inspection of Figures 2–5 shows.

Given the compatibility of the G5BP binding channel structure to the conformational requirements of bound DNA, it was relatively straightforward to meld the two macromolecular elements into a protein–nucleic acid complex model. From a relatively coarse initial fitting, specific interactions consistent with the available data were brought into play to achieve a finer fit. We chose to use guanidine bases throughout since this purine moiety would impose the greatest number of spatial constraints during model building. Upon maximization of mutually compatible contacts, the protein model was then examined to determine if small conformational changes could enhance binding without significantly distorting the native structure of the protein. Minor polypeptide rearrangements were reconstructed in two regions leading to the model for the G5BP–DNA complex shown in Figure 6.

Details of Protein–Nucleic Acid Complexation. The orientation of the single-stranded DNA with respect to the surface of the gene 5 DNA binding protein is shown schematically in Figure 6. The general course of the nucleic acid chain is that suggested in the representation of Figure 4. That is, the two DNA strands are positioned just inside the extended DNA binding loops and held apart by a smaller protrusion of atoms

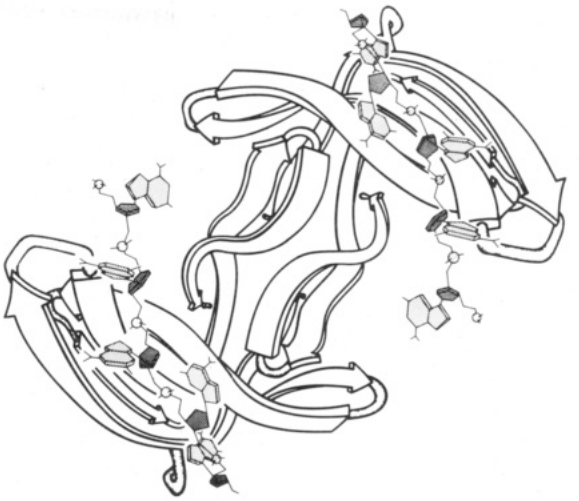

FIGURE 6: Stylized drawing of the G5BP dimer and the course of bound single-stranded DNA. Features of the bound chains are coded as follows: striped overlay, phosphate groups; cross-hatched overlay, sugar moieties; dotted overlay, base rings. Differentiation of polypeptide chains and the location of specific amino acids can be made by comparison with the stereo drawings of Figures 2 and 4, all of which view the G5BP dimer from a similar orientation. A total of 10 nucleotides is illustrated, 5 in each of the 2 bound DNA strands.

at the molecular 2-fold axis. Strand separation is approximately 30 Å, and the two DNA chains run essentially antiparallel to one another.

For the most part, each strand is positioned across the central three-stranded β sheet which forms the core of the G5BP molecule. Major interactions are also made with the complex and DNA binding loops. As Figure 6 shows, the β-bend tips of the dyad loops are also in the vicinity of the DNA strands. Note, however, that this latter interaction occurs within the 2-fold-related G5BP monomer binding channel as opposed to the monomer unit to which the dyad loop is a part. These dyad loop interactions are essential components in DNA complexation. Cross-linked binding channel interactions of this sort provide further evidence that the G5BP dimer is the smallest species capable of binding DNA.

It is apparent from Figure 6 that the phosphate backbone is bound closely to the surface of the protein throughout its course, whereas the base rings project somewhat further from the surface. The DNA conformation is almost fully extended and with bases unstacked. Even in this extended state, it was necessary to utilize five nucleotides to traverse the full length of each binding channel. The average phosphate to phosphate distance in the model is very nearly that expected for a fully extended polynucleotide. Thus, the structure is most consistent with those investigations which have found five bases bound per G5BP monomer (Pratt et al., 1974; Gray, C. W., et al., 1982; Cavalieri et al., 1976; Torbet et al., 1981).

The most prominent interactions between bound DNA and the G5BP are detailed in Figure 7. Only one binding channel is illustrated, since its 2-fold-related counterpart has exactly the same chemical and physical environment (Figure 2). The base of nucleotide 1 stacks coplanar with the aromatic side chain of Tyr-41, and the side chain of Tyr-34 stacks upon base 2. Lys-46 is positioned such that its free amino group is approximately equidistant from the phosphates of bases 1 and 2. Nucleotide 3 lies flat against the exposed side of the phenylalanyl side chain of residue 73. Note that this residue is a part of the polypeptide chain of the 2-fold-related G5BP monomer. The phosphate groups of bases 2 and 3 are located on either side of the side chain of Arg-16. Tyr-26, at the tip

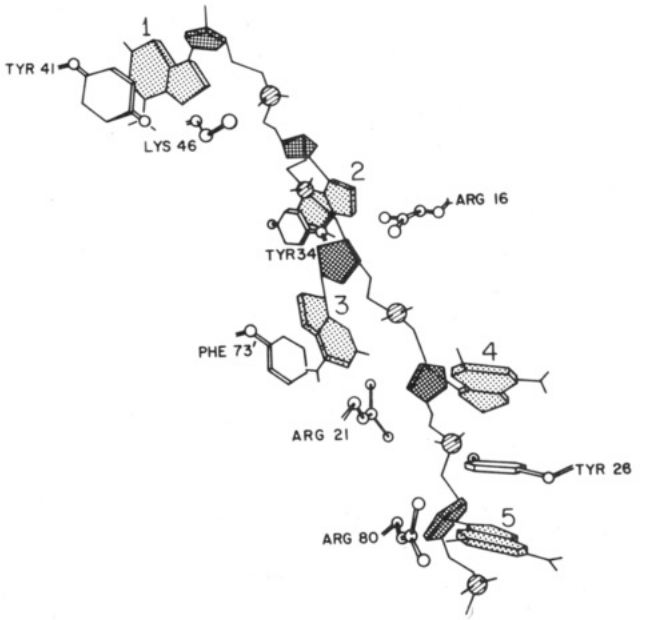

FIGURE 7: Schematic representation of one binding cleft on the G5BP dimer and the major interactions formed with DNA. The 2-fold symmetry-related binding channel exhibits an identical interactive environment. Two types of chemical interactions appear to be responsible for complexation. Aromatic side chains of the protein stack upon base rings of the DNA, and the phosphate backbone is neutralized by a combination of lysyl and arginyl side chains. The distribution of nucleotide binding sites is such that the DNA must be in a near fully extended state. To facilitate discussion in the accompanying text, each nucleotide has been arbitrarily numbered starting at the top of the drawing.

of the DNA binding loop, intercalates between base rings 4 and 5. The guanidinium group of Arg-21 lies near phosphate 3, and the same group of Arg-80 is positioned between phosphates 4 and 5. Besides these specific interactions, we observe that there are groupings of hydrophobic side chains clustered on the sides of base rings not stacked against aromatic groups and additional hydrophilic residues in the vicinities of bound ribose rings and phosphate groups. Two general mechanisms appear responsible for complexation. Base rings stack on protein aromatic side chains, and the bound phosphate backbone is fixed by appropriately positioned lysyl and arginyl side chains. A molecular model of a G5BP binding channel with bound DNA is illustrated in Figure 8.

Optimization of the binding interactions was effected by allowing minor conformational changes in the native high-resolution G5BP structure. The most extensive of these involved the tip of the DNA binding loop. In particular, the side chain of Tyr-26 was reoriented so that this residue pointed more directly between the base rings of nucleotides 4 and 5. This rearrangement also served to draw the DNA binding loop over the DNA strand. In doing so, arginines-16 and -21 assumed positions adjacent to the phosphates of nucleotides 2 and 3. These proposed structural changes are found in Figure 9, which illustrates the conformation of the DNA binding loop before and after complexation. The root mean square change in position of the 15 α-carbon atoms involved in the conformational change is 4.33 Å, with the greatest movement being that of glycine-23 which shifts 10.1 Å. The center of the aromatic ring of tyrosine-26 moves a total of 5.6 Å.

Movement of the DNA binding loop is easily accomplished since, for the most part, it is extended away from the core of the G5BP molecule (Figures 2–5) and therefore makes few contacts with the remainder of the protein. The polypeptide segment most affected involves residues 21–27. This β loop,

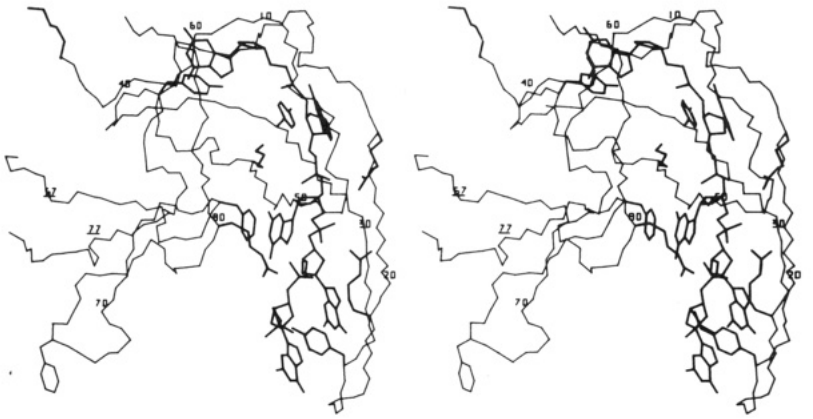

FIGURE 8: Stereo drawing of a DNA binding cleft with a bound oligomer of five bases in length. The polypeptide chain of the G5BP monomer which forms most of the binding channel surface is drawn. An additional binding element, the dyad loop of a 2-fold-related G5BP monomer (residue numbers underlined), is illustrated as well. Specific interactive groups discussed in the text are drawn with dark lines along with the bound DNA to accent their positions.

FIGURE 9: Stereo drawing of the G5BP binding channel in the vicinity of the DNA binding loop and in the region about the tip of the complex loop. Dashed lines indicate the native positions of polypeptide chains, while solid lines show the conformations of these same chains upon DNA complexation.

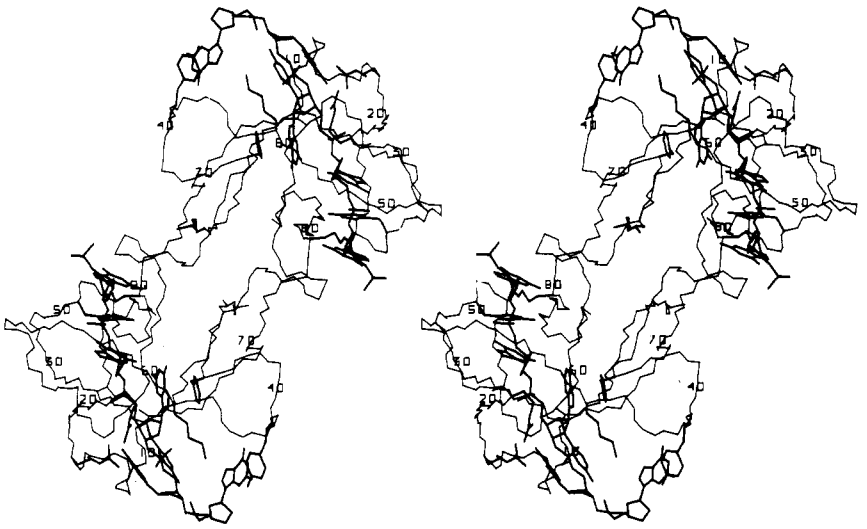

FIGURE 10: Stereo drawing of the polypeptide chain backbone of the G5BP dimer structure (light bonds) with two single strands of bound DNA (dark bonds). The protein conformation is that after modeling the DNA complex. For comparison, see the structure of the native G5BP dimer illustrated in Figure 2. Every tenth residue of both G5BP monomers has been numbered, and side chains interacting with DNA are drawn with bold dark lines.

and particularly residues 21–27, exhibits considerable mobility even in the crystalline state, where they are characterized by uniquely high temperature factors of nearly twice the mean for the remainder of the molecule (Brayer & McPherson, 1983). Given its positioning over the DNA binding channel, it is probable that the DNA binding loop makes initial contact with free nucleic acid strands and therefore may require the observed flexibility to sense and fix the DNA. In our model of the G5BP–DNA complex, the extensive contacts formed by this loop with the DNA would cause it to coalesce into a considerably more stable conformation.

The structure of G5BP, in the absence and presence of DNA, was examined in an effort to determine how the proposed conformational change in the DNA binding loop might be induced by polynucleotide binding. We suggest that the linchpin may be Arg-80. In the uncomplexed state, the guanidinium group of this residue lies midway between Asp-79 and Glu-30 (Figure 2). The bridge to Glu-30 represents the last, and essentially the only, interaction between the core of the G5BP molecule and the DNA binding loop before it freely extends into the surrounding solution. Our model building shows that when single-stranded DNA is inserted into the binding cleft, the phosphate groups of nucleotides 4 and 5 fall immediately above and below the guanidinium group of Arg-80 where they would vigorously compete for its attention. The resultant rearrangement of this side chain to accommodate the phosphate groups would disrupt the original Asp-79/Arg-80/Glu-30 bridge. Thus, when DNA enters the binding cleft, the last remaining constraint that may have been imposed on the DNA binding loop is removed, thereby initiating the conformational change (Figure 9).

A modest shift in the native position of Tyr-41 was also allowed to optimize the protein–nucleic acid fit. For the most part, rearrangement was restricted to residues 40–42 which occur at the tip of the complex loop and are relatively open to solvent (Figure 2). These conformational changes are also detailed in Figure 10. The high-resolution refinement of G5BP (Brayer & McPherson, 1983) has shown that these residues also have considerable mobility as is apparent from their above-average thermal motion parameters. The side chains of Lys-46 and Phe-73 were also repositioned as illustrated in Figures 9 and 10. We feel the conformational changes proposed are realistic, minimal, and consistent with expectations from solution studies (Day, 1973; Coleman et al., 1976; Hibers et al., 1975).

Correlation of Proposed Model with Data from Other Techniques. With regard to the extensive research of G5BP–DNA complexation using noncrystallographic methods, a number of correlations with the proposed model can be made. It was demonstrated by chemical modification and NMR experiments that three G5BP tyrosines occupy surface positions while two others are internal (Pretorius et al., 1975; Coleman et al., 1976; Alma et al., 1981). Upon complexation with DNA, there are marked changes in the CD tyrosyl bands (Day, 1973) and NMR spectra, which have been interpreted as indicating that all three surface tyrosines are involved in binding (Pretorius et al., 1975; Coleman et al., 1976; Garssen et al., 1977, 1978; Coleman & Armitage, 1978). Tyrosyl chemical shifts indicate that stacking of these aromatic groups on the DNA bases occurs (O'Connor & Coleman, 1983). Nuclear Overhauser effects imply that nucleotide sugar protons are near the ring protons of at least two of the aromatic residues for which resonances undergo substantial shifts during complex formation (Alma et al., 1981).

Chemical modification studies show three of five G5BP tyrosine residues can be reacted with either tetranitromethane or *N*-acetylimidazole (Anderson et al., 1975). DNA binding is lost upon derivatization, and conversely, the presence of DNA protects the side chains from modification. Subsequent peptide analysis has identified the modified tyrosine rings as belonging to residues 26, 41, and 56. These chemical modification results are consistent with the spectral evidence implying that three tyrosine residues play a pivotal role in DNA complexation.

Examination of the three-dimensional structure of G5BP (Figures 2 and 4) shows, in agreement with the data discussed above, that there are three surface and two internal tyrosine side chains. These are residues 26, 34, and 41 and residues 56 and 61, respectively. Furthermore, all three surface tyrosine residues are found in close proximity to the putative DNA binding channels (Figure 4). The structure of G5BP presented in Figure 2 is not, however, entirely in agreement with the assignment of surface tyrosines derived from chemical modification experiments (26, 41, 56; Anderson et al., 1975). In the three-dimensional structure, tyrosine-26 and tyrosine-41 are fully exposed and would be expected to undergo chemical

modification. Both occur at opposite extremes of the DNA binding cleft, and it is plausible that their reaction alone would be sufficient to explain the loss of DNA binding. In the G5BP structure, Tyr-56 is largely buried in the monomer hydrophobic core near Tyr-61 (Brayer & McPherson, 1983). Its edge, however, is exposed to solvent, and it is possible that it could be nitrated. It would not, however, have been our choice as the next most accessible tyrosine. Tyrosine-34, at the center of the DNA binding cleft, seems a more reasonable candidate for modification as it appears more accessible than tyrosine-56, though less so than either tyrosine-26 or tyrosine-41 (Figures 2, 4, and 8). The only explanation we can offer is that there may be an orientational or environmental effect that prevents nitration of tyrosine-34. Indeed, leucine-45 is stacked directly against the aromatic ring on the interior side, and phenylalanine-13 is also hovering nearby. In addition, it does not protrude overtly into the solvent but lies back flat against the monomer hydrophobic core.

In the model we present here (Figures 6–10), tyrosines-26, -34, and -41 are directly involved in the binding of DNA through stacking of their respective aromatic groups. Good agreement with experimental results is thus realized both in the number of tyrosine residues involved and in the mode of interaction involved. Judging from their central role in each DNA binding channel, modification of these residues would almost certainly lead to loss of DNA binding.

Correlation of this model with experimental data can also be made in the case of phenylalanine residues. Several lines of evidence imply that a phenylalanine side chain is intimately involved in DNA complexation. Ultraviolet irradiation of the complex of G5BP and bacteriophage fd DNA induces the formation of a covalent cross-link between protein and DNA (Lica & Ray, 1977). Peptide analysis revealed this cross-link occurred between residues 70 and 77 in the polypeptide chain of G5BP, with Phe-73 being the most likely candidate. Reference to Figures 2 and 5 shows Phe-73 is indeed located in the DNA binding channel. Of the three phenylalanines in the amino acid sequence, Phe-73 is the only residue having direct access to DNA complexation regions.

This phenylalanine was utilized in the model-building process to stack on the base ring of nucleotide 3 (Figures 7 and 8). It originates from the 2-fold-related monomer and extends across the breadth of the dimer face in assuming its position. NMR evidence (Coleman & Armitage, 1978; Hilbers et al., 1978; Garssen et al., 1980; Alma et al., 1981; O'Connor & Coleman, 1983) indicates that ring protons of a phenylalanine residue show large shifts upon DNA complexation. As with the G5BP tyrosyl residues, the magnitude and direction of this effect are consistent with aromatic ring stacking on bound DNA bases.

Other irradiation studies utilizing thymidine-containing oligonucleotides and G5BP have identified a cross-link formed to Cys-33 (Anderson et al., 1975b; Paradiso et al., 1982). Although inaccessible to the bulky Ellman's reagent, this residue reacts with the Hg^{2+} ion, and this leads to protein unfolding and loss of DNA binding (Anderson et al., 1975). The presence of fd DNA prevents the reaction of Cys-33 with mercurials. These studies imply that Cys-33 is in or near the protein–nucleic acid interface. Reference to Figure 2 shows this to be consistent with the model in that this residue is for the most part buried in the hydrophobic interior but also borders the center of the DNA binding cleft. Thus, while it is unlikely that Cys-33 plays an active role in DNA complexation, perturbation of its native state by chemical modification would have dramatic consequences on the polypeptide chain conformation of G5BP as a whole and could readily produce loss of binding affinity. Its proximity to the G5BP binding surface is also consistent with the photo-cross-link that has been shown to form.

G5BP–DNA complexes are susceptible to dissociation as a function of increasing cation concentration. This suggests that positively charged amino acid side chains interact with the phosphate backbone of DNA. Indeed, the relative sensitivity of complex dissociation by various cations as measured by CD correlates with known cation affinities for nucleotide phosphates (Anderson et al., 1975). Abolition of DNA binding upon acetylation of the seven lysine residues of G5BP also supports this likelihood. It has further been observed that the side chains of some arginine residues undergo significant chemical shifts and line broadening upon G5BP complexation with $d(pA)_4$ or $d(pA)_8$ (Coleman et al., 1976). The immobilization of as many as two lysine side chains has also been detected by using NMR (Alma-Zeestraten, 1982).

The structure of G5BP (Brayer & McPherson, 1983) places all lysine and arginine side chains on the surface of the molecule. However, only two of these have immediate access to the DNA binding channel in the native state of the protein. These are Lys-46 and Arg-80. Nevertheless, the minor rearrangement of the DNA binding loop we suggest (Figure 9) would position arginines-16 and -21 in this region as well. Thus, from structural considerations, at least four positively charged protein side chains would play a role in DNA complexation.

The model of G5BP–DNA complexation (Figures 8 and 10) shows the role that Lys-46 and arginines-16, -21, and -80 would play in binding the phosphate backbone of oligonucleotides. With the possible exception of Arg-80, these charged side groups remain fairly open to the surrounding solvent. This is in keeping with the observation that DNA complexation provides no protection against the acetylation of lysine residues (Anderson et al., 1975) and subsequent complex dissociation.

The only other side chain which has been investigated is the single histidine residue at position 64 in the amino acid sequence. The proton resonances of this histidine remain unchanged upon formation of G5BP–DNA complexes (Alma et al., 1981). This is consistent with its position toward the back of the protein, removed from the general area of the DNA binding channels (Figures 2 and 4).

Intracellular Complexation. The refined native structure of G5BP and the DNA complexation model can be used to explain some aspects of the physiological behavior of this protein. Figure 10 shows the G5BP dimer structure, along with the two bound DNA chains. The closed single-stranded circle of bacteriophage fd DNA is collapsed during complexation with G5BP into what is essentially two noncomplementary antiparallel strands, covalently joined at each end (Denhardt, 1976; Coleman & Oakley, 1980). The nonlinear portion of the collapsed DNA is probably limited to the two hairpin turns at the ends of the helix.

The process by which this may occur is suggested by inspection of the three-dimensional structure of G5BP. The two channels crossing the face of the dimer (Figures 4, 7, and 10) bind to the linear portions of the collapsed phage DNA. Indeed, such interactions are likely to initiate compacting of the circular DNA as well. The antiparallel feature of the two strands is accommodated by the reversed direction of the dyad symmetry related DNA binding clefts which allows each bound DNA strand to experience exactly the same binding environment. The absence of sequence specificity allows G5BP

to bind anywhere along the entire length of the bacteriophage DNA. The covalently closed circular viral DNA has direct access to binding channels from the surrounding solvent; thus, the phage DNA need not thread its way through closed channels, not possible with circular DNAs. The two bound DNA strands would be separated by approximately 35 Å, and it is unlikely there would be any interaction between the two DNA strands, which is in keeping with evidence from other physical methods (Day, 1973; Alma et al., 1982).

Complexation of G5BP dimers on either side of the first interactive unit would be the subsequent step in forming the nucleoprotein complex observed in infected cells (Gray, C. W., et al., 1982). The cooperative nature of this process (Cavalieri et al., 1976; Alberts et al., 1972) suggests that dimer–dimer interactions are enhanced or allowed by protein conformational changes induced upon DNA complexation. This is consistent with the observation that in the absence of DNA, units larger than G5BP dimers are not observed. The present G5BP–DNA complexation model suggests that these conformational changes may be confined to the DNA binding loop and the complex loop (Figure 9).

The G5BP–DNA complexation model presented here serves to correlate and incorporate observations from a broad range of chemical and physical techniques. Furthermore, it provides a reasonable explanation for binding short oligonucleotides or the natural ligand of G5BP, the bacteriophage fd DNA. Therefore, the present model of protein–nucleic acid complexation should serve as a valuable structural framework for further studies of this system.

Acknowledgments

We thank D. Matthews and R. Feldmann for freely providing access to graphics monitors, F. Coffman for assistance in producing schematic drawings, Dr. R. Wing for provision of the DNA geometry idealization program, and G. DeLozier for providing the programs used to produce many of the illustrations.

Registry No. $d(pA)_4$, 15279-64-8; d(CTTC), 70029-17-3; $d(CG)_3$, 58927-26-7; d(CCG), 81994-13-0.

References

Alberts, B., Frey, L., & Delius, H. (1972) *J. Mol. Biol. 68*, 139–152.

Alberts, B. M., Amodio, F. J., Jenkins, M., Gutmann, E. D., & Ferris, F. L. (1968) *Cold Spring Harbor Symp. Quant. Biol. 33*, 289–305.

Alma, N. C. M., Harmsen, B. J. M., Hull, W. E., van der Marel, G., van Boom, J. H., & Hilbers, C. W. (1981) *Biochemistry 20*, 4419–4428.

Alma-Zeestraten, N. C. M. (1982) Ph.D. Thesis, University of Nijmegen, Holland.

Anderson, R. A., Nakashima, Y., & Coleman, J. E. (1975a) *Biochemistry 14*, 907–917.

Anderson, R. A., Nakashima, Y., & Konigsberg, W. (1975b) International Symposium on Photobiology, Williamsburg, VA, June 1975.

Brayer, G. D., & McPherson, A. (1983) *J. Mol. Biol.* (in press).

Cavalieri, S. J., Neet, K. E., & Goldthwait, D. A. (1976) *J. Mol. Biol. 102*, 697–711.

Coleman, J. E., & Armitage, I. M. (1978) *Biochemistry 17*, 5038–5045.

Coleman, J. E., & Oakley, J. L. (1980) *CRC Crit. Rev. Biochem. 7*, 247–389.

Coleman, J. E., Anderson, R. A., Ratcliffe, F., & Armitage, I. M. (1976) *Biochemistry 15*, 5419–5430.

Cuyper, T., Van der Ouderaa, F. J., & De Jong, W. W. (1974) *Biochem. Biophys. Res. Commun. 59*, 557–564.

Day, L. A. (1973) *Biochemistry 12*, 5329–5339.

Denhardt, D. T. (1975) *CRC Crit. Rev. Microbiol. 4*, 161–223.

Felsenfeld, G., Sandeen, G., & von Hippel, P. H. (1963) *Proc. Natl. Acad. Sci. U.S.A. 50*, 644–651.

Garssen, G. J., Tesser, G. I., Shoenmakers, J. G. G., & Hilbers, C. W. (1980) *Biochim. Biophys. Acta 607*, 361–371.

Gray, C. W., Kneale, G. G., Leonard, K. R., Siegrist, H., & Marvin, D. A. (1982) *Virology 116*, 40–51.

Gray, D. M., Gray, C. W., & Carlson, R. D. (1982) *Biochemistry 21*, 2702–2713.

Hendrickson, W. A., & Konnert, J. H. (1980) *Acta Crystallogr., Sect. A A36*, 344–350.

Herrick, G., & Alberts, B. (1976) *J. Biol. Chem. 251*, 2133–2141.

Hilbers, C. W., Garssen, G. J., Kaptein, R., Schoenmakers, J. G. G., & van Boom, J. H. (1978) in *Nuclear Magnetic Resonance Spectroscopy in Molecular Biology* (Pullman, B., Ed.) pp 351–364, D. Reidel Publishing Co., Dordrecht, The Netherlands.

Lica, L., & Ray, D. S. (1977) *J. Mol. Biol. 115*, 45–59.

McPherson, A., Molineux, I., & Rich, A. (1976) *J. Mol. Biol. 106*, 1077–1081.

McPherson, A., Jurnak, F. A., Wang, A. H., Molineux, I., & Rich, A. (1979) *J. Mol. Biol. 134*, 379–400.

McPherson, A., Wang, A. H., Jurnak, F. A., Molineux, I., Brayer, G. D., Fitzgerald, P. D., & Rich, A. (1982) in *Proceedings of the Second SUNYA Conversation in the Discipline Biomolecular Stereodynamics* (Sarma, R. H., Ed.) Vol. II, Adenine Press, New York.

Nakashima, Y., Dunker, A. K., Marvin, D. A., & Konigsberg, W. (1974) *FEBS Lett. 43*, 125.

O'Connor, T. P., & Coleman, J. E. (1983) *Biochemistry 22*, 3375–3380.

Oey, J. L., & Knippers, R. (1972) *J. Mol. Biol. 68*, 125–138.

Paradiso, P. R., & Konigsberg, W. (1982) *J. Biol. Chem. 257*, 1462–1467.

Pratt, D., Laws, P., & Griffith, J. (1974) *J. Mol. Biol. 82*, 425–439.

Pretorius, H. T., Klein, M., & Day, L. A. (1975) *J. Biol. Chem. 250*, 9262–9269.

Rasched, I., & Pohl, F. M. (1974) *FEBS Lett. 46*, 115–118.

Ray, D. S. (1977) *Compr. Virol. 7*, 105–178.

Torbet, J., Gray, D. M., Gray, C. W., Marvin, D. A., & Siegrist, H. (1981) *J. Mol. Biol. 146*, 305–320.