Protein Science (1995), *4*:187–197. Cambridge University Press. Printed in the USA.

Electrostatic potential distribution of the gene V protein from Ff phage facilitates cooperative DNA binding: A model of the GVP–ssDNA complex

YUE GUAN, HONG ZHANG, AND ANDREW H.-J. WANG

Biophysics Division and Department of Cell & Structural Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(RECEIVED September 16, 1994; ACCEPTED November 29, 1994)

Abstract

The crystal structure of the gene V protein (GVP) from the Ff filamentous phages (M13, f1, fd) has been solved for the wild-type and two mutant (Y41F and Y41H) proteins at high resolution. The Y41H mutant crystal structure revealed crystal packing interactions, which suggested a plausible scheme for constructing the polymeric protein shell of the GVP–single-stranded DNA (ssDNA) complex (Guan Y, et al., 1994, *Biochemistry 33*:7768–7778). The electrostatic potentials of the isolated and the cooperatively formed protein shell have been calculated using the program GRASP and they revealed a highly asymmetric pattern of the electrostatic charge distribution. The inner surface of the putative DNA-binding channel is positively charged, whereas the opposite outer surface is nearly neutral. The electrostatic calculation further demonstrated that the formation of the helical protein shell enhanced the asymmetry of the electrostatic distribution. A model of the GVP–ssDNA complex with the $n = 4$ DNA-binding mode could be built with only minor conformational perturbation to the GVP protein shell. The model is consistent with existing biochemical and biophysical data and provides clues to the properties of GVP, including the high cooperativity of the protein binding to ssDNA. The two antiparallel ssDNA strands form a helical ribbon with the sugar-phosphate backbones at the middle and the bases pointing away from each other. The bases are stacked and the Phe 73 residue is intercalated between two bases. The optimum binding to a tetranucleotide unit requires the participation of four GVP dimers, which may explain the cooperativity of the GVP binding to DNA.

Keywords: electrostatics; molecular modeling; protein–DNA interactions; single-stranded DNA-binding protein; X-ray crystallography

The interactions of the gene V protein from the Ff filamentous bacteriophages (M13, f1, and fd) with its viral single-stranded DNA play critical roles in the life cycle of those phages (Salstrom & Pratt, 1971). GVP is a dimeric protein of 87 amino acids (Fig. 1; Kinemage 1) and it binds to DNA in a highly cooperative manner without pronounced sequence specificity (Sang & Gray, 1987; Bulsink et al., 1988a; Fulford & Model, 1988).

In the presence of a ssDNA, GVP binds DNA cooperatively to form a helical protein–DNA complex. Gray (1989) has shown by electron microscopy that such helical complexes are likely left-handed with a pitch of ~90 Å. Each turn of the helix contains about eight GVP dimers with an outer diameter of ~80 Å. Other studies indicated that two antiparallel ssDNA strands occupy the interior of the helix (Gray et al., 1982) and GVP binds to polymeric DNA and oligonucleotide, respectively, at a ratio of four and three nucleotides per protein monomer (Bulsink et al., 1988b). More detailed studies regarding the interactions between GVP and DNA have been carried out (reviewed in Kansy et al., 1986). In particular, NMR and fluorescence spectroscopies have implicated several amino acids of GVP that are in contact with DNA in the GVP–ssDNA complex (King & Coleman, 1987, 1988; Dick et al., 1989; Stassen et al., 1992; Folkers et al., 1993).

This impressive level of knowledge on the system notwithstanding, a more definitive approach to visualize the protein–DNA interactions is to have the three-dimensional structure of the protein–DNA complex by either X-ray crystallography or NMR spectroscopy. As the first step toward this goal, we have recently determined the crystal structure of the wt-GVP (Skinner

Reprint requests to: Andrew H.-J. Wang, Biophysics Division and Department of Cell & Structural Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801; e-mail: ahjwang@uiuc.edu.

Abbreviations: GVP, gene V protein; wt-GVP, wild-type GVP protein; Y41H and Y41F, Tyr 41 → His 41 and Tyr 41 → Phe 41 mutant GVP; ssDNA, single-stranded DNA; RMSD, RMS deviation; SA, simulated annealing; MAD, multiple wavelength diffraction.

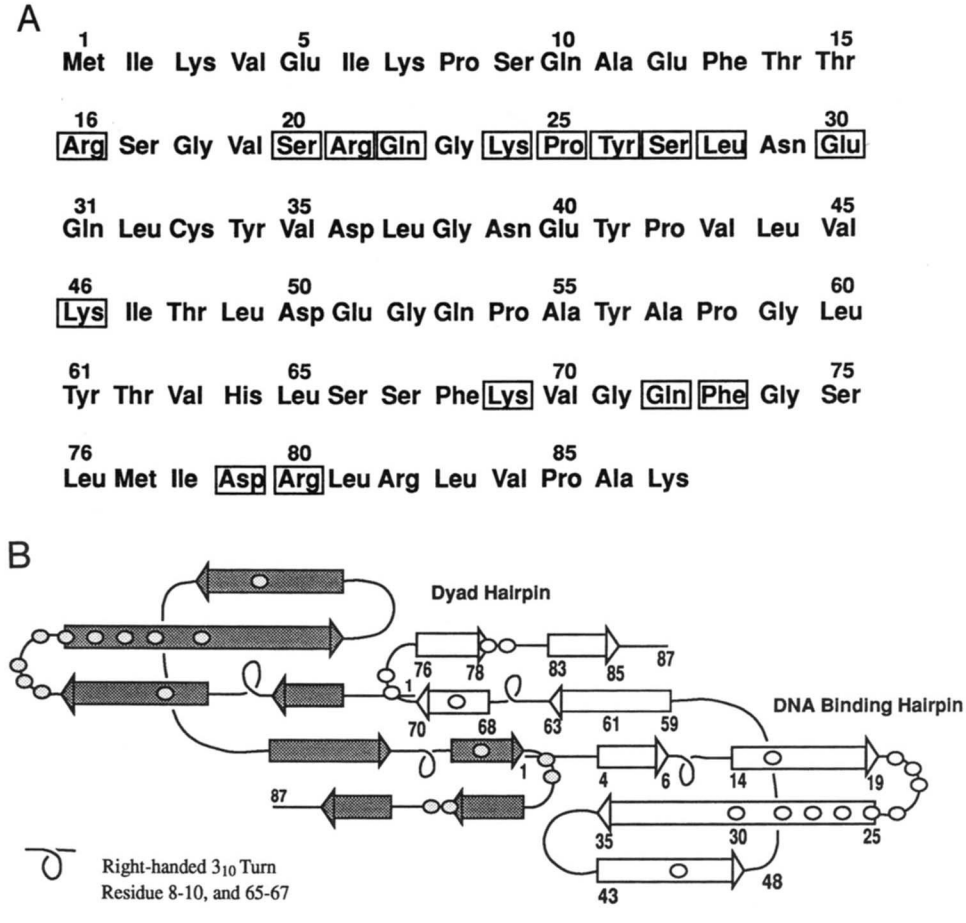

Fig. 1. A: Amino acid sequence of GVP (Cuypers et al., 1974) with those amino acids that are in contact with DNA marked in boxes. **B:** Topology diagram of the secondary structure of the gene V protein dimer. Amino acids that are in contact with DNA are marked in circles. Two β-hairpins are labeled according to their functions.

et al., 1994) as well as two mutants with changes at the tyrosine 41 position (Kinemage 2; Guan et al., 1994). The importance of Y41 is that it participates directly in the protein dimer–dimer contacts in the GVP–ssDNA complex (King & Coleman, 1988; Stassen et al., 1992). The three crystal structures revealed that Y41 indeed is involved intimately in the protein–protein contacts, but with two distinct modes (Guan et al., 1994; Skinner et al., 1994). In the wild-type and Y41F structures, the loop (residues 36–43) that contains the 41 position from one monomer protein interacts with the same loop from the crystallographic twofold symmetry-related monomer protein of a neighboring GVP dimer. In contrast, the Y41H mutant protein crystallized in a new crystal lattice and the H41-containing loop now interacts with the β-strand near the residue lysine 69. The new structural information further allowed us to construct a plausible model of the protein shell of the GVP–ssDNA complex (Guan et al., 1994). A similar model has been proposed using the three-dimensional structure of Y41H GVP determined by NMR (Folkers et al., 1994; Folmer et al., 1994).

Our earlier work did not include DNA in the model, even though the model of the protein shell clearly revealed a double-grooved channel in the interior of the protein polymer. DNA is likely to occupy this deep channel. It would be useful to have some ideas regarding the possible binding mode of DNA in the GVP–ssDNA complex. In the present work, we show that, by computer modeling studies, the DNA-binding channel of GVP can accommodate two antiparallel ssDNAs with four nucleotides per monomer protein. We have constructed a model of the GVP–ssDNA complex that explains satisfactorily the known biochemical and biophysical data so far (Kinemage 3).

Results

Electrostatic distribution of the GVP dimer

The structure of the dimer GVP protein is shown in the ribbon drawing (Fig. 2) with relevant amino acids that are involved in the binding to DNA displayed individually (Kinemage 3). Each monomer contains a distorted five-stranded β-barrel core with two prominently extended β-hairpins, the so-called dyad hairpin and the DNA-binding hairpin. The topology of the secondary structure of the dimer GVP is shown schematically in Figure 1B. The dimer GVP produces a concave inner surface covered with amino acid side chains from both hairpins, including those of R16, R21, K24, Y26, K46, R80, K69′, and F73′ (′ is from the symmetry-related monomer). Many of those amino acids are positively charged and their distribution on the surface of the GVP dimer is unusual. This is evident in the electrostatic potential calculation discussed below.

Figure 3 shows the electrostatic potential of two adjacent GVP dimers found in the crystal packing of the monoclinic C2 crystal of the wild-type and Y41F GVP. The concave inner surface has a large patch of positive potential (in blue color; $>11\ k_BT$). In this crystal lattice, the dimer-contact loops (residues 36–43)

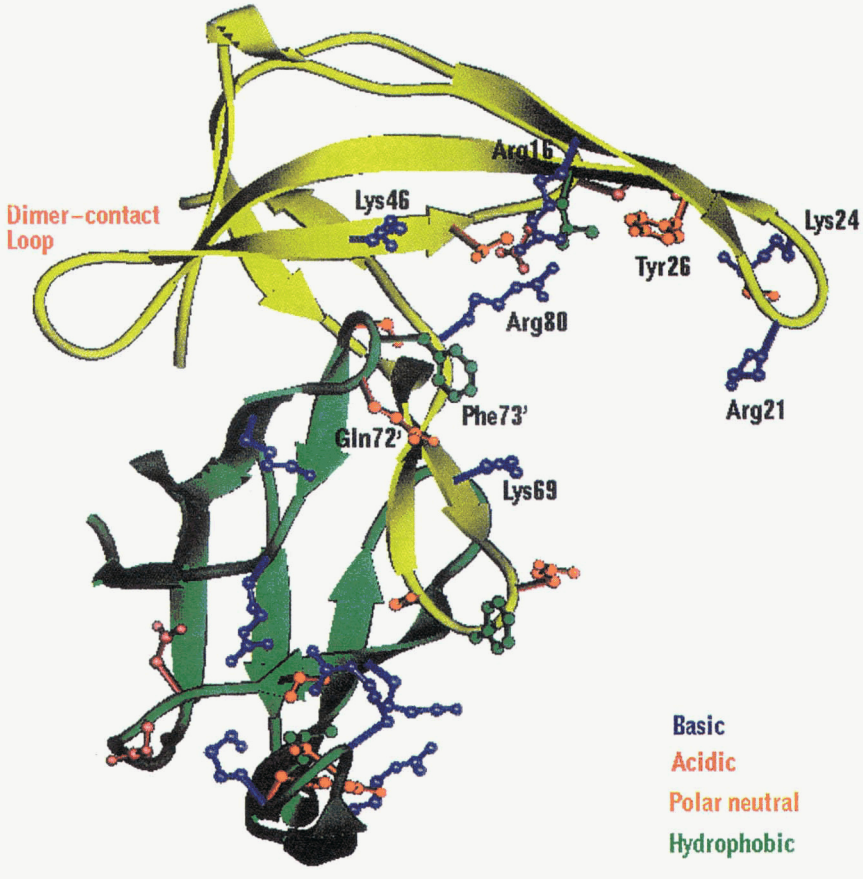

Fig. 2. Dimer structure of GVP. The two monomers are related by a crystallographic twofold axis that is tilted by ~45° to the paper. The picture was generated by RIBBON (Carson, 1987). Side chains of the amino acids that are in contact with DNA are shown.

are involved in the crystal packing and the two adjacent dimers are related by a unit cell translation along the *c*-axis direction. The aromatic side chain of Y41 (or F41) residue of one dimer intercalates into a cavity generated by the Y34, P42, and Y41 (or F41) amino acids of the neighboring dimer and vice versa (Guan et al., 1994). This mutually intercalating loop interaction creates a very tight binding between the two dimers, resulting in a virtual continuation of molecular surface at the dimer–dimer contact site (Fig. 3). In this case, the centers of the two dimers are separated by 42.6 Å, the length of *c*-axis. The electrostatic potential did not appear to be significantly affected in going from an isolated GVP dimer to the oligomerized GVP dimers linked through the direct loop–loop interactions. The other side of the dimer is essentially neutral (in white color), showing very little variation in the charge distribution (see below).

A dramatically different result appears when the polymer GVP protein shell is considered. As discussed before (Guan et al., 1994), we have constructed the protein shell on the basis of the crystal packing contacts in the Y41H crystal that involve the loop 36–43, Y56, and strands 59–68 and 78′–85′ near H64′ (see Fig. 1B for the locations of those amino acids). We have calculated the electrostatic potential of the GVP dimer in the protein polymer environment, exemplified by two neighboring dimers extracted from the polymer (Fig. 4). The blue color (therefore the positive potential) inside the channel appears to intensify significantly. This suggests that the tight and compact interactions between the neighboring dimers may reinforce the positive charge distribution between each other. This should be desirable for the DNA-binding interactions. This is more clearly seen in the electrostatic potential surface of the entire protein shell (Fig. 5). The inner surface has a continuous helical ribbon of positive potential, whereas the outer surface is essentially neutral due to a lack of strongly charged amino acids on this surface.

A model of the GVP–ssDNA complex

Each putative DNA-binding channel in the aforementioned protein shell model is about 30 Å in diameter on the average and is very protected from the solvent. This prompted us to ask whether those channels have sufficient room to accommodate two antiparallel ssDNA. Indeed, we were able to successfully place two DNA strands into the left-handed helical GVP protein shell using the procedure described in the Materials and methods section.

The global view of the model of the GVP–ssDNA complex is shown in Figure 6 and Kinemage 4, with the ribbon-like DNA molecules located close to the center of the helix. The two DNA strands are not plectonemically interwound and their sugar–phosphate backbones line up in the middle of the ribbon. The DNA (20-Å wide) ribbon forms a slender left-handed helix with an outer diameter of 40 Å. There are four nucleotides per protein monomer and the bases are stacked, but not base paired. The bases are oriented such that their base planes are nearly parallel (tilted by ~25°) to the helix axis. The pitch of the DNA he-

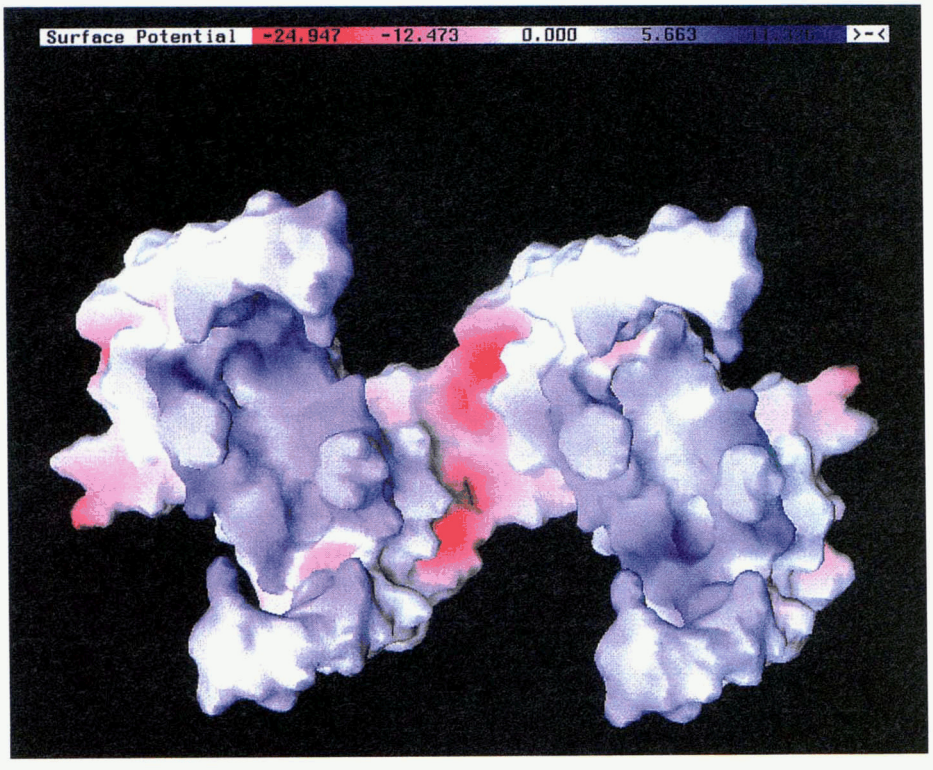

Fig. 3. Electrostatic potential distribution of the wt-GVP dimers found in the crystal packing of the monoclinic C2 unit cell viewed along the *b*-axis. The two dimers are related by the *c*-axis unit cell translation and are separated by 42.6 Å. The electrostatic potential distribution here is not significantly different from that of an isolated GVP dimer. The scale of the surface potential (in k_BT, where k_B is the Boltzmann constant) is shown at the top of the figure, with the blue color representing strong positive charge and the red color representing strong negative charge.

lix is 90 Å, the same as that of the protein shell. There are 32 DNA bases and 8 phenyl rings per turn of the helix.

The tight fit of the ssDNA in the channel is more vividly illustrated in Figure 7. It can be seen that each ssDNA occupies one side of the channel. There are four stacked bases separated by an intercalated phenyl ring (of Phe 73′). An important consequence of the interactions between the protein and DNA is the neutralization of the electrostatic potential. The inner surface of the protein shell now is almost neutral as evident from the very faint red color (Fig. 7). The DNA in the complex is completely buried, with the DNA-binding hairpins of GVP closing off the entrance to the channel.

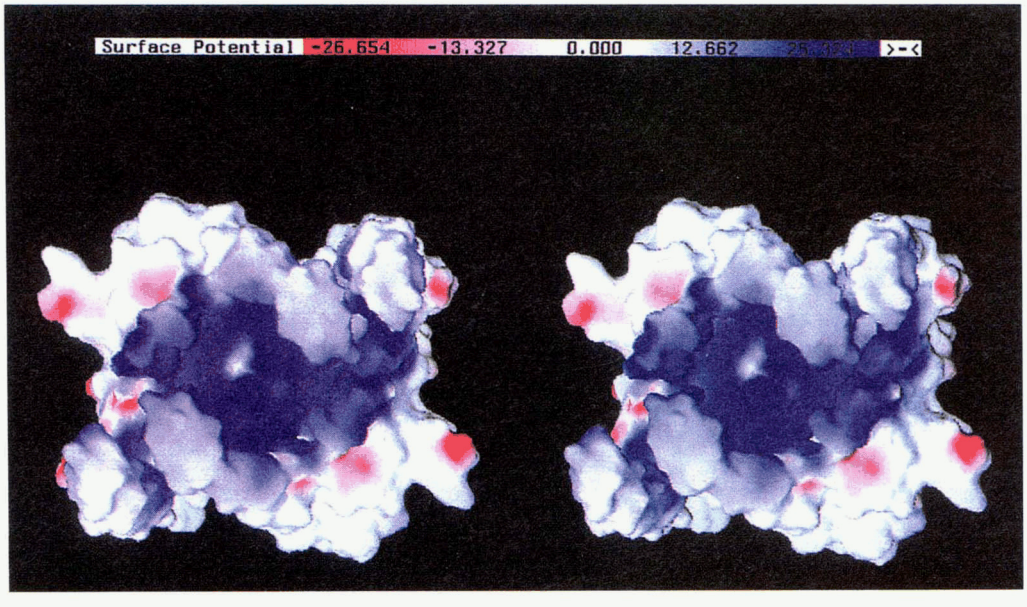

Fig. 4. Electrostatic potential distribution of two adjacent wt-GVP dimers found in the model of the protein shell of GVP polymer (protein only). The electrostatic potential distribution in the deep interior channel is significantly more positively charged in comparison with that of an isolated GVP dimer. The scale of the electrostatic potential has been adjusted (with maximum positive value being −22) relative to that used in Figure 3. This is done to achieve a more suitable contrast.

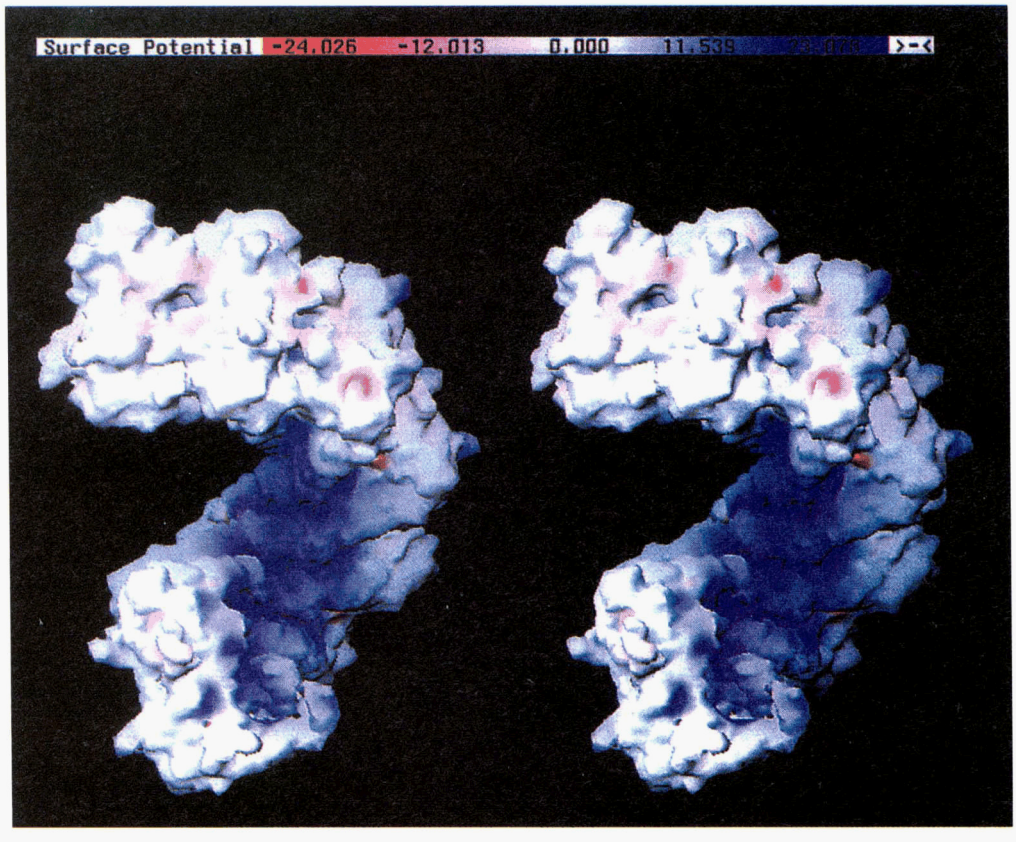

Fig. 5. Electrostatic potential distribution of one turn (eight GVP dimers) of the protein shell of GVP polymer.

Although the stoichiometry of DNA to protein is four nucleotides per GVP monomer, each tetranucleotide unit in fact interacts with several GVP monomers. This is more easily understood from Figure 8, in which three contiguous GVP dimers and two strands of dodecamer nucleotides are shown. The central four nucleotides are positioned in such a way that one side of the DNA molecule is primarily in contact with the inner wall of the (*n*)GVP dimer, but the other side is covered by the DNA-binding hairpin of the $(n-1)$GVP monomer. Additional contacts also exist between the DNA and the $(n+1)$GVP and, to a smaller extent, the $(n+2)$GVP dimers.

Therefore, each repeating tetranucleotide unit in the complex is surrounded by four different GVP dimers, with their detailed interactions depicted in Figure 9. Here we can see several important features. The phenyl ring of the Phe 73′ residue is intercalated between the $(n-1)$A4 and (*n*)A1 bases. The positively charged (*n*)K69′, (*n*)R80′, $(n-1)$R21, and possibly $(n+1)$R21′ side chains of GVP are intimately involved with the DNA phosphate groups. However, two positively charged residues from the DNA-binding hairpin, (*n*)R16 and $(n-1)$K24, appear to interact with bases and sugars only. The (*n*)Y26 tyrosyl side chain is in close contact with the $(n-1)$F73′ side chain and (*n*)A4 base, but is not intercalated between bases. Its phenolic hydroxyl group is close to (*n*)A3 and (*n*)R16 residues. The close contacts between DNA and the GVP dimer are summarized in Table 1. The amino acids that are in contact with DNA in the complex are identified as circles in the secondary structure representation of Figure 1B.

To facilitate the viewing of those DNA–protein interactions, a schematic diagram is shown in Figure 10. This approximate "cylindrical projection" of three contiguous dimers shows that amino acids F73′, K46, R16, L28, R80, Y26, K24, and finally R21 are along the pathway of the DNA trail from left to right. The distance between two helically related amino acid side chains, e.g., from $(n-1)$F73′ to (*n*)F73′, in the inner wall of the helix is about 17 Å, ideally suited for four stacked bases.

Discussion

Recognition and binding of DNA by proteins are important processes in life. Great advances have been made in the understanding of sequence-specific DNA recognition by proteins that have led to the identification of various DNA-binding structural motifs (helix-turn-helix, zinc-finger, leucine zipper, etc.) (Pabo & Sauer, 1992). Electrostatic potential of these proteins may play important roles in the protein–DNA recognition, as shown in the example of the *Escherichia coli* catabolic gene activation protein (CAP) binding to its DNA-binding site (Warwicker et al., 1987). Additionally, many proteins interact with DNA double helix in a nonsequence-specific manner. A conspicuous example is the histone octamer core, which is composed of highly positively charged histone proteins (Arents et al., 1991). How the right-handed B-DNA double helix interacts with the histone core may be dictated by the charge distribution on the curved surface of the histone core. Indeed, numerous recent work has suggested the importance of the electrostatic potential distribution

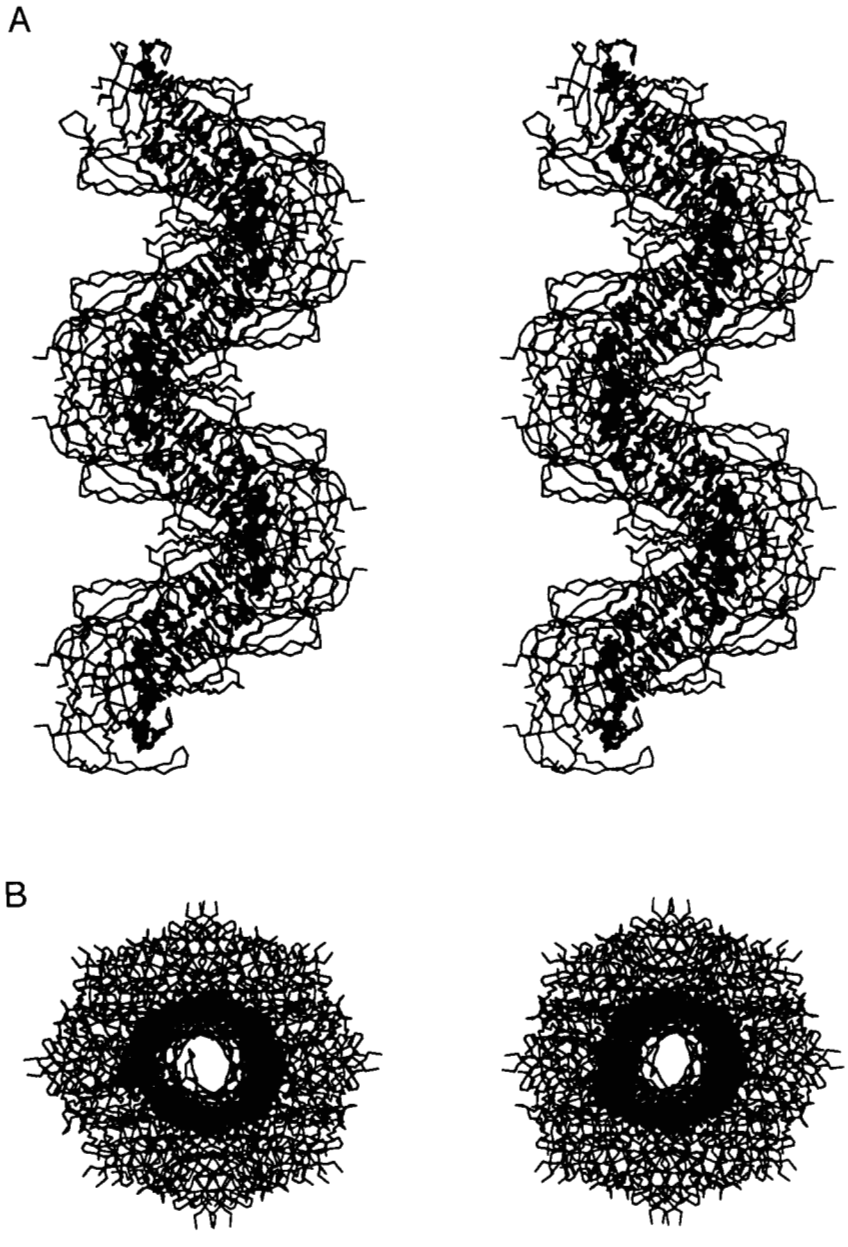

Fig. 6. Stereoscopic diagrams of two turns (16 GVP dimers) of the left-handed helix of the GVP–ssDNA complex. The pitch and the diameter of the helix are 90 Å and 80 Å, respectively. The DNA double helix (dark shaded) occupies near the center of the helix and they are not interwound, nor are they base paired. **A:** Side view. **B:** End view.

of proteins in their biological functions. Selected examples could be found in the electron transfer system of the plastocyanin–cytochrome *c* complex (Roberts et al., 1991), DNA-binding protein (Nikolov & Burley, 1994), and the enzyme acetylcholinesterase (Gilson et al., 1994).

The high-resolution crystal structures of the wt- (Skinner et al., 1994), the Y41F, and Y41H mutant (Guan et al., 1994 and this work) GVP proteins, and the high-resolution NMR structure of the Y41 mutant (Folkers et al., 1994) afford us a reliable molecular structure of the ssDNA-binding proteins from Ff phages. Our calculation of the electrostatic potential of the GVP in various environments provided insight into the DNA-binding properties. In an isolated GVP dimer, the highly asymmetric charge distribution is quite striking (Fig. 3). The deep concave surface has a large area of positive potential, whereas the opposite surface is neutral. This charge asymmetry should facilitate the binding of GVP to DNA using its concave surface.

When we used the GVP dimer embedded in the protein shell of the GVP-polymer for the electrostatic calculation, a strengthening of the positive charge in the DNA-binding channel was noted (Fig. 4). This may indicate that binding of the second GVP dimer next to the first GVP that is already bound to DNA enhances binding to DNA for both GVP dimers. This may contribute to the observed cooperativity of the GVP binding to DNA.

Our model of the GVP–ssDNA explains much of the existing biochemical and biophysical data. For instance, its global structural parameters are consistent with those obtained from the electron microscopy studies on the GVP–ssDNA complex, i.e., a left-handed superhelix with about eight dimers per turn of helix. The pitch of the helix ranges from 60 to 120 Å, and the diameter is about 80 Å (Gray, 1989). Our model shows that a ribbon of two ssDNA resides inside the helix (Fig. 6), but not wrapping around the outside of the helix. This is consistent with

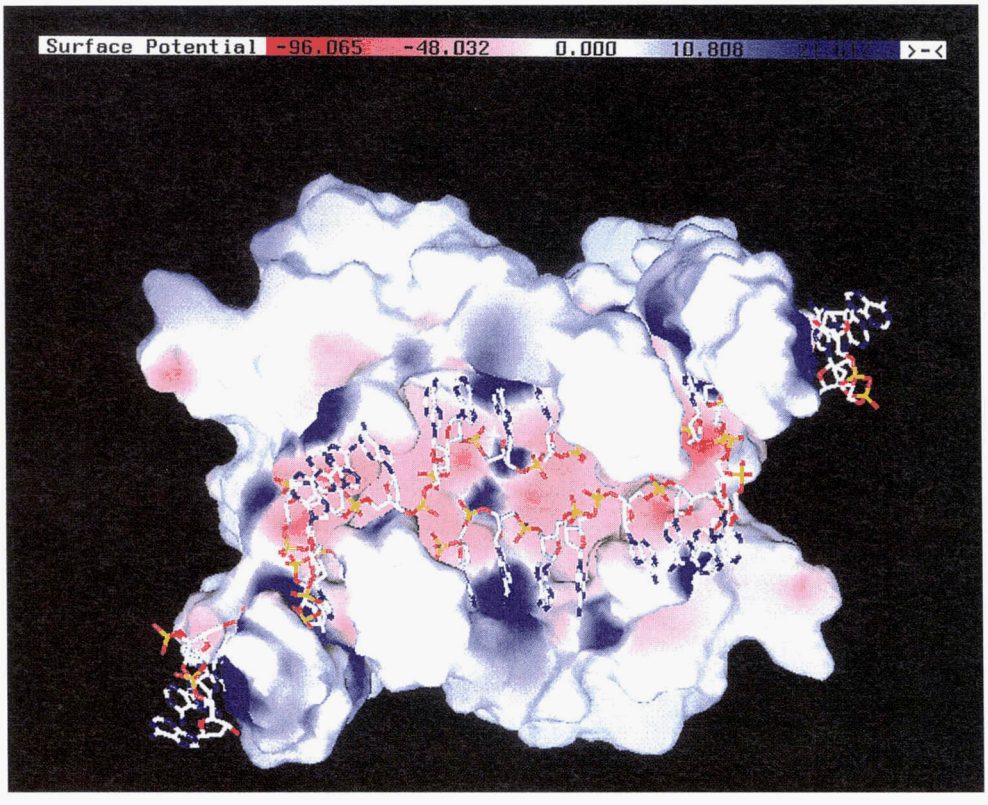

Fig. 7. Electrostatic potential distribution of two adjacent wt-GVP dimers found in the model of the GVP–ssDNA complex. Two ssDNA hexadecamer nucleotides are included in the calculation. The electrostatic potential distribution in the deep interior channel of the protein shell is almost completely neutralized by the negatively charged DNA phosphates.

the neutron-scattering studies, which deduced a cross-sectional radius of gyration of 17.6 ± 3.0 Å for DNA (Gray et al., 1982). Figure 6B shows that the inner radius of the DNA helix is 12 Å and the outer radius is 20 Å, in striking agreement with the neutron-scattering data. Our model also fully took into account the importance of the Y41 residue in the assembly of the polymer GVP proteins by intimately involved Y41 in the GVP dimer–dimer interactions, as suggested by earlier spectroscopic studies (King & Coleman, 1988; Folkers et al., 1994), and this has been elaborated previously (Guan et al., 1994). It should be noted that various models of the GVP–ssDNA complex, based on the incorrect crystal structure of GVP, have been proposed previously (Brayer & McPherson, 1984; Hutchinson et al., 1990). We believe that those earlier models are erroneous and should be discarded. In contrast, our model appears to be similar to a recent GVP–DNA complex model that has been described very briefly by Folmer et al. (1994).

The GVP dimer did not require any substantial conformational changes in order to accommodate the DNA strands. This is consistent with the conclusion derived from the experimental

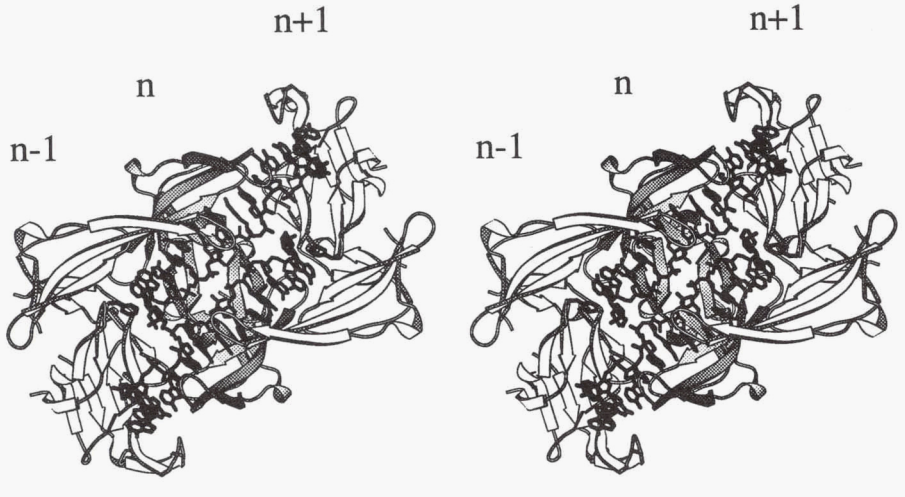

Fig. 8. MOLSCRIPT (Kraulis, 1991) drawing showing three GVP dimers bound to two dodecamer ssDNA strands.

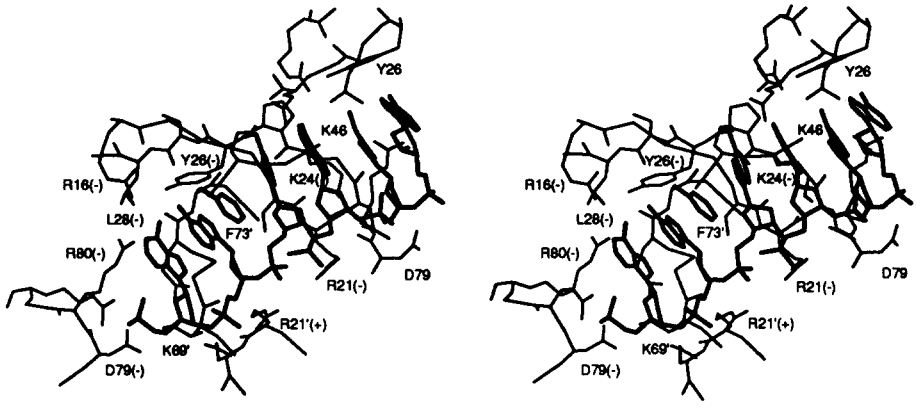

Fig. 9. Local interactions of the protein–DNA contacts in the model of the GVP–ssDNA complex. Amino acids labeled (+) and (−) are from the $(n + 1)$ and $(n - 1)$ dimers, respectively. Prime (′) denotes the amino acid is from the intradimer twofold symmetry-related monomer.

CD spectroscopic data (Sang & Gray, 1989). The RMSD between the wt-GVP crystal structure and the model GVP structure is 1.00 Å, with the only significant changes occurring near the tip of the DNA-binding hairpins (amino acids R21, Q22, G23, K24) and near the Phe 73 position.

The new model of the complex also explains the possible role of several amino acids that are involved in DNA binding (e.g., Y26, F73, and R80; Kinemage 3). NMR studies of the spin-labeled oligo(dA) and the lanthanide shift reagent $Gd(DOTP)^{+5}$ binding to DNA have placed the following amino acids, Ser 20, Tyr 26, Leu 28, Phe 73, in close distance, and Arg 16, Gly 18, Arg 21, Ser 27, Thr 48, Asp 50, Ser 75, and Arg 80 in intermediate distance, to DNA (van Duynhoven et al., 1993; Folmer et al., 1994). Inspection of Table 1 and Figure 9 indicates that our model is quite congruent with the observed data. NMR and fluorescence data suggested that Phe 73 and Tyr 26 rings became buried in a hydrophobic environment upon the binding of GVP to DNA (King & Coleman, 1987; Stassen et al., 1992). Our model indicates that the Phe 73 is intercalated between DNA bases, whereas Tyr 26 is enclosed in a hydrophobic pocket.

An important finding here is that the binding of a tetranucleotide repeating unit requires the participation of four contiguous GVP dimers (Figs. 8, 10). This could explain the high cooperativity of the GVP binding to DNA ($\omega > 500$) (Bulsink et al., 1988a). Our model may also help explain the $n = 3$ binding mode for the oligonucleotide binding by GVP. In Figure 8, there are three GVP dimers bound to two single strands of a $(dA)_{12}$ dodecamer DNA. Note that some portions of the DNA (e.g., the four nucleotides in the upper left region) are exposed and available for additional GVP binding. It is likely that one more GVP at the $(n - 2)$ site, making a total of four GVP dimers, will achieve the optimum binding affinity for a dodecamer DNA. This results in an $n = 3$ binding ratio because there are 12 nucleotides for 4 GVP monomers.

Can this model address the question on the formation of the filament produced by GVP and M13 circular ssDNA, as seen under the electron microscope? We believe that a plausible mechanism can be put forth. The DNA ribbon in the channel has the two antiparallel DNA sugar–phosphate backbones close to each other. The closest distances between the two symmetry-related backbones are from O1P1–O1P1′ (3.1 Å) and O2P1–O2P1′

Table 1. *Amino acid residues in contact with DNA (<3.5 Å) in the GVP–ssDNA complex model*

	Dimer GVP[a]			
DNA[b]	$(n - 1)$	(n)	$(n + 1)$	$(n + 2)$
B 1	P25(B), Y26(B)	F73′(S), K46(S)		
S 1	R21(S), K24(S)	F73′(S)		
P 1	S20(S), R21(S)		R21′(S)	
B 2	K24(S)	R16(S), E30(S)		
S 2				
P 2	R21(S)			
B 3		R16(S), L28(S)		
S 3	Q22(S), K24(S)		Q72′(S)	Q22′(S)
P 3		D79(B,S), R80(S)	Q72′(S), K69′(S)	
B 4		Y26(S), S27(B)	Q72′(S), F73′(S)	
S 4				R21′(S)
P 4			Q72′(S)	R21′(S)

[a] Distances between P3 and K69′(S), P1 and S20(S), and S3 and Q22(S) are 3.82 Å, 3.63 Å, and 3.72 Å, respectively. B denotes backbone atoms; S denotes side-chain atoms; and ′ denotes residues from monomer 2 of the dimer.

[b] B, S, and P denote base, sugar, and phosphate, respectively.

Table 2. *Crystal data and refinement parameters of three gene V proteins*

	Wild type	Y41F	Y41H
Space group	C2	C2	$P2_12_12_1$
a (Å)	75.81	74.95	77.19
b (Å)	27.92	28.01	83.94
c (Å)	44.40	42.74	28.81
β (°)	103.08	103.15	—
Resolution (Å)	1.6	1.6	1.8
Number of reflections ($\|F\| > 2\sigma$)	9,066	9,392	13,837
R_{merge} (%)[a]	4.3	3.7	4.5
Completeness of data	78%	81%	78%
R-factor	0.215	0.218	0.210
Number of water molecules	45	25	48
RMS bond length deviation (Å)	0.017	0.017	0.018
RMS bond angle deviation (°)	3.3	3.0	3.3

[a] $R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$.

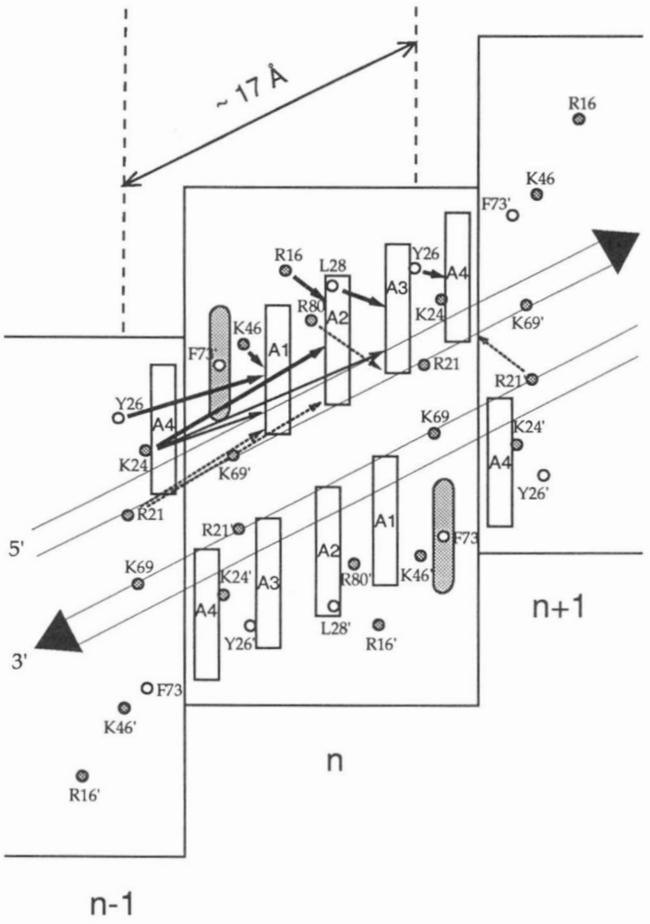

Fig. 10. A schematic diagram depicting an approximate "cylindrical projection" of three dimers of the model of the GVP-ssDNA complex. Two long arrows represent two antiparallel ssDNA sugar-phosphate backbones in the interior of the superhelix. The distance between two equivalent amino acids is around 17 Å, which accommodates a tetranucleotide with the bases stacked. The phenyl ring of Phe 73′ is intercalated between $(n-1)$A4 and (n)A1 bases. Interactions between amino acids and DNA are indicated with thick and thin solid arrows for binding to DNA bases and sugars, and dotted arrows to phosphates, respectively.

(3.1 Å). It is very easy to have one strand of DNA on one side of the channel making a sharp turn, reversing the strand in the opposite direction and switching into the twofold related site. The DNA structure in this case is essentially a "hairpin." This mechanism allows the GVP binding first to a high affinity site (e.g., a stretch of pyrimidine-rich sequence) of the close-circular M13 DNA, forcing a "hairpin" to form, and subsequently adding more proteins cooperatively to the remaining vacant DNA sites like a zipper action. Eventually all DNA sequences are occupied and another hairpin is formed to cap the other end of the filament. If there is more than one initiation site, a multi-armed GVP-ssDNA complex will form, as observed by electron microscopy (Gray, 1989).

Recently, the structure of another ssDNA phage Pf3 has been proposed in which the DNA in the virion has a very unusual double helical structure with the sugar-phosphate backbone in the center of the protein-DNA assembly (Liu & Day, 1994). If the M13 phage has an analogous virion structure, it is tempting to speculate that our GVP-ssDNA model, with the sugar-phosphate already in the center, may make it easy for the gene VIII protein (the coat protein) to displace the GVP dimer and bind to DNA without extensively rearranging the DNA conformation.

Before we conclude, it should be pointed out that, although our model is likely to be correct in general features because it can explain much known data satisfactorily, some small adjustments are probably needed. For one thing, GVP has little sequence specificity and our model is based on a poly(dA) sequence. Therefore, a mixed sequence DNA will inevitably alter the conformation of the bound DNA or the conformation of the amino acids that are in contact with DNA. Furthermore, the GVP-ssDNA complex is somewhat flexible (Gray, 1989). It is reasonable to assume that some degrees of variability of the protein dimer-dimer contacts as well as the protein-DNA contacts exist and are, in fact, desirable. Nevertheless, we feel that in the absence of the crystal structure of a GVP-DNA complex, our model provides an excellent framework for understanding the GVP-DNA interactions.

Materials and methods

Structure determination and refinement

The structure determination of the wild-type protein by the MAD method has been described before (Skinner et al., 1994). In order to have a more consistent comparison of the three protein structures (the wild type, plus the Y41F and Y41H mutants), we have improved the resolution of the structural analysis to higher resolution using diffraction data collected on a Rigaku R-Axis IIc image plate area detector system in our laboratory. The resolution of the data has been improved to 1.6 Å, 1.6 Å, and 1.8Å, respectively, for the wild type, the Y41F, and Y41H mutant proteins. All three structures have been refined by the SA procedure incorporated in X-PLOR (Brünger, 1992). During the refinement, some minor rebuildings of the model were required for the Y41H mutant. The crystal data and refinement summaries are listed in Table 2. The detailed structural refinement and structural analyses will be described elsewhere.

Model building of the GVP-ssDNA complex

The construction of the model of the GVP protein polymer in the GVP-ssDNA complex has been described previously (Guan et al., 1994). Briefly, the dimer-dimer contacts found in the Y41H crystal packing, in conjunction with existing biochemical and biophysical data, were used as the basis for model building. The final model of the protein shell in the complex is a left-handed helix with eight GVP dimers per turn of helix (pitch ~90 Å), and the diameter of the helix is approximately 80 Å.

In this model, a deep double-grooved channel residing in the interior of the protein helix is evident (see below). We wanted to show that those channels indeed have the suitable size and shape to accommodate the proper length of an ssDNA. In the active GVP-ssDNA complex found in the M13 phage-infected *E. coli*, there are four nucleotides bound per protein monomer (Bulsink et al., 1988b). We inspected the environment of the channel and noted that the repeating distance between two equivalent positions of the adjacent dimers is ~17 Å. We were intrigued to find that two neighboring Phe 73 residues are po-

sitioned such that four stacking bases can be accommodated between the two Phe 73 rings to make up the distance of 17 Å (5 × 3.4 Å). To simplify the model building, a single-stranded poly(dA) DNA was chosen. We shall denote the tetranucleotide between two Phe 73 rings as the reference site (n), with the adenines numbered as A1, A2, A3, and A4. The GVP dimer binding site on the 5′ side of the DNA is denoted as the site ($n - 1$), whereas the one on the 3′ side of DNA is denoted as ($n + 1$). Within a given dimer, one monomer is chosen as the reference with all amino acids numbered plainly, and the twofold related monomer with its amino acids numbered "prime." For example, the position 73 in the reference monomer is denoted Phe 73, whereas the symmetry-related site is denoted Phe 73′.

We first placed two adenine nucleotides across the Phe 73′ phenyl ring, i.e., the Phe 73′ ring is intercalated between two adenines. (At this time, we do not have information regarding the polarity of the DNA.) This established the locations of the ($n - 1$)A4 and (n)A1 nucleotides. The second tetranucleotide fragment was generated by the helical symmetry (−45° rotation [the minus sign means that a left-handed helix is generated] and ~11.2 Å translation along the superhelical axis direction). This further established the location of the (n)A4 nucleotide. The second and third nucleotides, i.e., (n)A2 and (n)A3, were then added between the (n)A1 and (n)A4 nucleotides. During this initial model building process, possible interactions between the positively charged amino acids (Arg and Lys) and the DNA phosphate groups were considered and care was taken to minimize bad contacts. Finally, all phosphate groups were properly joined together to produce a poly(dA) in one side of the channels. The symmetry-related DNA strand was then generated by a dyad operation.

The model was then energy minimized using X-PLOR (Brünger, 1992). One hundred cycles of Powell minimizations were performed to remove a small number of bad contacts. Both the helical symmetry and the noncrystallographic local twofold symmetry were preserved for each dimer during the energy minimization. The entire process was iterated until the energy minimization was converged. The final total energy for the complex is −9,200 kcal/mol.

The atomic coordinates of a single unit (one GVP dimer plus two tetranucleotides) of the GVP–ssDNA complex have been deposited at Brookhaven Protein Data Bank (entry number 1GPV).

Electrostatic potential calculation

The electrostatic potential was calculated using the program GRASP (Nicholls, 1993). This program uses a Poisson–Boltzmann solver, which is a similar but simpler version of that used by Delphi (Gilson et al., 1987). In order to minimize the end (boundary) effect, we did the calculations on oligomeric proteins that are either related by the crystallographic symmetry or by the helical symmetry of the model. Charged amino acid side chains were assigned with appropriate charges (+1.0 for Arg, Lys; −1.0 for Asp, Glu). No explicit hydrogen atoms were generated for the calculation. Water molecules were not included in the calculation; instead, the protein molecules were embedded in a box with the average dielectric constant of 80. The implicit ionic strength was set at 0.05 M. The buried area is given a dielectric constant of 2.

Acknowledgments

This work was supported by NIH GM-41612 and CA-52054 (A.H.-J.W.). We thank Dr. C.W. Hilbers and Dr. R. Konings of Nijmegen University for providing several GVP protein samples and useful comments. We also thank Dr. T. Terwilliger and his associates Drs. M. Skinner and D. Leschnitzer (Los Alamos National Laboratory) for their collaboration in the structure determination of the wild-type GVP by the MAD method. Useful comments from Dr. D. Gray and Dr. G.J. Thomas, Jr. are greatly appreciated.

References

Arents G, Burlingame RW, Wang BC, Love W, Moudrianakis EN. 1991. The nucleosomal core histone octamer at 3.1 Å resolution: A tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci USA 88*:10148–10152.

Brayer G, McPherson A. 1984. Mechanism of DNA binding to the gene 5 protein of bacteriophage fd. *Biochemistry 23*:340–349.

Brünger AT. 1992. *X-PLOR 3.1, a system for X-ray crystallography and NMR*. New Haven, Connecticut: Yale University.

Bulsink H, Harmsen BJM, Hilbers CW. 1988a. DNA binding properties of gene 5 protein encoded by bacteriophage M13. 1. The kinetics of the dissociation of gene-5-protein–polynucleotide complexes upon addition of salt. *Eur J Biochem 176*:589–596.

Bulsink H, Harmsen BJM, Hilbers CW. 1988b. DNA binding properties of gene 5 protein encoded by bacteriophage M13. 2. Further characterization of the different binding modes for poly- and oligodeoxynucleic acids. *Eur J Biochem 176*:597–608.

Carson M. 1987. Ribbon models of macromolecules. *J Mol Graphics 5*:103–106.

Cuypers T, van der Ouderaa FJ, de Jong WW. 1974. The amino acid sequence of gene 5 protein of bacteriophage M13. *Biochem Biophys Res Commun 59*:557–563.

Dick LR, Sherry AD, Newkirk MM, Gray DM. 1989. ^{13}C NMR of methylated lysines of fd gene 5 protein: Evidence for a conformational change involving lysine 24 upon binding of a negatively charged lanthanide chelate. *Biochemistry 28*:7896–7904.

Folkers PJM, Nilges M, Folmer RHA, Konings RNH, Hilbers CW. 1994. The solution structure of the Tyr 41 → His mutant of the single-stranded DNA binding protein encoded by gene V of the filamentous bacteriophage M13. *J Mol Biol 236*:229–246.

Folkers PJM, van Duynhoven JPM, van Lieshout HTM, Harmsen BJM, van Boom JH, Tesser GI, Konings RNH, Hilbers CW. 1993. Exploring the DNA binding domain of gene V protein encoded by bacteriophage M13 with the aid of spin-labeled oligonucleotides in combination with ^{1}H NMR. *Biochemistry 32*:9407–9416.

Folmer RHA, Nilges M, Folkers PJM, Konings RNH, Hilbers CW. 1994. A model of the complex between single-stranded DNA and the single-stranded DNA binding protein encoded by gene V of filamentous bacteriophage M13. *J Mol Biol 240*:341–357.

Fulford W, Model P. 1988. Bacteriophage f1 DNA replication genes. II. The roles of gene V protein and gene II protein in complementary strand synthesis. *J Mol Biol 203*:39–48.

Gilson MK, Sharp KA, Honig BH. 1987. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J Comput Chem 9*:1276–1278.

Gilson MK, Straatsma TP, McCammon JA, Ripoll DR, Faerman CH, Axelsen PH, Silman I, Sussman JL. 1994. Open "back door" in a molecular dynamics simulation of acetylcholinesterase. *Science 263*:1276–1278.

Gray CW. 1989. Three-dimensional structure of complexes of single-stranded DNA-binding proteins with DNA. Ike and fd gene 5 proteins form left-handed helices with single-stranded DNA. *J Mol Biol 208*:57–64.

Gray DM, Gray CW, Carlson RD. 1982. Neutron scattering data on reconstituted complexes of fd deoxyribonucleic acid and gene 5 protein show that the deoxyribonucleic acid is near the center. *Biochemistry 21*:2702–2713.

Guan Y, Zhang H, Konings RNH, Hilbers CW, Terwilliger TC, Wang AHJ. 1994. Crystal structures of Y41H and Y41F mutants of gene V protein from Ff phage suggest possible protein–protein interactions in the GVP–ssDNA complex. *Biochemistry 33*:7768–7778.

Hutchinson DL, Barnett BL, Bobst AM. 1990. Gene 5 protein–DNA complex: Modeling binding interactions. *J Biomol Struct & Dyn 8*:1–9.

Kansy JW, Clack BA, Gray DM. 1986. The binding of fd gene 5 protein to polynucleotides: Evidence from CD measurements for two binding modes. *J Biomol Struct & Dyn 3*:1079–1109.

King GC, Coleman JE. 1987. Two-dimensional ^{1}H NMR of gene 5 protein

indicates that only two aromatic rings interact significantly with oligodeoxynucleotide bases. *Biochemistry 26*:2929–2937.

King GC, Coleman JE. 1988. The Ff gene 5 protein-d$(pA)_{40-60}$ complex: ^{1}H NMR supports a localized base-binding model. *Biochemistry 27*: 6947–6953.

Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Liu DJ, Day LA. 1994. Pf1 virus structure: Helical coat protein and DNA with paraxial phosphates. *Science 265*:671–674.

Nicholls A. 1993. *GRASP: Graphical representation and analysis of surface properties*. New York: Columbia University.

Nikolov D, Burley SK. 1994. 2.1 Å resolution refined structure of a TATA box-binding protein (TBP). *Nature Struct Biol 1*:621–635.

Pabo CO, Sauer RT. 1992. Transcription factors: Structural families and principles of DNA recognition. *Annu Rev Biochem 61*:1053–1095.

Roberts VA, Freeman HC, Olson AJ, Tainer JA, Getzoff ED. 1991. Electrostatic orientation of the electron-transfer complex between plastocyanin and cytochrome *c*. *J Biol Chem 266*:13431–13441.

Salstrom J, Pratt D. 1971. Role of coliphage M13 gene 5 in single-stranded DNA production. *J Mol Biol 37*:181–200.

Sang BC, Gray DM. 1987. Specificity of the binding of fd gene 5 protein to polydeoxynucleotides. *J Biomol Struct & Dyn 7*:693–706.

Sang BC, Gray DM. 1989. CD measurements show that fd and Ike gene 5 proteins undergo minimal conformational changes upon binding to poly(rA). *Biochemistry 28*:9502–9507.

Skinner MM, Zhang H, Leschnitzer DH, Guan Y, Bellamy H, Sweet RM, Gray CW, Konings RNH, Wang AHJ, Terwilliger TC. 1994. Structure of the gene V protein of bacteriophage f1 determined by multiple wavelength X-ray diffraction of the selenomethionyl protein. *Proc Natl Acad Sci USA 91*:2071–2075.

Stassen APM, Harmsen BJM, Schoenmakers JGG, Hilbers CW, Konings RNH. 1992. Fluorescence studies of the binding of bacteriophage M13 gene V mutant proteins to polynucleotides. *Eur J Biochem 206*:605–612

van Duynhoven JPM, Nooren IMA, Swinkels DW, Folkers PJM, Harmsen BJM, Konings RNH, Hilbers CW. 1993. Exploration of the single-stranded-DNA-binding domains of the gene V proteins encoded by the filamentous bacteriophage Ike and M13 by means of spin-labeled-oligonucleotide and lanthanide–chelate complexes. *Eur J Biochem 216*: 507–517.

Warwicker J, Engelman BP, Steitz TA. 1987. Electrostatic calculations and model-building suggest that DNA bound to CAP is sharply bent. *Proteins Struct Funct Genet 2*:283–289.