Articles

Crystal Structures of Y41H and Y41F Mutants of Gene V Protein from Ff Phage Suggest Possible Protein–Protein Interactions in the GVP–ssDNA Complex[†,‡]

Yue Guan,[§] Hong Zhang,[§] Ruud N. H. Konings,[∥,⊥] Cornelis W. Hilbers,[∥] Thomas C. Terwilliger,[#] and Andrew H.-J. Wang*[,§]

Biophysics Division and Department of Cell and Structural Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, Laboratories of Biophysical Chemistry and Molecular Biology, University of Nijmegen, Nijmegen, The Netherlands, and Life Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received January 10, 1994; Revised Manuscript Received April 14, 1994[⊗]

ABSTRACT: Gene V protein (GVP) encoded by the filamentous phage Ff (M13, f1, fd) is a homodimeric protein of 87 amino acids that binds to single-stranded DNA (ssDNA) nonspecifically and cooperatively. The structure (monoclinic *C*2 form) of the wild-type protein has been determined and refined at 1.8-Å resolution [Skinner et al. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2071–2075]. The monomer structure consists of a somewhat distorted five-stranded β-barrel core with three prominent loops: a DNA-binding loop, a dyad loop, and a dimer contact loop. The amino acid residue at position 41 plays an important role in the dimer–dimer interactions of the protein–ssDNA complex. Two Y41 mutant structures have been studied by X-ray crystallography. The Y41F GVP structure has been refined to an *R*-factor of 0.180 at 2.2-Å resolution and is very similar to the wild-type (wt) structure (rmsd of all C_α atoms = 0.30 Å). In contrast, Y41H GVP forms a new crystal lattice in the space group $P2_12_12_1$ with a = 77.18 Å, b = 84.17 Å, and c = 28.62 Å. Its structure has been solved by the molecular replacement method and refined to an *R*-factor of 0.170 at 2.5-Å resolution. The two monomers of Y41H are crystallographically independent, and their structures remain similar to wt-GVP but with significant differences, particularly in the DNA-binding hairpin region. In both crystals, the loop (residues 36–43) that contains the Y41 residue is involved in the crystal dimer packings but in a different manner. The dimer–dimer contacts found in the wt-GVP crystal may be important for GVP aggregation in the absence of DNA. In the presence of DNA, the dimer–dimer contacts may switch to the type found in the Y41H crystal, allowing the GVP–ssDNA complex to form cooperatively. A model of the complex, consistent with existing biochemical and biophysical data, has been constructed from those crystal packing data.

Gene V protein (GVP)[1] encoded by the Ff filamentous bacteriophages, including M13, f1, and fd, is a single-stranded DNA (ssDNA) binding protein (Alberts et al., 1972; Pratt & Ehrdahl, 1968; Kowalczykowski et al., 1981; Chase & Williams, 1986). GVP is a small protein of 87 amino acids with its amino acid sequence shown in Figure 1. The protein exists as a dimer and binds to DNA in a highly cooperative manner without marked sequence specificity (Fulford & Model, 1988; Gray, 1989; Model et al., 1982; Zaman et al., 1992). Its primary biological role is to bind cooperatively, when a critical threshold protein concentration level has been reached, to the intermediate form of DNA replication produced from the replicative form (RF) of the phage genome and to keep it in the single-stranded form. DNA binding is triggered by the cellular GVP concentration in the host *Escherichia coli*.

Substantial biochemical and biophysical data have been accumulated regarding the interactions between GVP and DNA (Fulford & Model, 1988; Gray, 1989; Model et al., 1982; Zaman et al., 1992; Anderson et al., 1975; Dick et al., 1988; Bulsink et al., 1986, 1988; Kansy et al., 1986). NMR and fluorescence studies have suggested that certain amino acids of GVP are in contact with DNA in the GVP–ssDNA complex (Dick et al., 1988; Stassen et al., 1992a,b). Thus far, the positively charged R16 and R21, the aliphatic L28, and the aromatic Y26 and F73 residues appear to be firmly established as the amino acids involved (Stassen et al., 1992a,b; King & Coleman, 1988). Furthermore, it has been suggested that Y41 is involved in the protein dimer–dimer contacts in

[†] This work was supported by NIH Grants GM-41612 and CA-52054 (A.H.-J.W.) and the Netherlands Organization of Scientific Research (WWO).

[‡] The atomic coordinates of the structures have been deposited in the Brookhaven Protein Data Bank (entry numbers 1YHA and 1YHB).

* To whom correspondence should be addressed.

[§] University of Illinois at Urbana–Champaign.

[∥] Laboratory of Biophysical Chemistry, University of Nijmegen.

[⊥] Laboratory of Molecular Biology, University of Nijmegen.

[#] Los Alamos National Laboratory.

[⊗] Abstract published in *Advance ACS Abstracts,* June 1, 1994.

[1] Abbreviations: GVP, gene V protein; Y41H and Y41F, Tyr41 → His41 and Tyr41 → Phe41 mutant GVP; wt-GVP, wild-type gene V protein; NMR, nuclear magnetic resonance; rmsd, root mean square deviation; SA, simulated annealing. The one-letter and three-letter codes of amino acids are as follows: A and Ala, alanine; C and Cys, cysteine; D and Asp, aspartic acid; E and Glu, glutamic acid; F and Phe, phenylalanine; G and Gly, glycine; H and His, histidine; I and Ile, isoleucine; K and Lys, lysine; L and Leu, leucine; M and Met, methionine; N and Asn, asparigine; P and Pro, proline; Q and Gln, glutamine; R and Arg, arginine; S and Ser, serine; T and Thr, threonine; V and Val, valine; W and Trp, tryptophan; Y and Tyr, tyrosine.

0006-2960/94/0433-7768$04.50/0

```
 1 MIKVE IKPSQ AQFTT RSGVS 20
21 RQGKP YSLNE QLCYV DLGNE 40
41 YPVLV KITLD EGQPA YAPGL 60
61 YTVHL SSFKV GQFGS LMIDR 80
81 LRLVP AK
```

FIGURE 1: Amino acid sequence of GVP (Cuypers et al., 1974).

the GVP–ssDNA complex (King & Coleman, 1988; de Jong et al., 1989a,b; Folkers et al., 1991a; Stassen et al., 1992a,b).

Those solution studies suggested that GVP binds to polymeric DNA and oligonucleotide, respectively, at a ratio of four and three nucleotides per protein monomer [reviewed in Kansy et al. (1986)]. The binding of GVP to single-stranded phage DNA produces a superhelical protein–DNA complex. Electron microscopic studies of the GVP–ssDNA complex revealed that it is likely a left-handed superhelix with a pitch of ~90 Å (Gray, 1989). 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 superhelix (Gray et al., 1982).

To understand fully the role of GVP, it is necessary to have a definitive three-dimensional structure as a framework for the interpretation of various biochemical and biophysical data. Toward this goal, we have recently determined the crystal structure of the wild-type GVP using the multiple-wavelength anomalous diffraction (MAD) method (Skinner et al., 1994). The structure has been refined at 1.8-Å resolution to an R-factor of 0.189. The crystal structure is also fully consistent with the three-dimensional structure deduced from the multidimensional NMR analysis of Y41H GVP (van Duynhoven, 1992; Folkers et al., 1991a,b, 1994). These results have removed any confusion that has been associated with the earlier crystal structural analysis (Brayer & McPherson, 1983).

The crystal structure of the wt-GVP revealed that Y41 participates intimately in the protein–protein contacts in the $C2$ crystal lattice (Skinner et al., 1994). This residue is in a loop (residues 36–43) and is highly exposed to solvent in an isolated GVP dimer. Mutation of the Y41 position has a significant effect on the protein–protein interactions in solution (King & Coleman, 1988; de Jong et al., 1989a,b; Stassen et al., 1992a,b). For example, Y41F and Y41H mutant proteins are significantly more soluble than the wt-GVP. The structural role of Y41 responsible for those modified biophysical properties is not fully understood. To address the role of Y41 in dimer–dimer interactions, we have crystallized two GVP mutants, Y41F and Y41H, and solved their structures at 2.2- and 2.5-Å resolution, respectively. The Y41F mutant crystallized isomorphically to the wt-GVP, whereas the Y41H mutant crystallized in a new crystal lattice. The analyses of these two mutant crystal structures provide possible clues of protein–protein contacts in the absence and in the presence of DNA. The new structural information further allows us to construct a plausible model of the GVP–ssDNA complex.

MATERIALS AND METHODS

Structure Determination and Refinement. The Y41F and Y41H mutant proteins have been prepared as described before (Folkers et al., 1991a) and crystallized from solutions containing 0.5 mM GVP (monomer concentration), 4 mM Tris buffer (pH 7.5), and 15% PEG 4000, equilibrated against 15% PEG 4000 by vapor diffusion at room temperature. Single crystals selected for data collection were mounted in thin-walled glass capillaries and sealed with a droplet of the crystallization mother liquor. One large crystal of Y41F with the shape of an elongated plate having a rhombic cross section ($0.8 \times 0.3 \times 0.2$ mm) was used, while two somewhat smaller crystals of Y41H were needed.

Table 1: Crystal Data and Refinement Parameters for Mutant Gene V Proteins Y41F and Y41H

crystal form	monoclinic	orthorhombic
mutant gene V protein	Y41F	Y41H
water molecules	22	31
space group	$C2$	$P2_12_12_1$
a (Å)	74.97	77.18
b (Å)	27.90	84.17
c (Å)	42.55	28.62
β (deg)	103.09	
V (Å^3)	86707	185690
resolution (Å)	2.2	2.5
no. of reflections [$F > 3.0\sigma(F)$]	3210	3983
completeness (%)	74	85
R^{free} (%)	17.2	16.9
R (8.0–3.0 Å) (%)	16.4	16.9
R (%)	18.0	17.0
rms bond length deviation (Å)	0.017	0.021
rms bond angle deviation (deg)	3.5	4.3
rms dihedral angle deviation (deg)	28.0	28.8
rms improper deviation (deg)	1.7	1.9

The crystal of Y41F was found to be essentially isomorphous to the wt crystal (Table 1). The crystal of Y41H was found in the space group $P2_12_12_1$ with its unit cell volume (V = 185 690 Å^3) nearly twice that of the wt crystal (Table 1). This indicated that the entire Y41H GVP dimer is in the asymmetric unit of the unit cell. The diffraction data set was collected at room temperature on a Rigaku AFC-5R rotating-anode diffractometer, equipped with a copper anode and a graphite monochromator, at a power of 50 kV and 180 mA. An ω-scan mode was used for data collection with Cu Kα radiation (1.5406 Å). Data have been collected to 2.2- and 2.5-Å resolution, respectively, for the Y41F and Y41H crystals. Lorentz polarization, absorption, and decay corrections were applied to obtain the structure factor amplitudes.

Both Y41F and Y41H structures were solved by the molecular replacement method of X-PLOR (Brünger, 1992), using the wt structure as the starting model. Since the Y41F crystal is isomorphous to the wt crystal, the atomic coordinates from the wt-GVP model at early stage of refinement were used as the input for the X-PLOR simulated annealing (SA) refinement for the Y41F structure. The refinement proceeded smoothly and resulted in an R-factor of 0.180 using 3210 reflections [$F_o > 3\sigma(F_o)$] between 8- and 2.2-Å resolution with the inclusion of 22 water molecules.

The structure determination of the Y41H mutant was carried out with an extensive rotation–translation search by X-PLOR. Initially, we assumed the packing of the GVP in the $P2_12_12_1$ unit cell should be closely related to that in the $C2$ unit cell, due to the similarity of the unit cell dimensions ($a \sim a'$, $b \sim 2c'$, and $c \sim b'$). The first rotation search by X-PLOR using 15–4-Å resolution data with the wt dimer as the starting model was unsuccessful with the top 240 possible solutions as judged by the values of their correlation coefficients. A second search with a new starting orientation of the dimer GVP model (a 90° rotation around the a-axis) proved to be successful, with the correct solution having a correlation coefficient of 0.166. The translation search provided the best solution with an R-factor of 0.477. Interestingly, this solution placed the center of gravity of the GVP molecule in the unit cell at $x = 0.440$, $y = 0.131$, and $z = 0.429$, with the molecular 2-fold axis almost perpendicular to the c-axis and tilted 30° relative to the a-axis.

The initial R-factor of the Y41H structure from the best solution is 0.43 after the rigid-body refinement using 8–3-Å

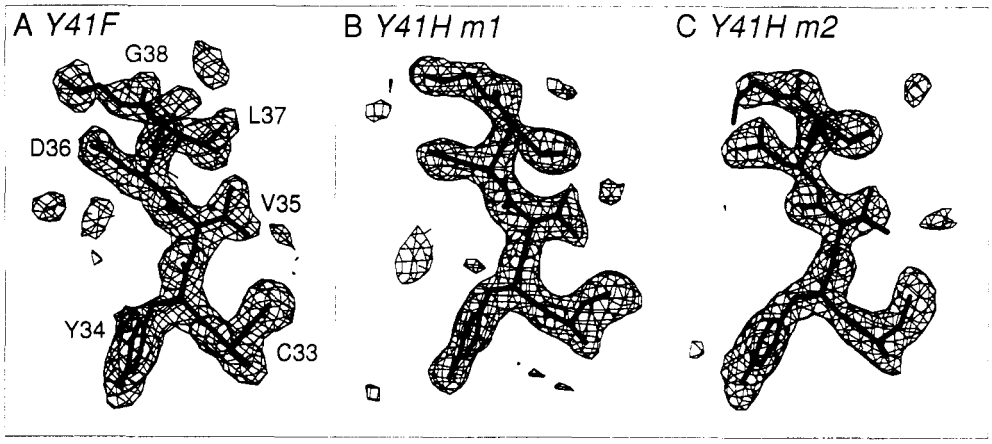

FIGURE 2: Omit $|F_o|-|F_c|$ Fourier electron density maps with residues 33–38 removed for Y41F (A), Y41H monomer 1 (B), and Y41H monomer 2 (C), respectively.

resolution data. A run of 10-ps SA plus the temperature factor refinement reduced the R-factor to 0.207. Nineteen water molecules were located by calculating $(2|F_o| - |F_c|)$ difference Fourier electron density maps using data between 8- and 2.5-Å resolution. Further refinement with the inclusion of those water molecules indicated that some regions of the model required manual rebuilding. A total of 40 omit or SA–omit Fourier maps were calculated. Each map was calculated on the basis of the model with five amino acid residues removed from the model, and some of those "omit models" were refined using simulated annealing refinement. During the rebuilding process, the DNA-binding region (residues 18–26) and the dimer–dimer contact loop region (residues 36–43) of monomer 1 for Y41H were carefully rebuilt using FRODO/TOM (Cambillau & Horjales, 1987; Jones, 1978). The final model for Y41H includes one Y41H dimer and 19 water molecules in the asymmetric unit with an R-factor of 0.170. The reliability of the two structures is reflected in the quality of the omit maps with residues 33–38 removed as shown in Figure 2. The crystal data and refinement summaries are listed in Table 1. The atomic coordinates of the structures have been deposited in the Brookhaven Protein Data Bank.

Model Building of the GVP Superhelix. The model of the GVP–ssDNA protein polymer in the complex was built using the wild-type GVP dimer as a building block. Existing biochemical and biophysical data were considered. In particular, we used the information derived from the electron microscopy (Gray, 1989); i.e., the complex is a left-handed superhelix with about eight GVP dimers per turn of helix (pitch ~90 Å). The diameter of the helix is approximately 80 Å but probably no larger than 100 Å.

As a first step, the GVP dimer was oriented such that the internal 2-fold axis is perpendicular to the superhelical axis and the amino acid residues involved in DNA binding (e.g., R16, R21, Y26, K69) are facing toward the helix axis. Since two types of dimer–dimer interactions were observed in the crystal packing (one for wt and Y41F, the other for Y41H), there are two possible ways of placing the second dimer relative to the first one.

The complex model described later was built using the dimer–dimer contacts found in the Y41H crystal packing, but with the wt molecular structure. The superhelix axis was placed ~20 Å away from the center of gravity of the GVP dimer. The orientation of the dimers was adjusted by rotating the dimer around the internal 2-fold axis so that the dimer–dimer contacts in the Y41H crystal packing could be preserved. The second dimer was then rotated −45° (the minus sign means that a left-handed helix is generated) and translated ~11.2 Å along the superhelical axis direction. In this case the dimer–dimer contacts involved the interactions of the residues from the 36–43 loop (e.g., E40, Y41) of one dimer to the middle of the hairpin of residues 50–70 (e.g., H64 and K69) of the neighboring dimer (Figure 4B). Each additional dimer following the second one was sequentially rotated by −45° and translated by 11.2 Å along the superhelical axis. The model built this way is a left-handed superhelix with eight dimers per superhelical turn and a pitch of ~90 Å per turn. It had an outer diameter of ~80 Å, close to the value deduced from the electron microscopy image (Gray, 1989). 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. Noncrystallographic 2-fold symmetry was preserved for each dimer during the energy minimization.

The model using the dimer–dimer contacts found in wt and Y41F crystal structures was also built. In that case the dimer–dimer contacts involved the interactions of the residues from the 36–43 loop (e.g., E40, Y41) from the two dimers (Figure 4A). In contrast to the first model, the center of gravity of the GVP dimer was placed ~40 Å away from the superhelix axis in order not to create any significant bad contacts. Following a similar procedure described above, the model was constructed, but it resulted in an outer diameter of ~120 Å with a significant hole at the center of the helix when looking down the helix axis. Attempts to make the outer diameter of the helix close to 80 Å inevitably created too many unacceptable, bad van der Waals contacts between proteins. Consequently, this model was rejected.

Protein–protein contact surface area was calculated using the solvent accessibility algorithm of Connolly (Langridge et al., 1981) with a rolling water molecule of 1.4-Å radius.

RESULTS

Structure of the Monomer of the Y41F Protein. The structure of the Y41F protein is very similar to that of the wt-GVP. A least squares fit of the two structures using all C_α atoms produced an rmsd of only 0.30 Å (0.91 Å for all atoms). A distorted five-stranded β-barrel makes up the core of the monomer with two prominently extended β-hairpins (Figure 3A,B). Figure 3C is a topology diagram of the secondary structure of the monomer which contains eight β-strands (numbered I–VIII) of varying length. The five-stranded β-barrel core has an arrangement of β-strands with the order of IV–III–I–V–VIII. The center of the core is made of all hydrophobic aliphatic amino acids, including the buried V4, I6, I47, L49, and L83 residues. Several nearby hydro-

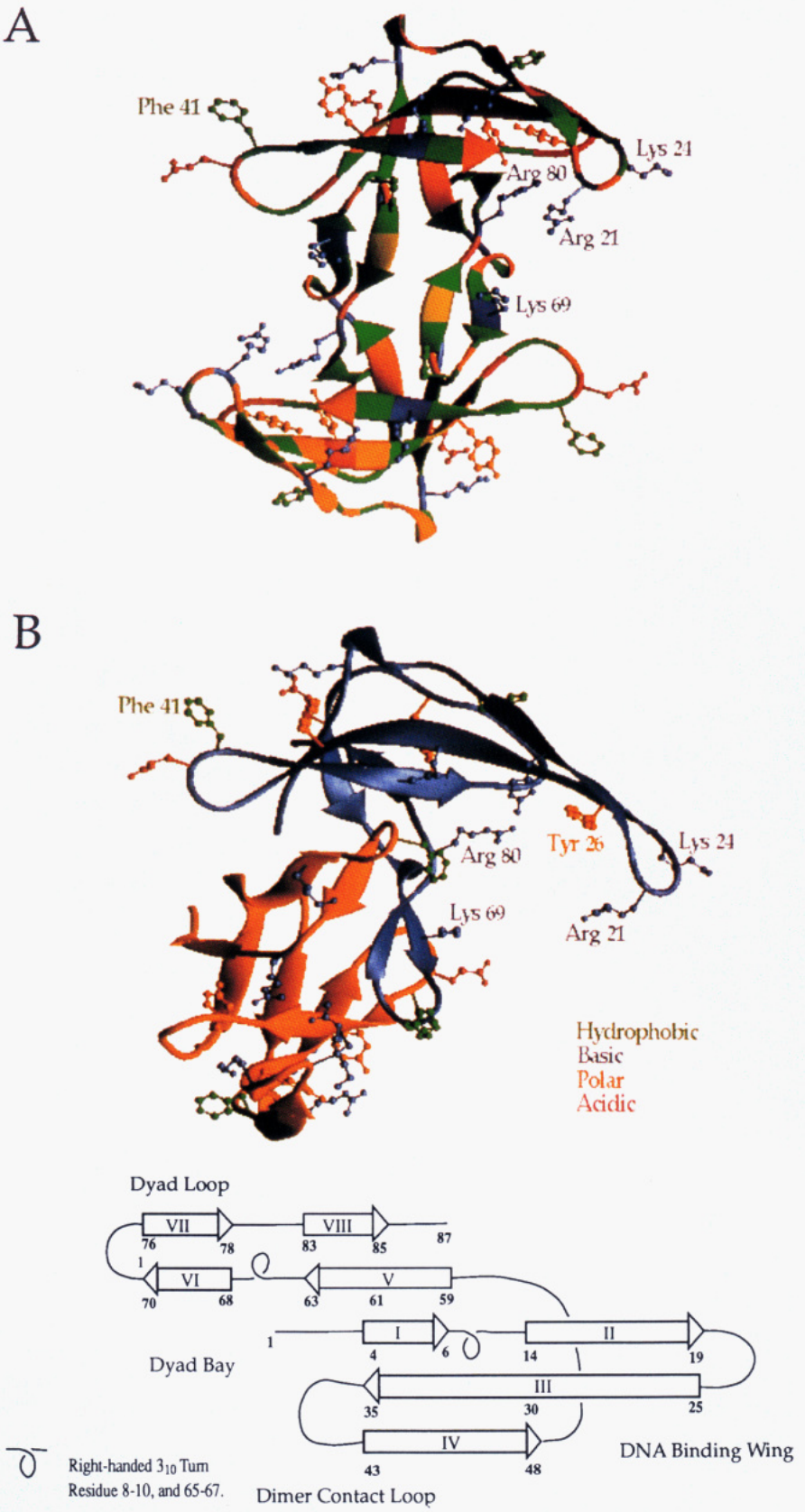

FIGURE 3: (A) Dimer structure of the Y41F GVP. The two monomers are related by a crystallographic 2-fold axis which is perpendicular to the paper. The picture was generated by RIBBONS (Carson, 1987). Some relevant amino acid side chains are shown. (B) Different view of the dimer structure of the Y41F GVP. (C) Topology of the GVP secondary structure. β-Strands are shown as large arrows, with connecting loop structures shown as thin lines. Two 3_{10} turns are seen. Several general regions of the GVP are labeled according their possible functions.

phobic residues, I2, V35, L37, V43, V45 and V63, are exposed in the monomer but become buried upon formation of the dimer structure. The hairpin of residues 13–31 (strands II–III) is implicated in the binding of ssDNA, whereas the stem of the other hairpin (residues 68–78 of strands VI–VII) is involved in the monomer–monomer interactions.

The analysis of the Φ/Ψ backbone torsion angles of the Y41F protein shows that they all fall in the energetically allowable regions with 66 of the Φ/Ψ values centered at $-90°/140°$, a combination associated with a typical β-structure.

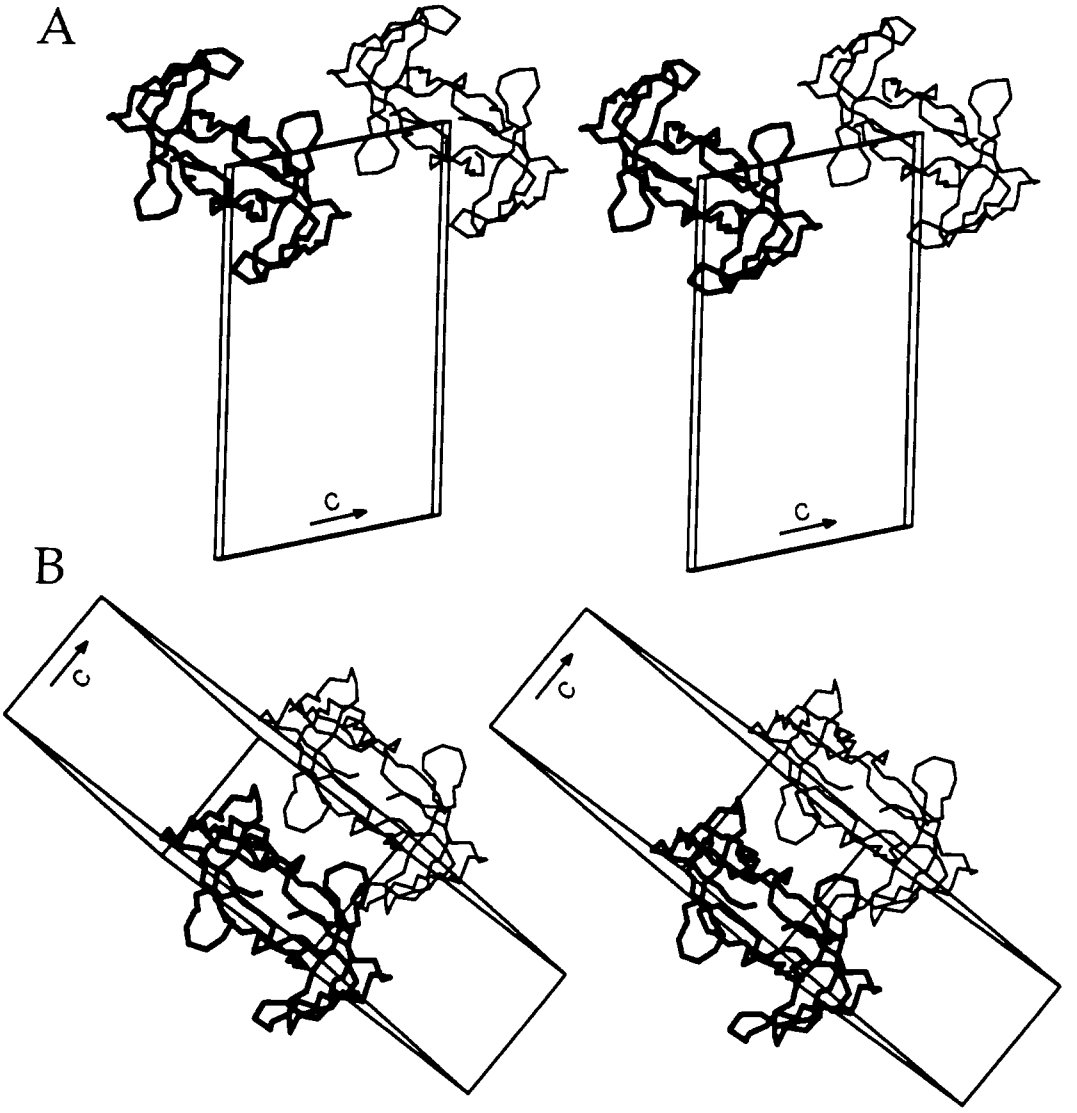

FIGURE 4: Comparison of two crystal packings. (A) The dimer–dimer contact in the monoclinic $C2$ unit cell of the wt and Y41F crystals. The view is along the b-axis which is perpendicular to the paper. The two dimers are related by an unit cell translation along the c-axis direction as well as a crystallographic 2-fold axis along the b-axis located at (0, 0, 0.5), and the distance between them is 42.6 Å. (B) The dimer–dimer contact in the orthorhombic $P2_12_12_1$ unit cell of the Y41H crystal. The two dimers are related by an unit cell translation along the c-axis direction. The distance between them is 28.5 Å.

Thirteen Φ/Ψ values are near $-60°/-10°$, most of which are associated with the two 3_{10} turns at positions 7–10 and 64–67. Four Φ/Ψ values are near $60°/0°$, and they all involve glycines (G23, G38, G52, and G74). The Φ/Ψ values for residues G18 and G59 are near $150°/-170°$ and for residues S17 and C33 are near $-140°/-180°$. In the refined Y41F model, the M1, A86, and K87 residues and the S20–K24 residues of the DNA binding loop are disordered.

The central five-stranded β-barrel fold has recently been recognized as a new protein-folding motif termed OB-fold (Murzin, 1993). Among the family of the OB-fold proteins are *Bacillus subtilis* major cold-shock protein (Schindelin et al., 1993), staphylococcal nuclease (Hynes & Fox, 1991), yeast aspartyl-tRNA synthetase (Ruff et al., 1991), heat-labile enterotoxin from *E. coli* (Sixma et al., 1991), and ribosomal protein S17 (Golden et al., 1993) in addition to the gene V protein.

Structure of the Monomers of the Y41H Protein. The Y41H protein crystallizes in a crystal form different from the wild-type and Y41F proteins. The entire dimer of Y41H protein is in the asymmetric unit of the $P2_12_12_1$ unit cell. While the overall structure remains similar to that of the wt-GVP, some differences are found. Specifically, the ssDNA binding hairpins of residues 13–31 (strands II–III of the two monomers forming the dimer in the asymmetric unit of the crystal) adopt different conformations. In the refined Y41H model, the M1, A86, and K87 residues and the G18–Y26 residues of the DNA-binding loop of monomer 1 are disordered, whereas the DNA-binding loop of monomer 2 is ordered with temperature factors in the range of 25 Å^2, not unlike the overall average temperature factor of 20 Å^2. Another region that has large deviations is in the monomer–monomer contact hairpin (strands VI–VII). The side chains of N39–P42 residues in loop 36–43 are slightly disordered with temperature factors in the range of 20–30 Å^2. The conformational differences are readily seen by the least squares fitting of the three monomers (one from Y41F and two from Y41H) (Figure 1S, supplementary material). The rmsd values (all C_α atoms) between Y41F and Y41H monomers 1 and 2 are 1.34 and 1.84 Å, respectively, whereas that between the two Y41H monomers is 1.50 Å.

The backbone Φ/Ψ angles of Y41H are in general slightly more spread than those of the wt or Y41F GVP, possibly due to its lower resolution (2.5 Å). Nonetheless, we noted that

some changes occurred. For example, the ssDNA binding strands II–III are slightly distorted with residues S17, G18, R21, and L28 of monomer 1 being out of the normal range of Φ/Ψ angles associated with a β-structure.

Structure of the Dimers of Y41F and Y41H Proteins. The biologically active GVP is a dimer as shown in Figure 3A, 3B, which we illustrate using the structure of the Y41F protein. The extended hairpin of strands VI–VII ("dyad loop") from one monomer approaches the 2-fold symmetry-related monomer at the "dyad bay" region near residues V35, L37, V43, and V45 from the β-barrel area and F68, I78, and L81 from the related dyad loop (Figure 3). Extensive hydrophobic interactions are involved in the stabilization of the dimer structure. Most notably, the side chains of V70 and L76 from one monomer are close to each other, and they are in van der Waals contacts with the side chains of V35, L37, V43, V45, I78, and L81 from the 2-fold symmetry-related monomer. Similarly, the side chain of F68 is surrounded by the side chains of L65, V63, L76, and I78 of the same monomer and the side chains of L37, F68, and I78 from the other monomer. The estimated contact surface area between the two mon[illegible] is ~369 Å^2.

In both mutant structures, the formation of the dimer creates a molecule having a near rectangular shape with the size of 25 × 35 × 50 Å plus two extended DNA-binding arms. All three dimer structures (wt, Y41F, and Y41H) are very similar to one another with the largest rmsd between the pair of structures of 2.46 Å (wt vs Y41H). The concave inner surface of the dimer is occupied by amino acids such as R16, R21, K24, Y26, K46, R80, F73′, and K69′ (the prime is from the symmetry-related monomer) that are likely involved in DNA binding. The implications of those amino acids in DNA binding are discussed later.

Dimer–Dimer Contacts in Crystals. The two different crystal forms (space group $C2$ for wt/Y41F GVPs and $P2_12_12_1$ for Y41H GVP) allow us to visualize two types of dimer–dimer packing interactions. In both crystals, the dimer–contact loops (residues 36–43) are involved in the crystal packing (Figure 4). In the wt and Y41F crystals, two adjacent dimers are related by a unit cell translation along the c-axis direction. The centers of the two dimers are separated by 42.6 Å, the length of c-axis, and the dimers are related by a crystallographic 2-fold axis in the b-axis direction (Figure 4A). In the Y41H crystal, two adjacent dimers are also related by an unit cell translation along the c-axis direction, but the centers of the two dimers are now separated by 28.6 Å, the length of the short c-axis (Figure 4B). Note that the relative position of the two symmetry-related 33–43 loops is dramatically different in the two cases. In the Y41H crystal, the top dimer slides by about 29 Å to the left (in going from the wt crystal in Figure 4A to the Y41H crystal in Figure 4B) so that the 36–43 loop is now near the H64 of the neighboring dimer (Figure 5B and Figure 2S, supplementary material).

We have analyzed the detailed dimer–dimer crystal contacts and found that they provided possible clues for the role of Y41 in the dimer–dimer contacts of the GVP–ssDNA complex. Figure 5A shows the manner in which F41 (or Y41 in wt-GVP) of one Y41F dimer interacts with the crystallographically symmetry-related dimer. The aromatic side chain of the F41 (or Y41) residue intercalates into a cavity generated by the Y34, P42, and F41 amino acids. There are no direct hydrogen bonds involving these four amino acids. A similar situation exists for the wt structure, suggesting that the Y41 phenolic OH does not directly participate in dimer–dimer contact through the hydrogen bond. The positively charged side chain of K7 lies between the negatively charged E5 and the neighboring E40, though no direct salt bridge interaction is seen.

In the Y41H crystal, there are two independent contact sites involving loop 36–43 and strands V/VIII near H64 for which the detailed interactions are shown in Figure 5B. Several close contacts (<3.4 Å) are found. Since the two monomers are no longer equivalent, the H41 residue from each monomer has somewhat different interactions. H41′ of monomer 1 in dimer 2 is close to Y56 of an adjacent dimer, and N39′ is close to S66. In contrast, E40 and H41 of monomer 2 in dimer 1 are close to V84′ and P85′ of the neighboring dimer (Table 2).

A Model of GVP in the Complex. We have used the crystal packing interactions as a starting point in exploring the possible dimer–dimer protein contacts in the GVP–ssDNA superhelical complex. Studies of electron microscopy on the complex suggested that it is 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). In addition, neutron scattering studies suggested that the DNA resides inside of the helix, not wrapping around the outside of the helix (Gray et al., 1982).

Initially, we examined the packing interactions of the wt-GVP crystal shown in Figure 4A. Using the procedure described in the Materials and Methods section, we were able to easily construct a left-handed superhelical GVP protein shell, which satisfies most of the known biochemical and biophysical data. For example, the left-handed GVP superhelix has eight dimers per turn of helix with a pitch of 90 Å, and the inner side of the helix has sufficient room to accommodate two ssDNA strands running in opposite directions. However, this helix can only have a diameter greater than 100 Å. When a helix with a diameter smaller than 100 Å is constructed, there are too many bad contacts between dimers, making the model unacceptable.

In contrast, when the packing interactions of the Y41H–GVP crystal shown in Figure 4B were used, a different but more plausible model (with an 80-Å diameter) could be easily constructed without any bad contact. Figure 6 shows two views of this model. No DNA has been included. To simplify the visualization of the GVP superhelix, a schematic of the superhelix is depicted with the GVP represented as elongated plates with two arms (Figure 7). This model is significantly different from those proposed earlier on the basis of an earlier crystal structure (Hutchinson et al., 1990; Brayer & McPherson, 1984) and is similar to the one we recently constructed on the basis of packing considerations of the (now correct) 3D structure (Skinner et al., 1994).

Figure 8 depicts a pair of dimers from the superhelix model viewing from the inside of the helix where DNA resides. There are 1257 contacts (<6 Å) between adjacent dimers, some of which are possible hydrogen bonds, listed in Table 2. Figure 9 is an approximate "cylindrical projection" of this dimer pair. By examining this figure, we note that the following amino acids are along the possible pathway of the DNA trail from left to right: F73′, K46, R16, L28, R80, Y26, K24, and finally R21. The distance between two helically related amino acid side chains in the inner wall of the superhelix is about 17 Å. For example, the distance from one F73′ to the next F73′ is 17 Å, easily spanned by three or four nucleotides. It should be pointed out that we do not know the polarity of the DNA molecule in the complex, nor do we know the exact contact between DNA and protein. Though one may speculate that positively charged amino acids interact with the backbone

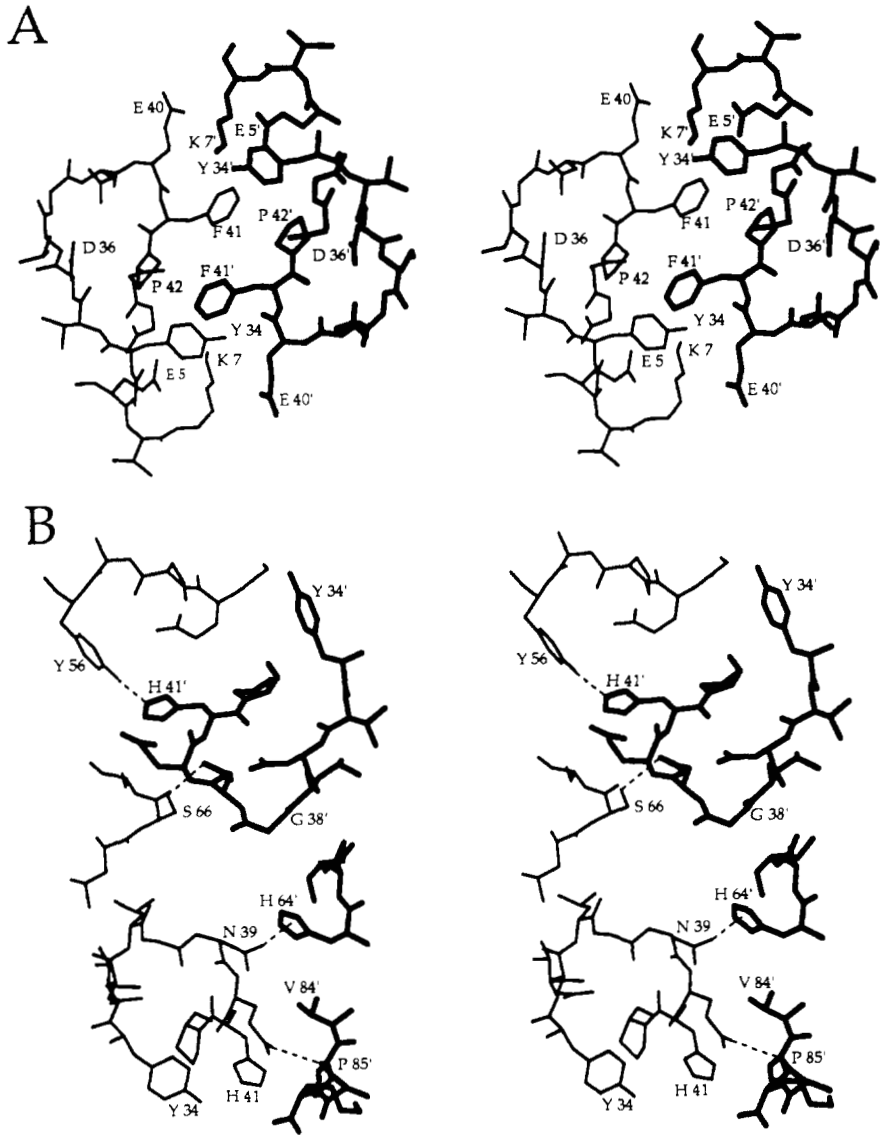

FIGURE 5: A detailed drawing showing the local interactions of dimer–dimer crystal contacts: (A) Y41F crystal; (B) Y41H crystal.

phosphates and the aromatic Y26 and F73′ residues (which are only ~4 Å apart) stack with two consecutive bases along the ssDNA, it remains to be determined experimentally. Furthermore, the DNA-binding arms are quite flexible, as reflected by their multiple conformations observed in crystals. It is likely that some movements of the DNA-binding arms are required upon the binding to DNA.

DISCUSSION

The structure determinations of the wt (Skinner et al., 1994), Y41F, and Y41H GVP proteins afford us a detailed view on a member of an important family of ssDNA binding proteins. While an earlier interpretation of the crystal structure of fd GVP has been published (Brayer & McPherson, 1983), recent NMR studies suggested that the amino acid assignment of that model could not be correct. The newly refined GVP crystal structures are now completely consistent not only with the published secondary structure assignment from NMR studies (Folkers et al., 1991a,b) but also with the three-dimensional structure determination of Y41H by NMR (Folkers et al., 1994).

The new structure also explains the possible role of several amino acids among different ssDNA binding proteins, in addition to those involved in DNA binding (e.g., Y26, F73, and R80). Many of the glycines (G23, G38, G52, G71, G74) and prolines (P25, P42, P58) occupy at or near the corners of loops. Y61 is important as mutation at this position seemed to affect the activity of GVP significantly (Folkers et al., 1991a,b). It turns out that the aromatic side chain of Y61 is forming two hydrogen bonds through its OH group to the peptide N and O of A57, thereby stabilizing the loop structure of residues 56–62 (Figure 3). This type of loop stabilization by the tyrosine residue appears to be a commonly occurring motif in protein structure (Richardson and Richardson, personal communication).

Some of the existing biochemical and biophysical data on GVP may be explained by the present three-dimensional structures of three GVPs (wt, Y41F, and Y41H). In particular, Y41 has been shown to be important in the GVP dimer–dimer interactions (King & Coleman, 1988; de Jong et al., 1989a,b; Folkers et al., 1991a). The wt-GVP is not very soluble in solution, presumably due to the facile aggregation of the protein. Mutation of Y41 to histidine or glycine leads to mutant proteins with significantly improved solubility properties. In fact, the NMR structural analyses have been performed using the more soluble Y41H GVP (Folkers et al., 1991a,b, 1994). Most of the Y41 mutant

Table 2: Dimer–Dimer Hydrogen-Bonding Interactions in the Crystal Packing of Y41H Mutant GVP and in the Models of the GVP–DNA Complex

Y41H mutant GVP (crystal packing)		GVP superhelix (starting model)		GVP superhelix (energy minimized)	
interactions	*d* (Å)	interactions	*d* (Å)	interactions	*d* (Å)
Y56OH–H41NE2	2.55	E51O–L44O	3.02		
(1.1)[a] (2.1)		(1.1) (2.1)			
S66OG–N39OD1	3.34	E51OE1–K46NZ	3.05	E51OE1–K46NZ	2.58
(1.1) (2.1)		(1.1) (2.1)		(1.1) (2.1)	
N39OD1–H64ND1	3.10	E51OE1–F730	3.21		
(1.2) (2.2)		(1.1) (2.2)			
E40OE1–P85O	3.23	E51OE1–G74N	3.25		
(1.2) (2.2)		(1.1) (2.2)			
		E51OE1–G74O	2.59		
		(1.1) (2.2)			
		Q53NE2–Y41OH	3.36	Q53NE2–Y41OH	3.36
		(1.1) (2.1)		(1.1) (2.1)	
		H64ND1–E40OE1	3.11		
		(1.1) (2.1)			
		H64NE2–E40OE1	2.07	H64NE2–E40OE1	3.04
		(1.1) (2.1)		(1.1) (2.1)	
		H64NE2–E40OE2	3.25	H64NE2–E40OE2	2.79
		(1.1) (2.1)		(1.1) (2.1)	
		K69NZ–D79OD2	3.26	K69NZ–D79OD2	2.52
		(1.1) (2.2)		(1.1) (2.2)	
				K69NZ–D69OD1	3.25
				(1.1) (2.2)	

[a] (1.1), (1.2), (2.1), and (2.2) represent monomer 1 in dimer 1, monomer 2 in dimer 1, monomer 1 in dimer 2, and monomer 2 in dimer 2, respectively.

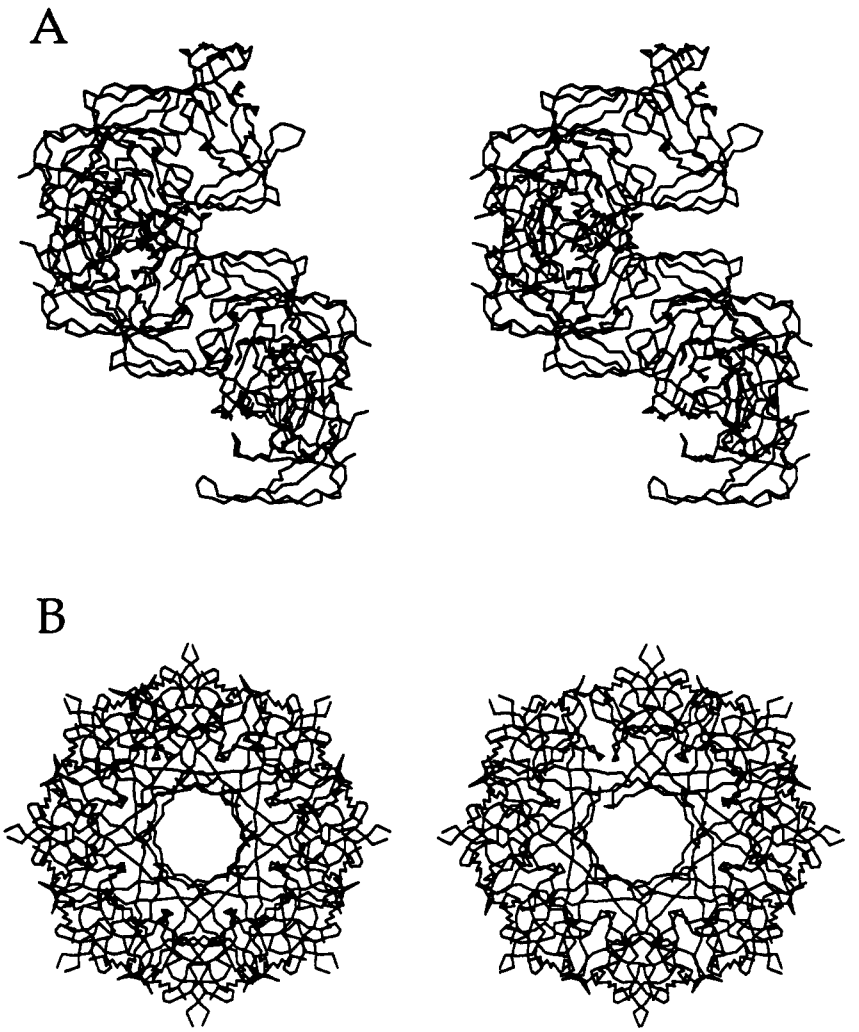

FIGURE 6: Stereoscopic diagrams of one turn (eight dimers) of the left-handed GVP superhelix. The pitch and the diameter of the helix are 90 and 80 Å, respectively. Panels: (A) side view; (B) end view.

proteins, including those studied in this work (Y41F and Y41H), retain cooperativity toward ssDNA binding to a varying degree (King & Coleman, 1988; Stassen et al., 1992b; T. C. Terwilliger, unpublished observations).

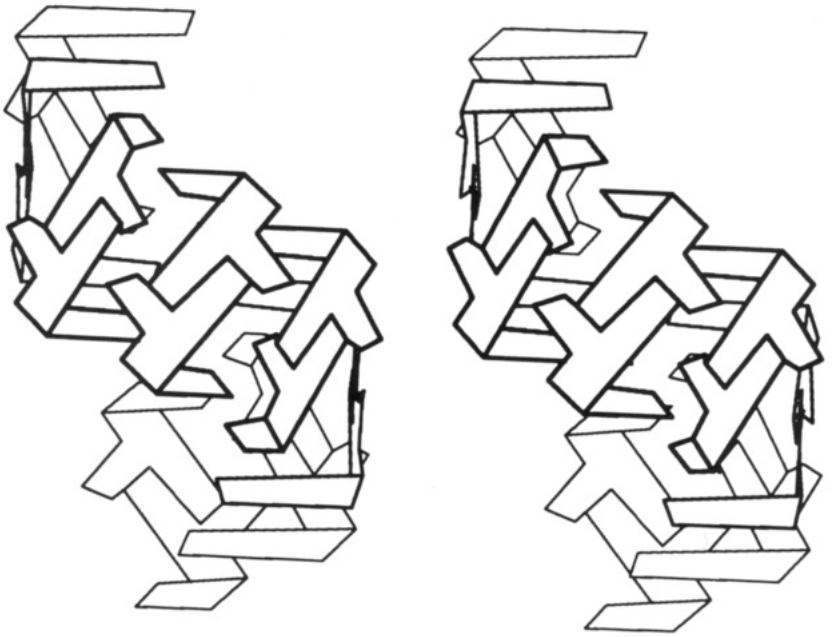

FIGURE 7: Schematic of the left-handed GVP superhelix. GVP is represented as elongated motifs which roughly resemble the shape of the GVP dimer.

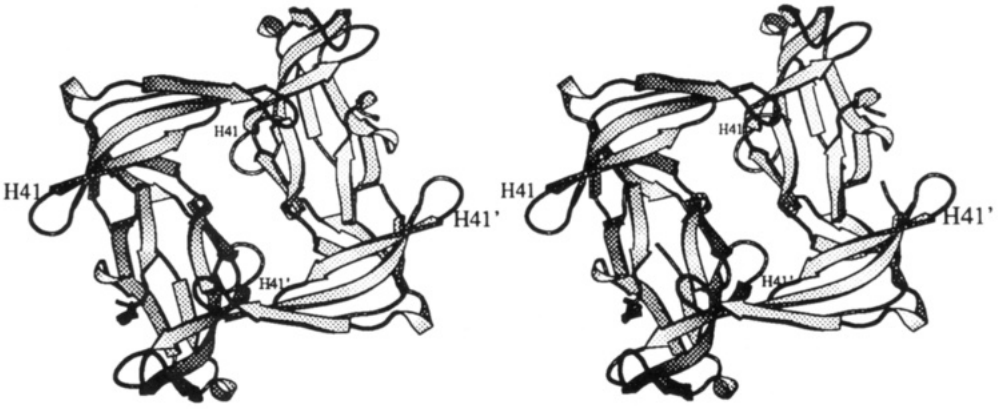

FIGURE 8: A Molscript (Kraulis, 1991) drawing showing the local interactions of the dimer–dimer contacts in the GVP superhelix.

Taken together, the information obtained here and previous data by others, we offer a plausible scenario for understanding the DNA-binding property of GVP. First, the crystal packing interactions for the wt and Y41F GVPs suggest that the insertion of the aromatic side chains of Y41/F41 into a hydrophobic pocket surrounded by Y34, P42, and Y41 amino acids from another dimer GVP plays an important role of the dimer–dimer aggregation of GVP in solution in the *absence* of DNA. When sufficient copies of DNA are present and the critical level of GVP concentration inside the cell has been achieved, GVP dimers bind to DNA and align themselves along the DNA molecule cooperatively. The contacts between dimers now switch from those seen in (or closely related to) the wt crystal packing (Figure 5A) to those in (or closely related to) the Y41H crystal packing (Figure 5B). The switching of the protein contact is facilitated by the somewhat larger interdimer contact surface (172 Å^2) derived from the Y41H contact mode and the newly created protein–DNA interactions. This model does not require any significant conformational changes in GVP in order to have a cooperative binding to ssDNA.

The present model (which is derived from crystal contact information) is very similar to the one in Skinner et al. (1994). The two models differ only in slight rotation about the 2-fold axis relating two adjacent dimers. We note that in our GVP superhelix models Y41 points away from the next dimer over. It is likely that Y41 is in a different configuration in the actual GVP–DNA complex, bringing it closer, perhaps, to H64″ and E40. Interestingly, suppression mutations of Y41F involve many amino acids either in the DNA-binding channel (L28V and F73W) or near the putative dimer–dimer contacts (G71S, G71N, S75C, M77A, and M77I) (T. C. Terwilliger, unpublished results).

The current model of the protein shell of the GVP–ssDNA complex (Figures 6 and 7) does not incorporate DNA. Although at present some information regarding the involvement of certain amino acids in DNA binding is available (e.g., R16, R21, Y26, L28, and F73′ have been implicated), it is not possible to unambiguously build the DNA molecule into the protein shell. Thus far, the only example of a known 3D structure of the polymeric single-stranded nucleic acid–protein complex is that of the tobacco mosaic virus (TMV) (Stubbs et al., 1977) in which a trinucleotide RNA per protein monomer binding mode is found. The interactions appeared to be between positively charged amino acids of the proteins and the negatively charged phosphate of RNA. However, a recent study of a single-stranded DNA fragment $(dT)_4$ bound to the Klenow fragment of DNA polymerase I showed that other types of amino acids, in addition of the positively charged

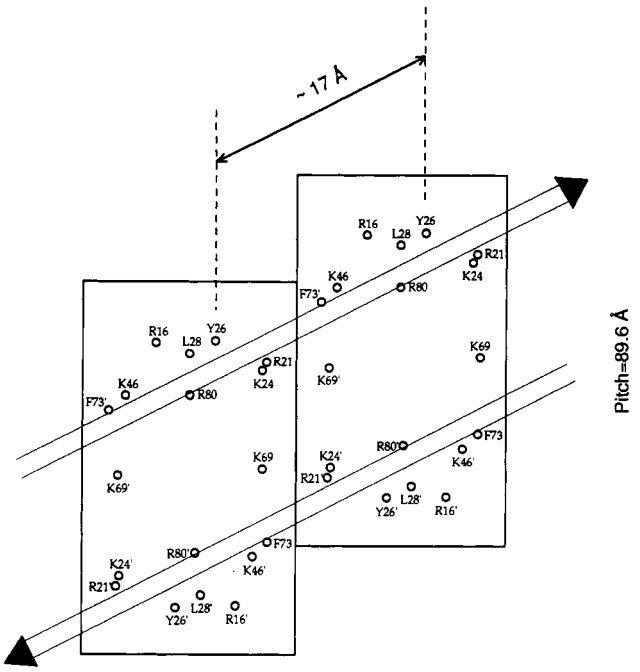

FIGURE 9: Schematic diagram depicting an approximate "cyclindrical projection" of a pair of dimers of the left-handed GVP superhelix. Two long arrows represent two antiparallel ssDNA in the interior of the superhelix. The polarity and the location of DNA are somewhat arbitrary. Note that the distance between two equivalent amino acids is around 17 Å, which can accommodate a tri- or tetranucleotide easily. The distribution of those amino acids which are likely involved in DNA binding is shown on the projection.

ones, are involved in the binding interactions (Beese & Steitz, 1991).

The elucidation of the exact mode of interactions would require the structure determination of a complex between GVP and certain DNA oligonucleotides of defined sequence. In the presence of oligonucleotides, some kinds of aggregate of GVP–oligonucleotide are formed, as evident by the broadening of the NMR resonances (van Duynhoven, 1992; van Duynhoven et al., 1993) and the large asymmetric volume of the GVP–oligonucleotide crystals (Fitzgerald et al., 1979). Inspection of Figure 9 suggests that the current model of the GVP superhelix does not describe the contiguous DNA-binding mode unequivocally. It is likely that some intermolecular rearrangements among dimers are required for the binding of polymer DNA versus oligomer DNA, resulting in a different DNA binding ratio of $n = 4$ for polynucleotides and $n = 3$ for oligonucleotides.

It should be pointed out that the model shown in Figure 7 is one of the few possible alternatives. A slightly different model has been constructed using a search modeling method on the basis of possible minimum bad contacts between neighboring dimers (Skinner et al., 1994). The difference between those two models [Figure 7 of this paper and Figure 3 of Skinner et al. (1994)] is the relative position of the two contact loops (residues 36–43). So far, there are no other definitive data to discriminate those two models, and both models can explain existing data equally well. It is worthy to note that the GVP–ssDNA complex is somewhat flexible as evident by the EM images (Gray, 1989). Therefore, it would not be surprising that some degree of variability of the dimer–dimer contacts is in fact desirable.

In conclusion, we have shown that the Y41 residue of GVP is intimately involved in the dimer–dimer interactions as revealed by two types of crystal packings. Analyses of those packing interactions further our understanding on how GVP binds to ssDNA and on a possible switching mechanism of GVP in going from the free state to the (DNA) bound state. The results from this work should be valuable in understanding the biological roles of GVP *in vivo* and in guiding us for the preparation of new GVP mutants.

SUPPLEMENTARY MATERIAL AVAILABLE

Two figures (1S and 2S) showing the least squares fitting of the three monomers (one from Y41F and two from Y41H) and the superposition of two different crystal packings found in Y41H and Y41F mutants (2 pages). Ordering information is given on any current masthead page.

REFERENCES

Alberts, B., Frey, L., & Delius, H. (1972) *J. Mol. Biol. 68*, 139–152.
Alma, N. C. M., Harmsen, B. J. M., de Jong, E. A. M., van de Ven, J., & Hilbers, C. W. (1983) *J. Mol. Biol. 163*, 47–62.
Anderson, R. A., Nakashima, Y., & Coleman, J. E. (1975) *Biochemistry 14*, 907–917.
Baas, P. D. (1985) *Biochim. Biophys. Acta 825*, 111–139.
Beese, L. S., & Steitz, T. A. (1991) *EMBO J. 10*, 25–33.
Brayer, G. D., & McPherson, A. (1983) *J. Mol. Biol. 169*, 565–596.
Brayer, G. D., & McPherson, A. (1984) *J. Biomol. Struct. Dyn. 2*, 495–509.
Brünger, A. T. (1992) X-PLOR 3.1, Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 6666, New Haven, CT 06511.
Bulsink, H., van Resandt, R. W. W., Harmsen, B. J. M., & Hilbers, C. W. (1986) *Eur. J. Biochem. 157*, 329–334.
Bulsink, H., Harmsen, B. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 176*, 597–608.
Cambillau, C., & Horjales, E. (1987) *J. Mol. Graphics 5*, 174–177.
Carson, M. (1987) *J. Mol. Graphics 5*, 103–106.
Chase, J. W., & Williams, K. R. (1986) *Annu. Rev. Biochem. 55*, 103–136.
Cuypers, T., van der Ouderaa, F. J., & de Jong, W. W. (1974) *Biochem. Biophys. Res. Commun. 59*, 557–563.
de Jong, E. A. M., van Duynhoven, J. P. M., Harmsen, B. J. M., Konings, R. N. H. & Hilbers, C. W. (1989a) *J. Mol. Biol. 206*, 119–132.
de Jong, E. A. M., van Duynhoven, J. P. M., Harmsen, B. J. M., Tesser, G. I., Konings, R. N. H., & Hilbers, C. W. (1989b) *J. Mol. Biol. 206*, 133–152.
Dick, L. R., Sherry, A. D., Newkirk, M. M., & Gray, D. M. (1988) *J. Biol. Chem. 263*, 18864–18872.
Fitzgerald, P. M. D., Wang, A. H.-J., McPherson, A., Jurnak, F. A., Molineux, I., Kolpak, F., & Rich, A. (1979) *J. Supramol. Struct. 10*, 479–489.
Folkers, P. J. M., Stassen, A. P. M., van Duynhoven, J. P. M., Harmsen, B. J. M., Konings, R. N. H., & Hilbers, C. W. (1991a) *Eur. J. Biochem. 200*, 139–148.
Folkers, P. J. M., van Duynhoven, J. P. M., Jonker, A. J., Harmsen, B. J. M., Konings, R. N. H., & Hilbers, C. W. (1991b) *Eur. J. Biochem. 202*, 349–360.
Folkers, P. J. M., van Duynhoven, J. P. M., van Lieshout, H. T. M., Harmsen, B. J. M., van Boom, J. H., Tesser, G. I., Konings, R. N. H., & Hilbers, C. W. (1993) *Biochemistry 32*, 9407–9416.
Folkers, P. J. M., Nilges, M., Folmer, R. H. A., Konings, R. N. H., & Hilbers, C. W. (1994) *J. Mol. Biol.* (in press).
Fulford, W., & Model, P. (1988) *J. Mol. Biol. 203*, 39–48.
Golden, B. L., Hoffman, D. W., Ramakrishnan, V., & White, S. W. (1993) *Biochemistry 32*, 12812–12820.

Gray, C. W. (1989) *J. Mol. Biol. 208*, 57–64.
Gray, D. M., Gray, C. W., & Carlson, R. D. (1982) *Biochemistry 21*, 2702–2713.
Hutchinson, D. L., Barnett, B. L., & Bobst, A. M. (1990) *J. Biomol. Struct. Dyn. 8*, 1–9.
Hynes, T. R., & Fox, R. O. (1991) *Proteins 10*, 92–105.
Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.
Kansy, J. W., Clack, B. A., & Gray, D. M. (1986) *J. Biomol. Struct. Dyn. 3*, 1079–1109.
King, G. C., & Coleman, J. E. (1988) *Biochemistry 27*, 6947–6953.
Kowalczykowski, S. C., Bear, D. G., & von Hippel, P. H. (1981) in *The Enzymes* (Boyer, P. D., Ed.) 3rd ed., Vol. 14, pp 373–444, Academic Press, New York.
Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946–950.
Langridge, R., Ferrin, T. E., Kuntz, I. D., & Connolly, M. L. (1981) *Science 211*, 661–666.
Model, P., McGill, C., Mazur, B., & Fulford, W. D. (1982) *Cell 29*, 329–335.
Murzin, A. G. (1993) *EMBO J. 12*, 861–867.
Pratt, D., & Ehrdahl, W. S. (1968) *J. Mol. Biol. 37*, 181–200.
Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitshler, A., Podjarny, A., Rees, B., Thierry, J. C., & Moras, D. (1991) *Science 252*, 1682–1689.
Schindelin, H., Marahiel, M. A., & Heinemann, U. (1993) *Nature 364*, 164–168.
Sixma, T. K., Pronk, S. E., Kalk, K. H., Wartna, E. S., van Zanten, B. A. M., Witholt, B., & Hol, W. G. J. (1991) *Nature 351*, 371–377.
Skinner, M. M., Zhang, H., Leschnitzer, D. H., Guan, Y., Bellamy, H., Sweet, R. M., Gray, C. W., Konings, R. N. H., Wang, A. H.-J., & Terwilliger, T. C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2071–2075.
Stassen, A. P. M., Zaman, G. J. R., van Deursen, J. M. A., Schoenmakers, J. G. G., & Konings, R. N. H. (1992a) *Eur. J. Biochem. 204*, 1003–1014.
Stassen, A. P. M., Harmsen, B. J. M., Schoenmakers, J. G. G., Hilbers, C. W., & Konings, R. N. H. (1992b) *Eur. J. Biochem. 206*, 605–612.
Stubbs, G., Warren, S., & Holmes, K. C. (1977) *Nature 267*, 216–221.
van Duynhoven, J. P. M. (1992) Ph.D. Thesis, University of Nijmegen, Nijmegen, The Netherlands.
van Duynhoven, J. P. M., Nooren, I. M. A., Swinkels, D. W., Folkers, P. J. M., Harmsen, B. J. M., Konings, R. N. H., & Hilbers, C. W. (1993) *Eur. J. Biochem. 216*, 507–517.
Zabin, H. B., & Terwilliger, T. C. (1991) *J. Mol. Biol. 219*, 257–275.
Zabin, H. B., Horvath, M. P., & Terwilliger, T. C. (1991) *Biochemistry 30*, 6230–6240.
Zaman, G. J. R., Kaan, A. M., Schoenmakers, J. G. G., & Konings, R. N. H. (1992) *J. Bacteriol. 174*, 595–600.