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

P. J. M. Folkers,[‡] J. P. M. van Duynhoven,[‡] H. T. M. van Lieshout,[‡] B. J. M. Harmsen,[‡] J. H. van Boom,[§] G. I. Tesser,[∥] R. N. H. Konings,[‡] and C. W. Hilbers*[,‡]

NSR Center (Nijmegen Son Research Center), Laboratory of Biophysical Chemistry, and Laboratory of Organic Chemistry, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands, and Department of Organic Chemistry, Gorlaeus Laboratories, University of Leiden, 2333 RA Leiden, The Netherlands

Received February 24, 1993; Revised Manuscript Received June 18, 1993[⊗]

ABSTRACT: The DNA binding domain of the single-stranded DNA binding protein gene V protein encoded by the bacteriophage M13 was studied by means of ^{1}H nuclear magnetic resonance, through use of a spin-labeled deoxytrinucleotide. The paramagnetic relaxation effects observed in the ^{1}H-NMR spectrum of M13 GVP upon binding of the spin-labeled ligand were made manifest by means of 2D difference spectroscopy. In this way, a vast data reduction was accomplished which enabled us to check and extend the analysis of the 2D spectra carried out previously as well as to probe the DNA binding domain and its surroundings. The DNA binding domain is principally situated on two β-loops. The major loop of the two is the so-called DNA binding loop (residues 16–28) of the protein where the residues which constitute one side of the β-ladder (in particular, residues Ser20, Tyr26, and Leu28) are closest to the DNA spin-label. The other loop is part of the so-called dyad domain of the protein (residues 68–78), and mainly its residues at the tip are affected by the spin-label (in particular, Phe73). In addition, a part of the so-called complex domain of the protein (residues 44–51) which runs contiguous to the DNA binding loop is in close vicinity to the DNA. The NMR data imply that the DNA binding domain is divided over two monomeric units of the GVP dimer in which the DNA binding loop and the tip of the dyad loop are part of opposite monomers. The view of the GVP–ssDNA binding interaction which emerges from our data differs from previous molecular modeling proposals which were based on the GVP crystal structure (Brayer & McPherson, 1984; Hutchinson et al., 1990). These models implicate the involvement of one or two tyrosines (Tyr34, Tyr41) of the complex loop of the protein to participate in complex formation with ssDNA. In the NMR studies with the spin-labeled oligonucleotides, no indication of such interactions has been found. Other differences between the models and our NMR data are related to the structural differences found when solution and crystal structures are compared.

During the last decades, our insight into cellular processes, such as DNA replication, transcription, and translation, has been significantly broadened. A considerable amount of biochemical research has been carried out in order to study the specific functions of key proteins involved in these processes, the DNA binding proteins. Although the general function of these proteins is known, the detailed molecular mechanisms underlying the regulation of a variety of processes at the DNA level remain to be elucidated for a large number of systems. However, progress has been made in the understanding of the recognition of specific double-stranded DNA (dsDNA)[1] sequences by DNA binding proteins. Several classes of these proteins have been identified on the basis of common functional domains which are involved in the recognition of dsDNA [for a review, see Freemont et al. (1991)]. The knowledge of the DNA binding proteins that recognize single-stranded DNA has not advanced that far. Very recently, the structure of the Rec A protein of *Escherichia coli* which binds to ssDNA as well as to dsDNA has been determined (Story et al., 1992). Of the class of proteins which solely bind to ssDNA, only the crystal structure of one of its members, namely, the gene V protein (GVP) of the filamentous bacteriophage Ff (M13, f1, fd), is known (Brayer & McPherson, 1983). The refined model of this dimeric protein reveals that the individual monomers are entirely composed of β-structure and that they are closely associated about a dyad axis. On the basis of the crystal structure and several physicochemical observations, models describing the complex formation of ssDNA with GVP have been proposed (Brayer & McPherson, 1984; Hutchinson et al., 1990).

Recently, however, NMR studies have indicated essential differences between the solution and crystal structures of GVP (van Duynhoven et al., 1990; Folkers et al., 1991b). Furthermore, in several papers, physicochemical observations contrasting the proposed ssDNA–GVP interaction models have been reported (King & Coleman, 1987, 1988; Folkers et al., 1991a). Therefore, we have embarked on studies concerning the structure of GVP and its interaction with ssDNA in solution. This paper reports of NMR experiments aimed at

* Correspondence should be addressed to the author at NSR Center, Laboratory of Biophysical Chemistry, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands. Phone: **31.80.652678. FAX: **31.80.652112.

‡ Laboratory of Biophysical Chemistry, University of Nijmegen.

§ Department of Organic Chemistry, University of Leiden.

∥ Laboratory of Organic Chemistry, University of Nijmegen.

⊗ Abstract published in *Advance ACS Abstracts,* August 15, 1993.

[1] Abbreviations: GVP, gene V protein; GVP Y41H, GVP with Tyr41 replaced by His; ssDNA, single-stranded DNA; dsDNA, double-stranded DNA; NMR, nuclear magnetic resonance; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser enhancement spectroscopy; TPPI, time-proportional phase incrementation; 1D, one-dimensional; 2D, two-dimensional.

0006-2960/93/0432-9407$04.00/0

extending our knowledge about the interaction of Ff GVP with ssDNA at the submolecular level. This has been achieved with the aid of spin-labeled oligonucleotides in combination with 1D- and 2D-NMR techniques (de Jong et al., 1988, 1989). The method involves the evaluation of "relative" and "absolute" difference 2D spectra, computed from the 2D spectra of GVP recorded in the presence and absence of (small amounts of bound) spin-labeled oligonucleotides, utilizing both the NOESY and the TOCSY technique (de Jong et al., 1988, 1989). We have predominantly used the "solubility" mutant M13 GVP Y41H (Folkers et al., 1991a; Stassen et al., 1992a) for this purpose, of which high-quality NMR spectra could be obtained and of which a majority of the proton resonances have been assigned (Folkers et al., 1991b). It is shown that these assignments could easily be checked as far as the DNA binding domain and its surroundings was concerned by means of 2D NOESY difference spectra recorded in H_2O.

We present here a complete analysis of the difference 2D spectra and a description of the DNA binding domain and its surrounding based on the solution structure of GVP. The results will be discussed in light of the proposed interaction models and other relevant physicochemical data.

MATERIALS AND METHODS

Isolation and Sample Preparation. Gene V of the bacteriophage M13 was cloned and expressed in *Escherichia coli* under the control of the arabinose-inducible promoter of *Salmonella typhimurium* (Stassen et al., 1992a). Mutant GVP was constructed by means of chemical mutagenesis techniques (Stassen et al., 1992a). The mutant GVP Y41H, which was investigated in the 2D-NMR studies described, contained a histidine instead of a tyrosine residue at position 41 of wild-type M13 GVP. Mutant GVP Y41H and wild-type GVP were prepared and isolated as described previously (Garssen et al., 1977; van Duynhoven et al., 1990). After dialysis against 1 mM cacodylate buffer (pH 6.9) containing 50 mM NaCl, the purified GVP was lyophilized and stored at −20 °C. Samples of M13 GVP, dissolved either in 99.8% D_2O or in 90% H_2O/10% D_2O, contained a few millimolar of NaCl and no buffer components.

Spin-Labeled Oligonucleotides. The spin-label trinucleotide, three adenyl residues to which the spin-label 4-hydroxy-2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO) was covalently attached via phosphodiester bonds at both the-5′ and 3′ ends of the oligonucleotide, was synthesized, purified, and characterized as described by Claesen et al. (1986). We will refer to this spin-label as *(dA)$_3$*, where the asterisk refers to the attached spin-label. The concentration of *d(A)$_3$* in H_2O (at pH 7) was determined from its absorption at 260 nm, using the absorption coefficient for the adenine trimer (Cassani & Bollum, 1969) corrected for a slight extra absorption by the TEMPO moiety, i.e., 37 500 M^{-1} cm^{-1}.

The oligonucleotide d(A)$_{16}$ was synthesized using the phosphotriester method (Arentzen et al., 1979). The oligonucleotide was used as the Na^+ salt. The concentration of d(A)$_{16}$ was determined at 260 nm using the absorption coefficient 155 000 M^{-1} cm^{-1}.

Spin-Labeled Oligonucleotide Binding Experiments. The binding experiments were carried out by adding small amounts of a concentrated solution of the spin-labeled oligonucleotide (up to 0.05 molar equiv) to the protein solutions. The experiments were executed in a manner similar to that described previously in a study of the IKe phage-encoded gene V protein (de Jong et al., 1988, 1989).

Table I: 2D Experiments Performed Using Spin-Labeled *d(A)$_3$*

sample[a]	concn (mM)	solvent	expt	mixing time (ms)	molar equiv of *d(A)$_3$*[b]
GVP Y41H	2.5	H_2O	NOESY	150	0.05
GVP Y41H	3.0	H_2O	NOESY	150	0.02
GVP Y41H	4.0	D_2O	NOESY	100	0.05
GVP Y41H	4.0	D_2O	TOCSY	26	0.05
wild-type GVP	1.5	D_2O	TOCSY	34	0.03

[a] The pH was adjusted to 5.2 in all samples. [b] Amount of added *d(A)$_3$* in molar equivalents (defined with respect to the monomer protein concentration).

The *d(A)$_3$* (D_2O) solutions contained a few millimolar salt and were prior to the experiments adjusted to the pH of the protein sample. The concentration of mutant GVP Y41H ranged from 2 to 4 mM (monomer), and the wild-type GVP concentration amounted to 1.5 mM (monomer). The pH was adjusted to 5.2 with diluted DCl. An overview of the 2D experiments which were performed is given in Table I.

NMR Spectroscopy. ^{1}H-NMR experiments were performed at 600 MHz on a Bruker AM600 spectrometer interfaced to an Aspect 3000 computer. The data were processed on an Aspect 3000 workstation. All two-dimensional spectra were recorded in the pure phase absorption mode by making use of TPPI (Marion & Wüthrich, 1983) with the carrier at the position of the solvent resonance. TOCSY (or HOHAHA) (Bax & Davis, 1985) were recorded in D_2O solution with mixing times of 26 and 34 ms. The mixing was carried out with the MLEV17 pulse scheme with delays before and after the 180° pulse of the scheme to compensate for cross-relaxation peaks (Griesinger et al., 1988). NOESY (Jeener et al., 1979) spectra were recorded with mixing times of 100 and 150 ms. In the NOESY(H_2O) experiments, suppression of the solvent resonance was achieved by replacing the last pulse of the sequence by the jump–return sequence (Plateau & Guéron, 1982). For experiments in D_2O, only continuous irradiation of the residual water resonance was used during the recycle delay. Optimization of the receiver phase was performed to eliminate base-line distortions (Marion & Bax, 1988). Furthermore, additional base-line corrections were performed on the spectra recorded in H_2O solution, after Fourier transformation in two dimensions. Typically, 400–512 increments of 2K data points were recorded. The data sets were processed with either Gaussian or squared sine functions, and zero-filling was applied to obtain spectra with a resolution of 8.1 Hz/point.

Difference Spectra. 1D and 2D difference spectra were calculated by subtraction of the spectra recorded before and after the addition of the spin-labeled oligonucleotide to the protein sample. Apart from absolute difference spectra for which the cross-peak intensities are given by

$$I_a(\text{diff}) = I(\text{absence}) - I(\text{presence})C$$

also relative difference spectra were calculated for which I_r-(diff) is given by

$$I_r(\text{diff}) = 100[I(\text{absence}) - I(\text{presence})C/I(\text{absence})]$$

I(absence) and I(presence) are the intensities of the peaks in the absence and presence, respectively, of the spin-labeled oligonucleotide. In the calculation of relative difference spectra, a minimum absolute intensity level, I(minimum), for peaks in the unperturbed spectrum was defined to discriminate between noise and peaks. C is a constant factor used to correct for systematic deviations in cross-peak intensities (de Jong et al., 1988).

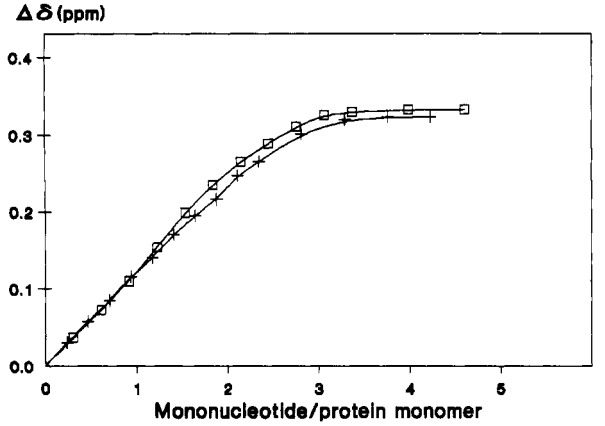

FIGURE 1: Plots of the shift ($\Delta\delta$) in ppm of the ϵ-protons of Tyr26 of wild-type (+) and mutant GVP Y41H (□), respectively, as a function of added oligonucleotide $d(A)_{16}$. The amount of $d(A)_{16}$ added is expressed as the ratio of oligonucleotide vs protein (in mononucleotide and protein monomer units).

The subtraction of the two-dimensional spectra was achieved using a home-written PASCAL program running on the Aspect 3000 computer.

RESULTS

General Remarks. The majority of the experiments which will be discussed below have been applied to mutant Y41H of M13 GVP because it exhibits significantly improved solubility characteristics over wild-type GVP (Folkers et al., 1991a). For comparison, however, we have also performed NMR difference measurements utilizing the spin-labeled oligonucleotides with a 1.5 mM sample of wild-type GVP in D_2O solution. The validity of the approach was further checked by comparing the binding of nonmodified oligonucleotides to wild-type GVP and mutant GVP Y41H, respectively. To this end, we have titrated both mutant and wild-type proteins with a variety of oligonucleotides with different chain lengths and recorded ^{1}H-NMR spectra in a similar fashion as described by Alma et al. (1982). The binding affinity gradually increases with increasing oligonucleotide chain length for both protein molecules [see also Alma et al., (1982)]. The spectral changes which occur as consequence of the addition of different oligonucleotides to both molecules are, however, very similar. In Figure 1, the shifts of the aromatic $C^\epsilon H$ resonances of tyrosine-26 of both wild-type and mutant GVP Y41H are shown as a function of added oligonucleotide $d(A)_{16}$ (see legend of Figure 1). The results obtained indicate that mutant GVP Y41H binds to ssDNA with virtually the same affinity as the wild-type protein. Additional fluorescence binding experiments have revealed, however, that the binding affinity of mutant GVP Y41H for polynucleotides is decreased in comparison to wild-type GVP, which can be mainly attributed to a lower cooperativity factor (Stassen et al., 1992b).

The spin-labeled oligonucleotide bound to a macromolecule selectively broadens the resonances which are in close vicinity of the spin-label. Difference spectra computed from spectra obtained in the absence and presence of the spin-labeled ligand will thus display the resonances originating from the residues which are part of the DNA binding domain and its surroundings. Both "absolute" and "relative" 2D difference spectra were calculated. The absolute difference spectra display in particular the difference effects of strong cross-peaks. The relative difference spectra instead give a more reliable view of the genuine perturbation of the cross-peaks in the 2D spectra. For a more detailed explanation of the experiment, we refer to papers published previously (de Jong et al., 1988, 1989).

The difference spectra also reveal some small shifts that are caused by the change in sample conditions upon addition of the spin-labeled DNA. It was suggested previously that the shift effects are related to a specific aggregation effect of GVP which is indicative of the occurrence of specific protein–protein interactions between dimeric gene V protein molecules (Folkers et al., 199a). A change in the aggregation state of the protein is accompanied by small shifts of resonances of residues which possibly are involved in protein–protein interaction.

By plotting both the positive and negative levels, we checked the origin of the difference effects in absolute difference 2D spectra to discriminate paramagnetic relaxation effects from shift effects. The degree of aggregation of GVP is also dependent on the total protein concentration (Folkers et al., 1991a). Hence, the shift effects were also investigated by means of protein concentration dependent 2D-NMR measurements using wild-type GVP samples.

Analysis of the Paramagnetic Difference Spectra. In Figure 2, two TOCSY(D_2O) spectra of 4 mM GVP Y41H are shown, the first recorded in the absence of *$d(A)_3$* and the second being the absolute difference spectrum calculated from the spectra recorded in the absence and presence of 0.05 molar equiv of *$d(A)_3$*. It is noted that the residues which are significantly perturbed by the spin-labeled oligonucleotides also shift significantly in the presence of nonmodified oligonucleotides. The addition of the spin-label and calculation of the difference spectra result in a tremendous reduction of connectivities. The connectivities observed in the TOCSY(D_2O) spectrum could be assigned by virtue of the nearly complete assignment of mutant GVP Y41H made by us previously (Folkers et al., 1991b). Apart from obtaining difference spectra in D_2O, it was also possible to perform the experiments in H_2O with samples of GVP Y41H. Particularly, useful in this respect is the recording of NOESY(H_2O) difference spectra because those give the possibility of making sequence-specific assignments for a special subset of the protein sequence. Figure 3 shows the potential of such an experiment in an absolute difference plot. Almost continuous sequential walks can be made for three stretches of the protein sequence. The major stretch extends from residue Gln10 to residue Gln30; the resonances of the first five residues of this stretch appear in the spectrum because of the occurrence of shifts instead of a broadening of the cross-peaks. The sequential walk is secured by many sequential connectivities because of the long mixing time used. The strongest perturbations were found for the part comprising residues Arg16–Leu28. A second stretch of residues which could be tied up in the NOESY(H_2O) difference spectra starts at the α resonance of Val45 and ends at residue Glu51. The intraresidual NOE contacts of Leu44 are missing in the difference spectrum because of its absence in the original NOESY(H_2O) spectrum. However, the TOCSY(D_2O) spectrum shown in Figure 2 reveals a part of the Leu44 connectivity pattern, and therefore this residue is still to be considered as being part of or close to the DNA binding domain. A third stretch of residues which could be identified in a sequence-specific manner comprises residues Lys69 through Arg80. The sequential walk is, however, interrupted at residue Leu76. The connectivity pattern expected for Leu76 is hardly visible in all recorded 2D spectra. The sequential $d_{\alpha N}(i,i+1)$ and $d_{\beta N}(i,i+1)$ connectivities between Ser75 and Leu76, are, however, present in the NOESY(H_2O) difference spectra. A very small sequential $d_{\alpha N}(i,i+1)$ contact between Leu76 and Met77 permits the

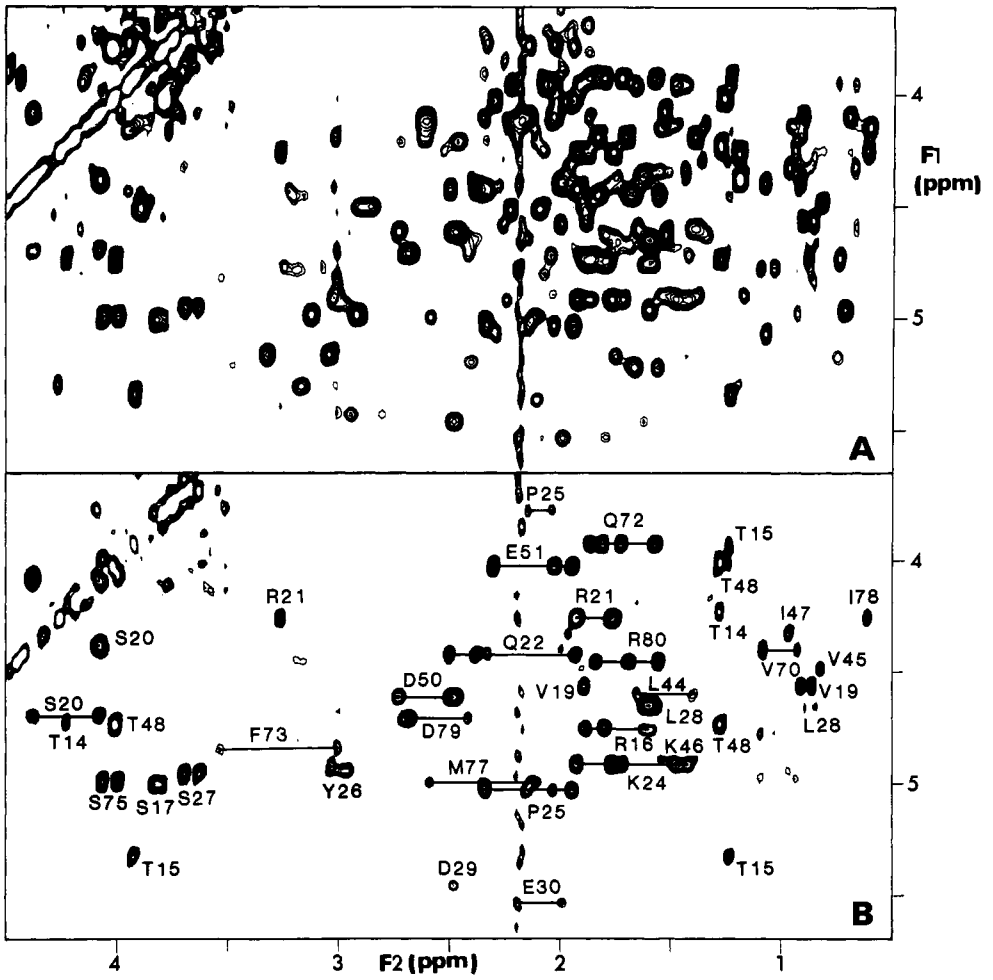

FIGURE 2: (A) Part of the 26-ms TOCSY(D_2O) spectrum of 4 mM GVP Y41H recorded at a temperature of 298 K. (B) Absolute difference plot calculated by subtraction of the spectrum in (A) and a TOCSY spectrum recorded after addition of 0.05 molar equiv of *$d(A)_3$*. The assignment of all connectivity patterns is indicated. Relayed connectivities are indicated by continuous lines.

further continuation of the sequential walk. The results obtained are in exact accordance with the assignments reported previously (Folkers et al., 1991b).

Apart from the sequential connectivities, many long-range connectivities emerge in the difference spectrum. Most of these had already been assigned and are of the $d_{\alpha N}(i,j)$ type which are indicative of the antiparallel β-sheet structure which is abundantly present in GVP Y41H (Folkers et al., 1991b). However, not all long-range NOE connectivities which emerge in the difference spectra could be unambiguously assigned before.

Very interesting in this respect is the contact between the α resonance of Ser75 and the amide resonance of Lys46. Before considering this particular NOE contact, we note that the secondary structure elements which were identified from the 2D experiments previously (Folkers et al., 1991b) could be extended after careful examination of the 2D ^{1}H-NMR, 3D ^{1}H-NMR, and also 3D heteronuclear NMR data involving ^{15}N-labeled GVP Y41H (data not shown). It turned out that the N-terminal segment of GVP (residues 3–6) is connected to one strand of the complex loop (residues 33–36) as judged from several backbone to backbone, backbone to side chain, and side chain to side chain NOE contacts (see Figure 4). In addition, the amides of residues 5 and 34 and to a lesser extent of residues 7 and 36 have a slow rate of exchange with the solvent and therefore likely are hydrogen-bonded to the carboxyl oxygens of the opposite strand in the antiparallel β-sheet structure (see Figure 4).

The secondary structure elements as deduced so far from the NMR data are present within each monomeric unit of the GVP dimer (Folkers et al., 1991b). This secondary structure, however, does not include the unambiguously assigned NOE between residues Lys46 and Ser75. It is therefore very likely that this NOE is a so-called intermonomer NOE which connects the backbones of both monomeric units (see Discussion). After this crucial NOE was assigned, a very small NOE which connects the amides of residues 76 and 46 could also be identified. The complete secondary structure including the intermonomer contacts is presented in Figure 4. The NOE data and the fact that the amide of Lys46 exchanges very slowly (data not shown) suggest the presence of a hydrogen bond between the carbonyl of Gly74 and the amide of Lys46.

Other long-range NOE connectivities which appeared in the NOESY(H_2O) difference spectra are the NOE contacts between the $C^\beta H$ and $C^\gamma H$ of Thr48 and the side chain amide proton of Arg80. The presence of these NOE connectivities was corroborated by the fact that in NOESY(D_2O) difference spectra, NOE connectivities between $C^\beta H$ and $C^\gamma H$ of Thr48 and the $C^\delta H$ resonances of Arg80 were found.

Description of the DNA Binding Domain. After assigning all connectivity patterns in the 2D difference spectra, we are able to consider the location of the binding domain in more detail. The relative difference 2D spectra are used for this purpose. All residues which are affected by the spin-label are shown in Figure 5 and are arranged within the framework of the elucidated secondary structure (see Figure 4). The

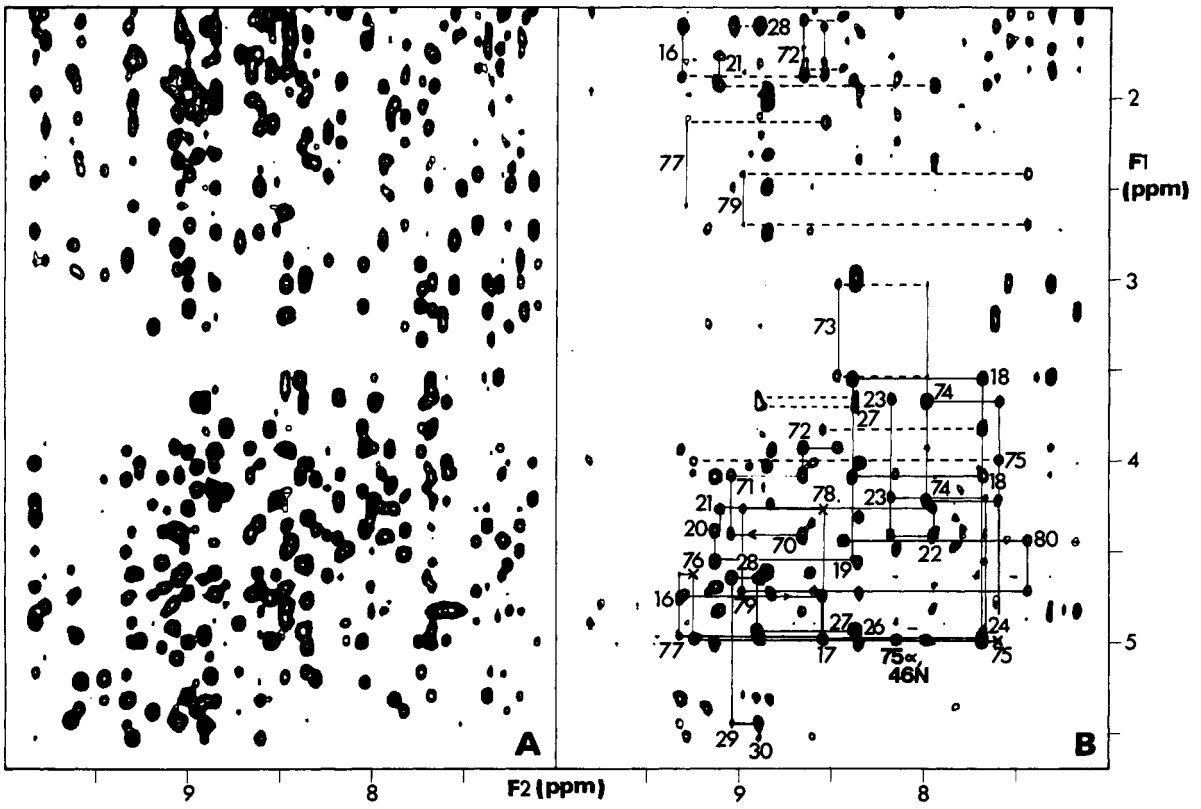

FIGURE 3: (A) Part of a 150-ms NOESY(H_2O) spectrum of 2.5 mM GVP Y41H recorded at a temperature of 298 K. (B) Absolute difference plot calculated by subtraction of the spectrum in (A) and a NOESY spectrum recorded after addition of 0.05 molar equiv of *d(A)$_3$*. For the stretches comprising residues 16–30 and 70–81, $d_{\alpha N}(i,i+1)$ and $d_{\beta N}(i,i+1)$ are indicated by continuous and dashed lines, respectively. The intraresidual $C_\alpha H(i)$–NH(i) NOE connectivities are labeled. Missing intraresidual $C_\alpha H(i)$–NH(i) NOE connectivities are marked with an x.

percentage intensity decrease of the cross-peaks of the residues was taken from the TOCSY(D_2O) data of GVP Y41H and is included in Figure 5 as well. Some of the cross-peaks of the relatively strongly coupled α resonances of the glycines are missing in the TOCSY spectrum because of a significant contribution of antiphase magnetization to these cross-peaks. Therefore, we have extracted the values for the percentage intensity decrease of the cross-peaks for several of the glycines from NOESY(D_2O) and NOESY(H_2O) difference spectra after a full comparison of the data obtained in the different NMR difference experiments was made. The actual percentages which were found in the different experiments are not exactly the same. The reason for this is related to differences in signal filtering in the different types of experiments. The rotational correlation time of the GVP system (τ_c = 15 ns) is in the range where the paramagnetic effect on T_2 (and $T_{1\rho}$) is considerably stronger than that on T_1 (de Jong et al., 1988). In TOCSY experiments, we are dealing with so-called T_2-filtering or actually $T_{1\rho}$ filtering, and in NOESY experiments, we are dealing with an effect on the T_1 relaxation time. Furthermore, there are also differences in the length of the mixing times used in the experiments and the amounts of spin-label which have actually been used. Apart from differences in the actual percentage values, the overall pattern in the percentage intensity decrease of the residue cross-peaks which emerges in all data sets available is very much the same.

It can be seen immediately that the signals of the aromatics Phe73 and Tyr26 are filtered out to 100%. Apart from these two aromatic residues, the resonances of the serine at position 20 and the leucine at position 28 are also strongly affected in the presence of the spin-label. Therefore, we can conclude that the average distance between the free electron of the bound spin-labeled molecule and the nuclei of the Ser20, Tyr26, Leu28, and Phe73 spin systems is the shortest of all spin systems for which cross-peaks are observed in the TOCSY contour plots of the unperturbed protein.

Different segments of the protein sequence are affected by the spin-label. The protein segment running from residue 16 to residue 28 which had been designated earlier as the DNA binding loop is very strongly affected by the spin-label. Examination of Figure 5 clearly reveals the distinct differential influence of the spin-label on the spin systems of this loop. The strongest difference effects are connected with residues in its β-ladder. The residues which constitute the tip of the loop are influenced to a lesser extent. The β-ladder comprises two sides: one side is formed by the side chains of the residues Ser17, Val19, Pro25, and Ser27; the other side is formed by the side chains of residues Arg16, Gly18, Ser20, Lys24, Tyr26, and Leu28. The distance dependence of the spin-label-induced relaxation implies that the distance between the spin-label and the residues which are part of the surface constituting residues Arg16, Gly18, Ser20, Lys24, Tyr26, and Leu28 must on the average be shorter than that between the spin-label and the opposite surface. The differences become even more clear when the individual connectivities within each residue are evaluated. Especially, the relative intensity differences of the connectivities belonging to protons which are positioned further up in the side chains become apparent. For example, in residue Lys24 which is on the surface closest to the spin-label, the connectivities of the protons further up in the side chain become more perturbed by the spin-label, while in, e.g., residue Val19 these connectivities are less perturbed by the spin-label. This further indicates that the TEMPO group of the bound spin-labeled oligonucleotide is in close contact with one of the two surfaces. The data presented in Figure 5 show

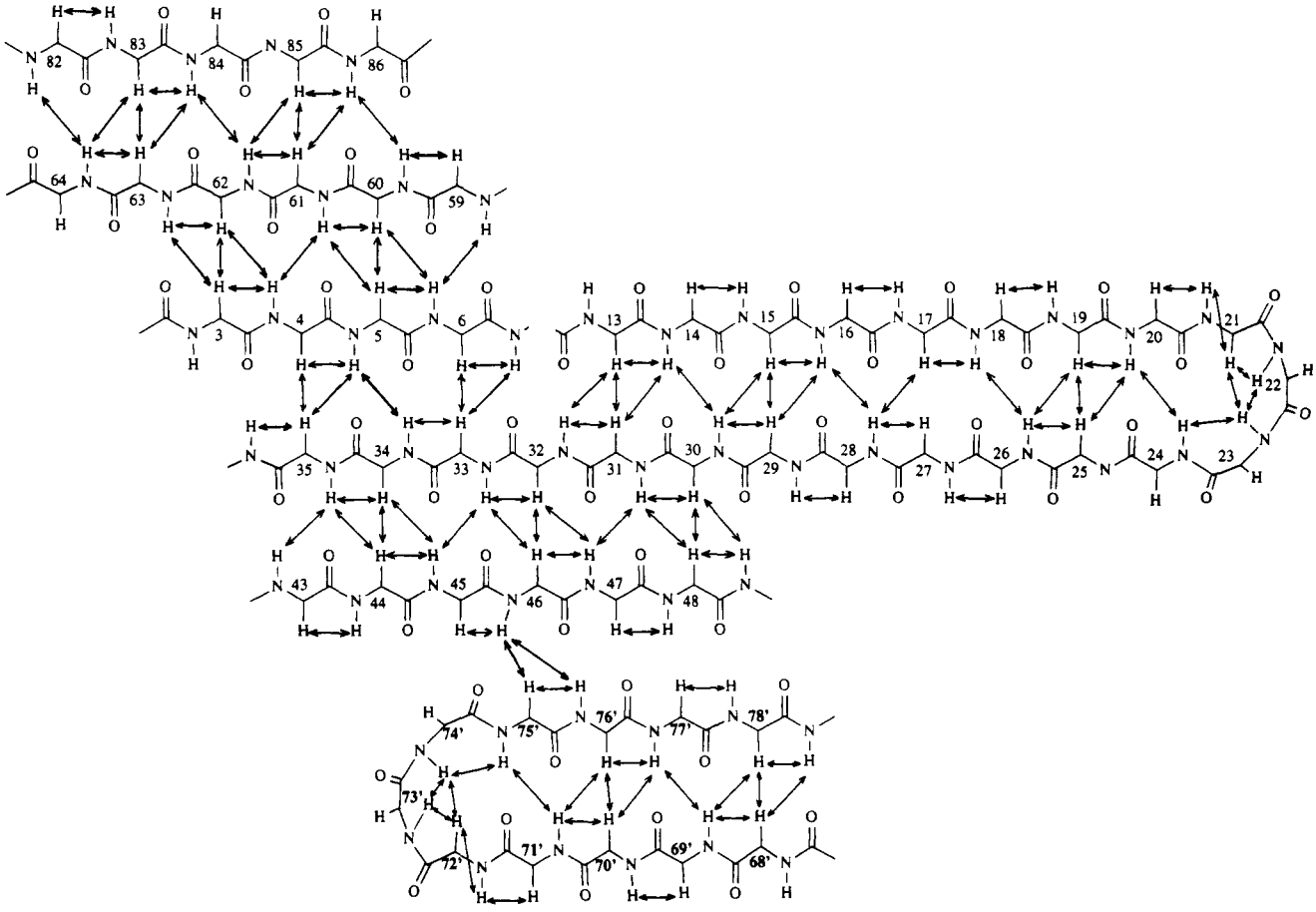

FIGURE 4: Schematic representation of the complete secondary structure of M13 GVP Y41H. The sequential and long-range NOEs are indicated by arrows. The prime notation refers to residues that derive from the second monomer.

no discernible difference between the perturbation of residues Arg16 and Ser17.

This is partly due to the fact that no connectivities are present in the TOCSY spectrum involving protons which reside further up in the side chain of the Arg16 residue. The NOESY(H_2O) difference spectra reveal, however, a stronger perturbation of the connectivities belonging to the Arg16 spin system than the ones belonging to the Ser17 spin system.

The protein segment neighboring the DNA binding loop comprising residues 44 through 51 is also affected by the spin-labeled oligonucleotide. A part of this protein segment (residues 43–48) is arranged in an antiparallel fashion with respect to a stretch comprising residues 30–35 (Folkers et al., 1991b). The other residues are part of a turn. The typical character of the antiparallel β-structure where the side chains of the amino acids alternatingly point upward and downward is also in the case of this β-strand (residues 43–48) reflected in the relaxation-induced perturbation pattern. The surface of the β-strand formed by the side chains of residues Leu44, Lys46, and Thr48 is closer to the spin-label than the surface formed by the side chains of residues Val45, Ile47, and Leu49. The residues in the turn, i.e., Asp50 and Glu51, are both moderately affected by the spin-label.

It is very striking that the opposite β-strand comprising residues 30–35 of this particular antiparallel β-sheet (formed by residues 43–48 and 30–35) is not affected by the spin-label, except for a weak effect on Glu30.

The protein segment running from residue Lys69 to Arg80 which is part of the so-called dyad domain (Brayer & McPherson, 1983) of the protein also emerges in the paramagnetic difference spectra. The majority of the residues in this segment are part of a so-called β-loop (residues 68–78). It can be seen immediately from Figure 5 that in contrast, to the DNA binding loop, the residues in the tip of this loop are more significantly perturbed than the residues which constitute the β-ladder. Especially the phenylalanine at position 73 is very strongly affected. The stretch of residues 69–71 which is part of one of the β-ladders of the loop is only weakly affected by the spin-label, and a clear alternating pattern in the perturbation of its residues cannot be seen. The stretch comprising residues 75–78 is also weakly affected by the spin-label except for residue Ser75. It seems likely that the spin-label is closer to the surface formed by the side chains of residues 75 and 77 than the surface formed by residues 76 and 78. Residues Asp79 and Arg80, which are part of a turn following the above-mentioned β-loop, are both moderately perturbed by the spin-label.

Although not many differences were expected between the results of the spin-label oligonucleotide binding experiments obtained for mutant GVP Y41H and the wild-type molecule, a comparison was made between these two. To this end, we have recorded TOCSY(D_2O) spectra of 1.5 mM wild-type GVP in the absence and presence of the spin-labeled oligonucleotide. Although the experimental conditions used in the experiments performed on Y41H GVP and wild-type GVP were not exactly the same (see Materials and Methods), the actual percentage intensity decrease values of the residue cross-peaks are almost identical. Therefore, it can be concluded that both wild-type and Y41H GVP share the same DNA binding domain.

Instead, differences between the two proteins are related to their interactions between gene V protein dimers. Two-

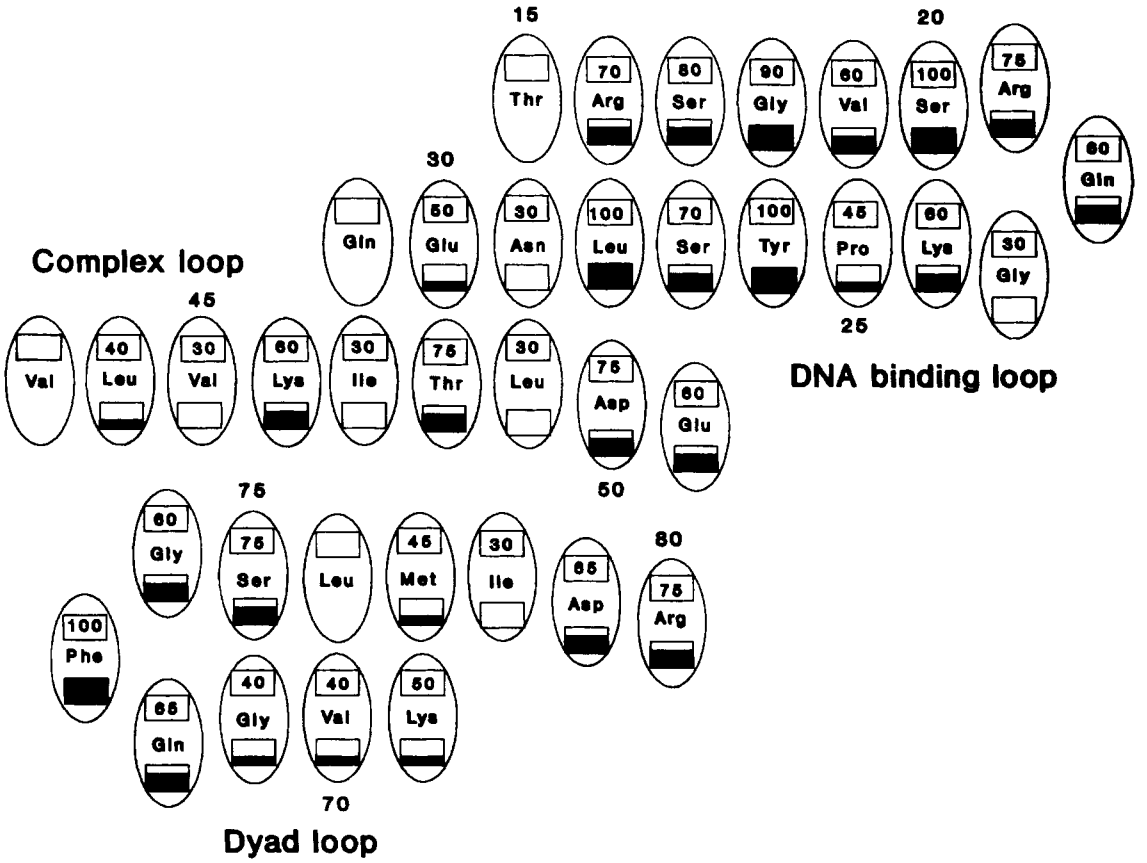

FIGURE 5: Schematic representation of the percentage intensity decrease in cross-peak intensities in M13 GVP Y41H. The numbers in the boxes above the residues denote the average of the percentage intensity decrease of that residue's cross-peaks. Additionally, the boxes under the residues are representative of the degree of the decrease in the cross-peak intensities as follows: (■) boxes refer to strongly affected residues (>85%); (▣) boxes refer to moderately affected residues (60–80%); (▤) boxes refer to weakly affected residues (35–55%); (□) boxes refer to very weakly affected residues (<35%).

dimensional protein concentration-dependent NOESY(D_2O) measurements were performed on wild-type GVP in order to obtain more information about residues possibly involved in dimer–dimer interaction. Apart from the aromatic resonances of Tyr41 which shift quite significantly depending on the protein concentration [e.g., see Folkers et al. (1991a)], shift effects could be observed for residues Phe13, Tyr34, Gly38, Asn39, Glu40, and Pro42.

Similar shift effects could also be observed in 1D and 2D paramagnetic difference spectra. The addition of the spin-labeled oligonucleotide causes a change in the sample conditions (Folkers et al., 1991a). The NOESY(H_2O) difference spectrum of Y41H GVP shown above displayed shifts for residues 10–15. In other experiments, small shifts were also observed for residues 3, 4, 5, 35, 36, 60, 61, 62, 83, 84, and 85.

DISCUSSION

Recently, we reported the ^{1}H-NMR assignments for the GVP mutant Y41H of the single-stranded DNA protein encoded by gene V of the filamentous bacteriophage M13 (Folkers et al., 1991b). The progress made in the ^{1}H-NMR structural studies of M13 GVP Y41H has also enabled us to further explore its DNA binding domain using spin-labeled oligonucleotides in combination with ^{1}H-NMR.

The paramagnetic difference spectra obtained with the aid of the spin-labeled oligonucleotide ligands not only provide information about the location of the ssDNA binding domain but also result in a vast data reduction which in itself is very helpful in providing additional sequential and long-range NOEs. We would like to emphasize that the spectral changes which occur upon binding of spin-labeled oligonucleotides to GVP are congruent with the changes observed in the binding of different nonmodified oligonucleotides ranging from 2 to 60 in length (Alma et al., 1982; King & Coleman, 1987, 1988; de Jong et al., 1989; Folkers et al., 1991a). Previously, it had already been shown that the difference spectra acquired with GVP protein samples dissolved in D_2O simplify the residue-specific assignment of residues located in the binding groove to a great extent (de Jong et al., 1989). In addition, the spin-label experiments in H_2O presented in this paper are of great help in the sequential assignments. A very nice example is the assignment of Phe73. The resonance position of the amide proton of Phe73 almost coincides with that of four other residues (Asp36, Leu65, Ser67, and Leu81) which complicates its assignment significantly (Folkers et al., 1991b). The difference spectra in H_2O, however, only display a limited number of connectivities at that particular chemical shift position, making the assignment of Phe73 rather straightforward (see Figure 3).

The secondary structure of M13 GVP Y41H consists of two β-loop structures and a large five-stranded antiparallel β-sheet. The presence of the five-stranded antiparallel β-sheet in the solution structure of M13 GVP Y41H has not been reported as such by us previously (Folkers et al., 1991b). Examination of some recently recorded 2D-NMR data sets under different conditions (higher temperature), and some homo- and heteronuclear 3D-NMR data sets, revealed the NOE contacts shown in Figure 4 (see Results) which join the previously identified triple-stranded antiparallel β-sheet comprising three stretches of the peptide chain (Lys3–Ile5, Leu60–Val63, and Leu83–Pro85) with the previously identified

double-stranded antiparallel β-sheet which is part of the so-called complex domain of the protein (residues 30–48). The similarity between the overall folding in the NMR structure and the available crystal structure of wild-type GVP indicates that the secondary structure elements as deduced by NMR are present within the monomeric unit (Brayer & McPherson, 1983; Folkers et al., 1991b). It was, however, noted previously that the NMR data and those derived from X-ray crystallography are different at the local level. Therefore, we have specifically compared the newly derived parts of the secondary structure which emerged from our NOE evaluation with the distances between the relevant backbone atoms from the crystal coordinates. The crystal data do not reveal the β-structure element connecting the strands comprising residues 3–6 and residues 33–35. In the crystal structure, residues Pro8 and Val35 are in close proximity of one other, suggesting a four-residue shift of the crystal data with respect to the NMR data.

The NOESY difference spectra reveal several long-range NOEs. Apart from the long-range NOEs of the $d_{\alpha N}(i,j)$ and $d_{\alpha\alpha}(i,j)$ type which make up the secondary structure elements in GVP, not many long-range NOEs were observed. The reason for this is related to the fact that the spin-label is a surface probe and therefore does not reveal long-range information of residues which constitute the central core of the protein.

The long-range NOEs particularly mentioned under Results were not assigned before. The long-range NOE between the α resonance of Ser75 and the NH resonance of Lys46 which is of the $d_{\alpha N}(i,j)$ type is not accounted for in the elucidated antiparallel β-structure elements. Therefore, this particular NOE together with the NOE connecting the amides of Lys46 and Leu76 can only be explained as being intermonomer NOEs, i.e., NOEs between backbone atoms which are part of different monomer units. The presence of these NOEs together with the fact that the amide of Lys46 exchanges slowly with the solvent suggests the presence of a hydrogen bond between the amide of Lys46 and the carbonyl of Gly74. Again, the NMR data are not congruent with the crystal data. In the crystal structure, residue Lys46 and residue Phe73 of the symmetry-related monomer are in close proximity of one another. Furthermore, intermonomer hydrogen bonds are found between the amide of Gly74 and the carbonyl of Lys46 as well as the carbonyl of Phe73 and the amide of Lys46, suggesting a one-residue shift with respect to the NMR data.

The observed NOEs between Thr48 and Arg80 probably are intramonomer contacts as judged from preliminary structure calculations based on our NMR data (unpublished results) as well as from the available crystal structure.

The results of our spin-label oligonucleotide binding study allow a comparison with experimental results from a variety of physicochemical studies performed over the years, all intended to obtain more insight in the DNA binding domain of GVP. In most of these studies, special attention was paid to positively charged and aromatic residues. It has, for example, been suggested from chemical modification studies that Lys24, Lys46, and Lys69 are involved in binding to ssDNA (Dick et al., 1988). Indeed, our spin-labeled oligonucleotide binding studies indicate that Lys24, Lys46, and to lesser extent Lys69 are in close vicinity to the spin-label. The other three lysines present in GVP, Lys3, Lys7, and Lys87, are not affected by the spin-label, and the secondary structure model proposed by us also does not indicate that these residues are in close vicinity to the structural elements which are involved in ssDNA binding. The positively charged arginines at positions 16, 21, and 80 have also been implicated in ssDNA binding (Brayer & McPherson, 1984), which was indeed corroborated in our binding studies. Arg82 is not involved in ssDNA binding although a very small perturbation effect caused by the oligonucleotide spin-labels on the side chain amide resonance of this particular residue has been observed (data not shown). Of the aromatic residues, we and others have already argued that the only aromatic residues which are involved in ssDNA binding are the tyrosine at position 26 and the phenylalanine at position 73 of the molecule (King & Coleman, 1987; Folkers et al., 1991a). NMR binding studies with longer ssDNA fragments which may be more relevant with respect to complex formation under cooperative binding conditions reveal a similar view (King & Coleman, 1988). Finally, the hydrophobic residue Leu28 has also been implicated to interact with the nucleotides (King & Coleman, 1987). The paramagnetic difference spectra shown in this paper provide additional information about the DNA binding domain and its surroundings. Residues Ser20, Thr48, Asp50, Glu51, and Ser73, for example, had never been implicated as being part of or close to the DNA binding domain. The DNA binding domain comprises residues which constitute two antiparallel β-loops and a β-strand which runs contiguous to one of the β-loops. The so-called DNA binding loop (residues 16–28) is the most important loop as judged from the relaxation-induced perturbation pattern. This is corroborated by the fact that this part of the amino acid sequence of the protein is conserved best when the amino acid sequences of closely and distantly related ssDNA binding proteins are compared (van Duynhoven et al., 1993). The spin-label is in close vicinity of one side of the β-ladders of the loop constituting the even-numbered residues such as Ser20, Tyr26, and Leu28. The other β-loop is part of the dyad domain of the protein. In this β-loop, mainly the tip is affected by the spin-labeled oligonucleotide. Both these β-loops are relatively flexible as judged from the fact that the exchange of their amide protons is relatively fast (Folkers et al., 1991b). A part of the complex domain of the protein consisting of a β-strand (residues 44–49) and part of a loop (residues 50 and 51) is also part of the DNA binding domain.

As mentioned earlier, the paramagnetic difference spectra contain the intermonomer NOEs between residue Lys46, on the one hand, and residues Ser75 and Leu76, on the other hand. The fact that these NOE contacts appear in the NOESY(H_2O) difference spectrum forces us to conclude that the DNA crosses the monomer–monomer interface of the GVP dimer. In other words, this observation implies that the DNA binding domain of GVP is divided over its two functional units, a phenomenon which has been suggested previously but never has been unambiguously proven (Brayer & McPherson, 1984).

The location of the putative DNA binding domain of GVP has been proposed earlier on the basis of molecular modeling utilizing the crystal structure in conjunction with solution data (Brayer & McPherson, 1984). The NMR data clearly predict a different DNA binding path than this model describes. This is in part due to the structural differences between solution and crystal structures. Another reason is that in the X-ray model some amino acid residues are implicated in ssDNA binding that are not affected by the presence of the spin-labeled oligonucleotides in our NMR experiments.

Very prominent structural differences between solution and crystal structures are found in the DNA binding loop of GVP

of which the residues are clearly at a different position in the two structures (van Duynhoven et al., 1990). The fact that a great deal of flexibility exists in the β-loop indicates that the position of the loop relative to the core of the protein may change upon binding. A rearrangement of the residues of the binding loop as predicted in the X-ray modeling study is, however, not expected and probably not needed when considering the solution data. According to the NMR data, the DNA is positioned on one of the sides of both strands of the DNA binding loop and is not running along one of the strands of that β-loop as the X-ray model suggests. The DNA binding path runs along a part of the complex domain, of which residues 46 and 48 and also the loop residues 50 and 51 are positioned on the same surface as all even-numbered residues of the DNA binding loop, a feature which is lacking in the GVP crystal structure. In the solution and crystal structures, residue Arg80 which takes part in ssDNA binding is positioned close to residue Thr48 and probably stems from the dyad domain (residues 62–82) of the same monomer unit as Thr48. In the X-ray model, the DNA binding path crosses the monomer–monomer interface to interact with the tip residue Phe73 of the dyad loop of the symmetry-related monomer and then runs on along the loop of the complex domain, thereby interacting with Tyr-34 and -41. The NMR data indicate that the binding path indeed crosses the monomer–monomer interface to interact with residues constituting the tip of the dyad loop. However, the complex loop is not involved in ssDNA binding. As a consequence, the path is shorter than the chosen stoichiometry of 5 nucleotides per monomer of the proposed model, probably being 3–4 nucleotides per monomer in length, congruent with the outcome of several binding studies (Alma et al., 1983; Bulsink et al., 1986; Kansy et al., 1986).

More recently, an alternative model for the complex has been proposed on the basis of the crystal structure (Hutchinson et al., 1990). The model does not implicate Tyr41 in ssDNA binding but suggests the involvement of Tyr34 in ssDNA binding under cooperative binding conditions. The importance of Tyr34 as well as the close proximity of Leu83 in the DNA binding channel in their model could not be confirmed in our studies with the spin-labeled oligonucleotides.

As reported earlier, especially residue Tyr41 and maybe also residue Tyr34 are involved in protein–protein interactions between GVP dimers (Folkers et al., 1991a). It is of interest in light of modeling studies describing the full GVP–ssDNA superhelix to investigate which protein surfaces on the protein are responsible for the strong interaction between the GVP dimers which may form the physical basis for the observed cooperativity in the binding of the protein to ssDNA.

Hence, two-dimensional protein concentration-dependent NOESY(D_2O) measurements have been used to shed some light on the possible protein–protein interaction surfaces. The increase of the protein concentration leads to an increase in aggregation which is accompanied by small changes of the chemical shift positions of the proton resonances of several residues. The problem with these experiments is that due to spectral overlap and broadening of all spectral lines at relatively high protein concentrations, only significantly shift effects are observed in not too crowded regions of the spectrum. In fact, this problem is rather similar to the interpretation of, for example, shift effects in normal oligonucleotide protein binding experiments.

Careful analysis of the origin of the difference effects in the paramagnetic difference spectra obtained with the use of the spin-labeled oligonucleotide ligands revealed some additional shift effects. Apart from the residues constituting the complex loop (residues 36–42) of the protein, it seems likely that the surface formed by residues of three stretches of the polypeptide chain (Lys3–Ile5, Leu60–Val63, and Leu83–Pro85) is involved in protein–protein interactions as well.

In conclusion, it could be shown that GVP predominantly utilizes two β-loops for complex formation with ssDNA which are divided over both functional units of the GVP dimer. In the study, further differences between the crystal and solution structures have become apparent. Therefore, the elucidation of the complete three-dimensional solution structure is now imperative to obtain a more precise view of the ssDNA–GVP interaction.

ACKNOWLEDGMENT

We thank Mrs. A. J. Jonker and Mrs. C. W. J. M. Prinse for their efforts in isolation and purifying wild-type and mutant GVP. The NMR experiments were performed at the SON Hf-NMR facility (Nijmegen, The Netherlands). Mr. J. J. M. Joordens and Dr. S. S. Wijmenga are acknowledged for expert technical assistance.

REFERENCES

Alma, N. C. M., Harmsen, B. J. M., van Boom, J. H., van der Marel, G., & Hilbers, C. W. (1982) *Eur. J. Biochem. 122*, 319–326.

Alma, N. C. M., Harmsen, B. J. M., de Jong, E. A. M., van der Ven, J., & Hilbers, C. W. (1983) *J. Mol. Biol. 163*, 47–62.

Arentzen, R., Van Boeckel, C. A. A., Van der Marel, G., & van Boom, J. H. (1979) *Synthesis*, 137–139.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–366.

Brayer, D. G., & McPherson, A. (1983) *J. Mol. Biol. 169*, 565–596.

Brayer, D. G., & McPherson, A. (1984) *Biochemistry, 23*, 340–349.

Bulsink, H., van Resandt, R. W. W., Harmsen, B. J. M., & Hilbers, C. W. (1986) *Eur. J. Biochem. 157*, 329–334.

Cassani, G. R., & Bollum, F. J. (1969) *Biochemistry 8*, 3928–3936.

Claesen, C. A. A., Daemen, C. J. M., & Tesser, G. I. (1986) *Recl. Trav. Chim. Pays-Bas 105*, 116–123.

de Jong, E. A. M., Claesen, C. A. A., Daemen, C. J. M., Harmsen, B. J. M., Konings, R. N. H., & Hilbers, C. W. (1988) *J. Magn. Reson. 80*, 197–213.

de Jong, E. A. M., van Duynhoven, J. P. M., Harmsen, B. J. M., Konings, R. N. H., & Hilbers, C. W. (1989) *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.

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*, 139–147.

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.

Freemont, P. S., Lane, A. N., & Sanderson, M. R. (1991) *Biochem. J. 278*, 1–23.

Garssen, G. J., Hilbers, C. W., Schoenmakers, J. G. G., & van Boom, J. H. (1977) *Eur. J. Biochem. 81*, 453–463.

Gray, D. M., Sherry, A. D., Teherani, J., & Kansky, J. W. (1984) *J. Biomol. Struct. Dyn. 2*, 77–91.

Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870–7872.

Hutchinson, D. L., Barnett, B. L., & Bobst, A. M. (1990) *J. Biomol. Struct. Dyn. 1*, 1–9.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.

Kansy, J. W., Clack, B. A., & Gray, D. M. (1986) *J. Biomol. Struct. Dyn. 3*, 1079–1110.

King, G. C., & Coleman, J. E. (1987) *Biochemistry 26*, 2929–2937.

King, G. C., & Coleman, J. E. (1988) *Biochemistry 27*, 6947–6953.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Marion, D., & Bax, A. (1988) *J. Magn. Reson. 80*, 528–533.

Plateau, P., & Guéron, M. (1982) *J. Am. Chem. Soc. 104*, 7310–7311.

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.

Story, R. M., Weber, I. T., & Steitz, T. A. (1992) *Nature 355*, 318–325.

van Duynhoven, J. P. M., Folkers, P. J. M., Stassen, A. P. M., Harmsen, B. J. M., Konings, R. N. H., & Hilbers, C. W. (1990) *FEBS Lett. 261*, 1–4.

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.* (in press).