Assignment of the ^{1}H NMR Spectrum and Secondary Structure Elucidation of the Single-Stranded DNA Binding Protein Encoded by the Filamentous Bacteriophage IKe†

J. P. M. van Duynhoven,‡ P. J. M. Folkers,‡ C. W. J. M. Prinse,‡ B. J. M. Harmsen,‡ R. N. H. Konings,‡,§ and C. W. Hilbers*,‡

Nijmegen SON Research Centre for Molecular Design, Structure and Synthesis, Laboratory for Biophysical Chemistry, and Laboratory of Molecular Biology, Faculty of Science, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands

Received July 26, 1991

ABSTRACT: By means of 2D NMR techniques, all backbone resonances in the ^{1}H NMR spectrum of the single-stranded DNA binding protein encoded by gene V of the filamentous phage IKe have been assigned sequence specifically (at pH 4.6, T = 298 K). In addition, a major part of the side chain resonances could be assigned as well. Analysis of NOESY data permitted the elucidation of the secondary structure of IKe gene V protein. The major part of this secondary structure is present as an antiparallel β-sheet, i.e., as two β-loops which partly combine into a triple-stranded β-sheet structure, one β-loop and one triple-stranded β-sheet structure. It is shown that a high degree of homology exists with the secondary structure of the single-stranded DNA binding protein encoded by gene V of the distantly related filamentous phage M13.

IKe and Ff are distantly related and conjugative plasmid-specific filamentous bacteriophages that have *Escherichia coli* as host (Peeters et al., 1983; Model & Russel, 1988). Among other characteristics, they differ from each other with respect to their plasmid specificity. Ff (a synonym used for the closely related bacteriophages M13, fd, and fl) infects only cells that contain plasmids of the Inc F incompatibility groups while for host cell penetration of phage IKe conjugative pili of the Inc N incompatibility group are required. Filamentous phages contain a circular single-stranded genome that after host cell penetration is replicated according to a rolling circle type replication mechanism. For the sequestering of the viral strands from the double-stranded DNA replication cycle, the phage-encoded single-stranded DNA binding protein encoded by gene V (i.e., GVP)[1] is absolutely required. GVP is a dimeric molecule (molecular mass ca. 20 kDa) that binds cooperatively to the newly synthesized viral strands when a particular threshold concentration is reached. In addition to its indispensable role in DNA replication, it has been demonstrated for GVP encoded by phage Ff that it also fulfills, at the level of translation, a gene regulatory function in the infection process (Model et al., 1982; Yen & Webster, 1982; Zaman et al., 1990). Because of their unique ssDNA and RNA binding characteristics, the biochemical and biophysical properties of both IKe GVP and M13 GVP have been studied extensively in order to elucidate the mechanisms by which these macromolecules interact. Despite the fact that the amino acid sequences of IKe and M13 GVP (88 and 87 amino acids, respectively) are only homologous for 44%, a number of studies have, however, demonstrated that their DNA binding properties are very much alike (de Jong et al., 1987a). Also, the nucleoprotein helices formed by M13 and IKe GVP with ssDNA appear indistinguishable when studied by electron microscopy (Gray, 1989). At a more detailed level, far less insight has been achieved in the structural aspects of the protein–ssDNA interaction. NMR studies aimed at structure elucidation of GVP and its complex with ssDNA have lagged behind similar studies of other systems for two reasons. The high molecular weight of the dimer species gives rise to short T_2 relaxation times, a phenomenon resulting in poor resolution of NMR spectra. Moreover, the strong tendency of the proteins to aggregate, a trait most likely related to their property of cooperative binding to ssDNA, worsens the problem. Concentrations higher than 1–2 mM result in line widths unacceptable for 2D NMR studies.

So far, ^{13}C and ^{1}H NMR studies of GVP–DNA complexes revealed the involvement of charged (Dick et al., 1988; Alma et al., 1981) and aromatic residues (King & Coleman, 1987; Alma et al., 1982; de Jong et al., 1987b) in ssDNA binding. Recently, the structure of a particular amino acid segment of IKe (de Jong et al., 1989a) and of M13 GVP (van Duynhoven et al., 1990) was elucidated by 2D NMR. Both segments, encompassing residues 13–32 in M13 GVP and residues 16–29 in IKe GVP, form a regular β-loop which takes part in ssDNA binding (van Duynhoven et al., 1990; de Jong et al., 1989b). Folding of related amino acid sequences found to be present in nonevolutionary related ssDNA binding proteins into similar β-loops has indicated that the residues involved in ssDNA binding in M13 and IKe GVP are highly conserved (de Jong et al., 1989b). This has led us to postulate that in binding of these proteins to ssDNA a common structural motif is involved. Since then, further NMR studies on the Y41H mutant of M13 GVP have enabled a nearly complete assignment of the ^{1}H NMR spectrum. It was found that the secondary structure of the mutant GVP (mainly consisting of antiparallel β-sheet elements) was identical to that of the wild-type M13 GVP (Folkers et al., 1991) but was different at a number of places from the secondary structure proposed for M13 GVP on the

† Supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Advanced Research (NWO).

* Address correspondence to this author.
‡ Laboratory for Biophysical Chemistry.
§ Laboratory of Molecular Biology.

[1] Abbreviations: 2D NMR, two-dimensional nuclear magnetic resonance; NOESY, nuclear Overhauser enhancement spectroscopy; TOCSY, total correlation spectroscopy; TPPI, time-proportional phase incrementation; GVP, gene V protein; ssDNA, single-stranded DNA; CIDNP, chemically induced dynamic nuclear polarization.

0006-2960/92/0431-1254$03.00/0

basis of crystallographic studies (Brayer & McPherson, 1983). This paper reports the nearly complete assignment of the ^1H NMR spectrum of IKe GVP and the subsequent elucidation of its secondary structure in solution. It is shown that the secondary structures of IKe and M13 GVP resemble each other in almost every aspect. Evidence will be provided that the close homology found for their secondary structures may be extrapolated to their tertiary structures.

Materials and Methods

Protein Isolation and Purification. Isolation and purification of IKe GVP with the aid of a DNA–cellulose column were as described (Peeters et al., 1983). Instead of gel filtration on a G75 column, the protein was concentrated and further purified by cation-exchange chromatography on an S Sepharose Fast Flow column (Pharmacia). To remove high molecular weight contaminants, the protein was purified by gel filtration on a Superdex 75 (HR 10/30) FPLC column (Pharmacia) at pH 6.0 (1 mM cacodylate buffer) and high salt concentration (250 mM KCl). After dialysis against H_2O, the purified protein was lyophilized and stored at −20 °C. The concentration of IKe GVP was determined by the UV absorption at 276 nm using an absorption coefficient of 7100 M^{-1} cm^{-1}. The final concentration of IKe GVP in the NMR samples was 1–3 mM; no buffer components were added. The sample contained only a few millimolar of NaCl. Using aliquots of diluted ^2HCl, the pH was adjusted to 4.6–4.8.

2D NMR Spectra. The NMR experiments were performed at 600 MHz on a Bruker AM 600 NMR spectrometer interfaced to an ASPECT 3000 computer. Clean-TOCSY (MLEV17) (Bax & Davis, 1985; Griesinger et al., 1988), DQF-COSY (Rance et al., 1983), P-COSY (Marion & Bax, 1988b), and NOESY (Jeener et al., 1979) experiments were conducted. Data manipulation for obtaining P-COSY spectra from the raw data was done on an ASPECT 3000 computer using a PASCAL program kindly provided by Dr. L. Orbons. TOCSY spectra were recorded at 298 K with mixing times ranging from 24 to 30 ms. At the same temperature, a set of NOESY spectra was obtained with mixing times of 75, 100, and 175 ms. In addition, a set of spectra was measured at 303 K, which was limited to two NOESY spectra measured for D_2O and H_2O solutions (mixing times of 100 and 150 ms, respectively) and two TOCSY spectra recorded for H_2O and D_2O solutions as well (30- and 40-ms mixing times, respectively). To identify the amide protons that exchange slowly with H_2O, a 100-ms NOESY spectrum was recorded (at 298 K) of a sample of IKe GVP that freshly had been dissolved in D_2O and that previously had been equilibrated in H_2O. In the NOESY(H_2O) experiments, the solvent resonance was suppressed by applying a semiselective jump–return (Plateau & Gueron, 1982) pulse sequence. In the TOCSY(H_2O) measurements, both continuous irradiation and a semiselective jump–return pulse sequence (Bax et al., 1987) were used for this purpose. For experiments in D_2O, only continuous irradiation was applied to suppress the water resonance. In this case, the irradiation frequency was phase-coherent with the carrier frequency (Zuiderweg et al., 1986). Prior to the recording of the 2D experiment, the receiver phase was adjusted so as to minimize base-line distortions (Marion & Bax, 1988a). In all experiments, the carrier was placed at the position of the water signal, and TPPI (Marion & Wüthrich, 1983) was used for signal accumulation in the t_1 dimension. After multiplication with shifted square sine bell or Gaussian functions, the data were Fourier-transformed in the phase-sensitive mode. Zero-filling was used for improving the digital resolution (8.0 Hz/point). Additional base-line corrections were performed using polynomial functions of standard Bruker software installed on a Bruker X32 computer.

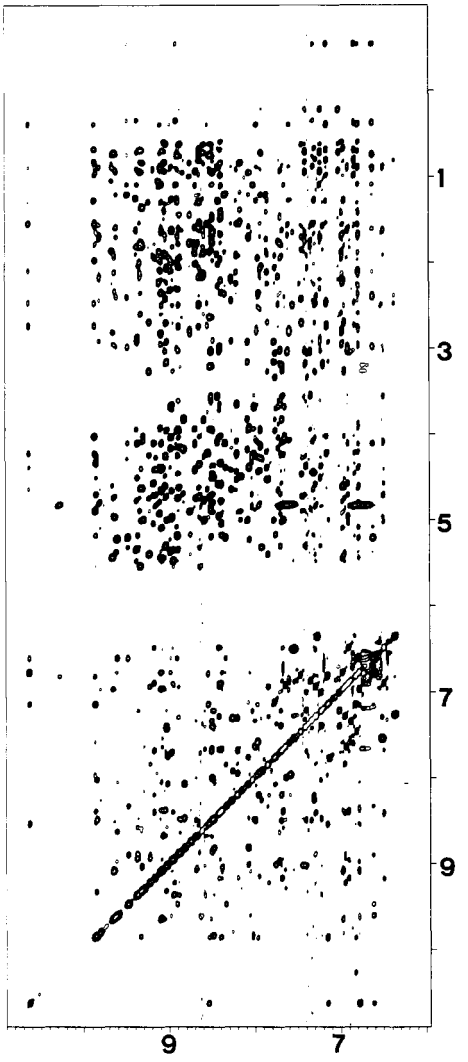

FIGURE 1: 600-MHz NOESY spectrum of IKe GVP dissolved in 90% H_2O/10% D_2O at a concentration of 2.5 mM and a pH of 4.7. The mixing time was 150 ms and the temperature 298 K. The H_2O resonance was suppressed by a jump–return pulse sequence.

Results

General Remarks. ^1H NMR experiments on IKe GVP can only be performed under very specific conditions, i.e., at ≈298 K, at salt concentrations kept as low as possible, and in a restricted pH region (4.6–4.8; de Jong et al., 1987). High salt concentrations cause severe aggregation of the protein, and at pH 6–8, even precipitation may occur. A moderate elevation of the temperature (to 303 K) improves the appearance of the ^1H NMR spectra, but the increased rate of denaturation does not allow measurements beyond a single 2D experiment, which takes 8–16 h on a 600-MHz spectrometer. Therefore, in the present study, the number of spectra recorded at this temperature was limited, but it sufficed to solve all ambiguities encountered in the data set obtained at 298 K. To illustrate the quality of the NOESY data obtained at 298 K, a NOESY spectrum of a 90% H_2O/10% D_2O sample is shown in Figure 1. Due to the relatively short relaxation times of the ^1H spins of IKe GVP, a considerable loss of intensity occurs during the mixing time of TOCSY experiments. For the observation of the resonances of the nonexchangeable protons, at 298 K, mixing times up to 30 ms proved to be feasible (Figure 2); they could be prolonged up to 40 ms at 303 K. The amide reso-

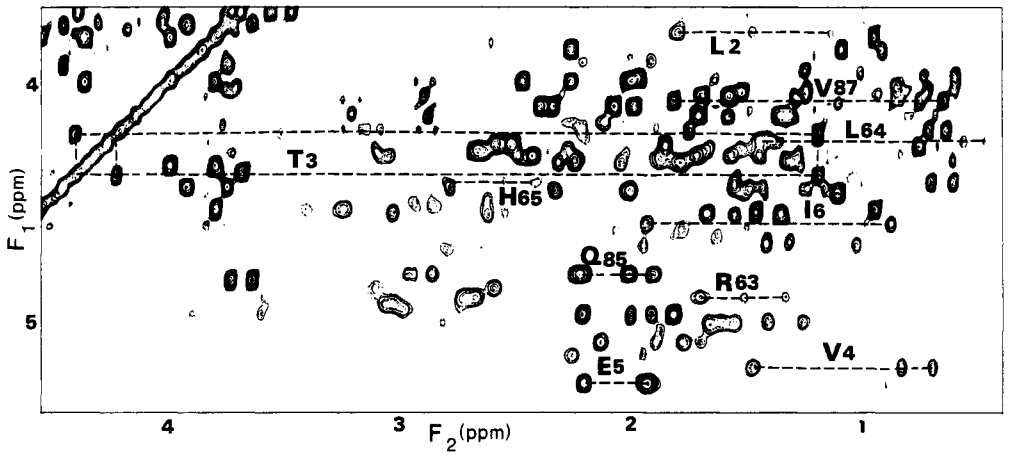

FIGURE 2: 600-MHz clean-TOCSY(MLEV17) spectrum of IKe GVP (1.5 mM, at pH 4.6, 298 K) dissolved in 99.8% D_2O. The mixing time was 30 ms. For clarity, only the spin systems belonging to residues in the triple-stranded β-sheet denoted as III (see text) are indicated.

nances, however, suffer from significant more loss of intensity during the spin lock time. Consequently, shorter mixing times have to be applied when working with H_2O samples (20 and 30 ms at 298 and 303 K, respectively). However, only at 303 K is the mixing time long enough that sufficient relay connectivities from the amide resonances can be observed (Figure 3). In many cases, the information in the COSY data, which are indispensable for discrimination between direct and relayed contacts in TOCSY spectra, is inaccessible. The antiphase components in the cross-peaks significantly reduce their intensity, and overlap of the cross-peaks hampers their interpretation. This problem is inherent to proteins of the size of IKe GVP. Though many long side chain spin systems manifested themselves in TOCSY spectra by extensive cross-peak patterns, the poor resolution of the COSY spectra precludes in many cases discrimination between β and γ resonances. Because of the poor appearance of most fingerpint (Nα) connectivities in H_2O-COSY spectra, residue-specific assignments for most amide resonances were obtained by combining the spin lock patterns observed in H_2O- and D_2O-TOCSY spectra.

Residue-Specific Assignments. In a previous 500-MHz study of IKe GVP (de Jong et al., 1987b), the aromatic ring proton resonances could be attributed to specific amino acids. Also, the sequential and/or residue assignments of five valines (out of nine) and four serines (out of seven) were reported. The quality of the 600-MHz data in the present study allows the assignment of the α and β resonances of all aromatic residues by combining NOE data with AMX patterns that could be recognized in TOCSY and/or COSY spectra. Moreover, we can now identify all but one valine and all but one serine. The quality of the TOCSY data furthermore allowed the unambiguous complete residue assignment of two leucines (out of seven), six glycines (out of seven), and all alanines. The reason for the missing serine and glycine spin systems could, in a later stage of the assignment procedure, be attributed to overlap of resonances.

The long side chain residues arginine and lysine manifest themselves in TOCSY spectra by long side chain connectivity patterns. Although a number of these were poorly resolved, the observation of a relay connectivity from an α resonance to δ or ϵ resonances (which have very characteristic positions around 3 ppm) for several residues enabled discrimination from other long side chain patterns. In this way, seven out of a total of ten R/K spin systems could be identified. The quality of the COSY data (vide supra) did not, however, permit an unambiguous discrimination between R and K spin systems.

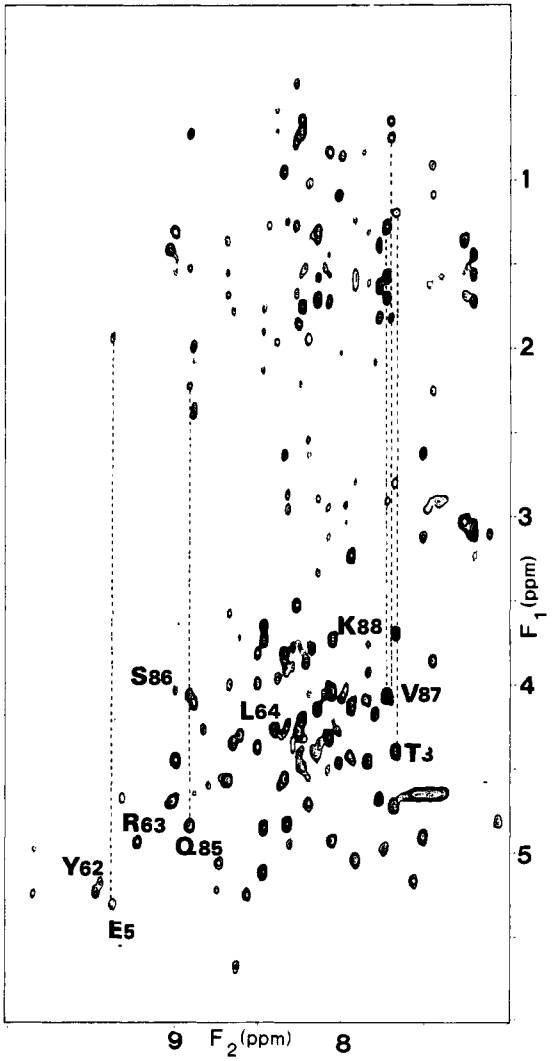

FIGURE 3: 600-MHz clean-TOCSY(MLEV17) spectrum of IKe GVP (1.5 mM, pH 4.6, 303 K) dissolved in 90% H_2O/10% D_2O. The spin systems in the triple-stranded β-sheet structure (denoted as III in the text) are indicated to illustrate the quality of the TOCSY spectrum at these conditions.

At this point in the assignment procedure, the combination of TOCSY and COSY data provided sufficient evidence for the unambiguous residue-specific assignment of an important part of the resonances. The remaining connectivity patterns could be attributed to certain classes of amino acids, e.g., D/N,

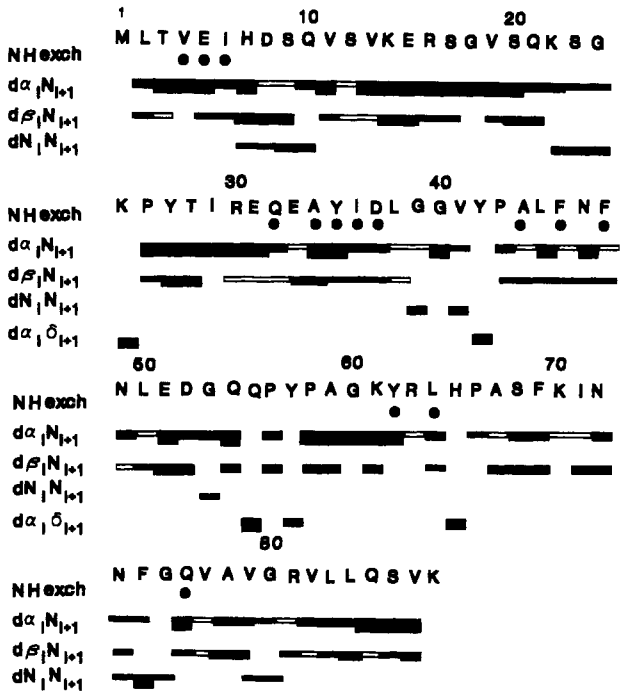

FIGURE 4: Sequentially assigned residues of IKe GVP together with a summary of $d_{\alpha N}$, $d_{\beta N}$, d_{NN}, and $d_{\alpha(\delta\delta')}$ contacts. The indicated NOE intensities correspond with a mixing time of 75 ms for the exchangeable protons and with a 100-ms mixing time for the nonexchangeable protons. Open bars indicate sequential contacts whose assignment and/or intensity determination was obscured by overlap with other resonances (see text). The filled dots indicate slowly exchanging amide protons.

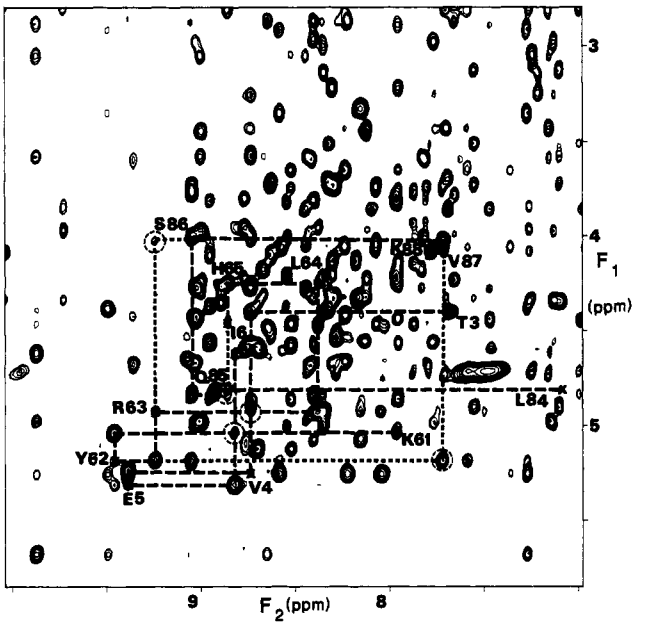

FIGURE 5: Detail of a 125-ms NOESY experiment showing the Nα region. The experiment was performed on an IKe GVP sample (2.5 mM, pH 4.6, 298 K, 90% H_2O/10% D_2O). The sequential walks for three consecutive sequences (3–6, 61–65, 84–88) are indicated by long dashes. These three strands form a triple-stranded β-sheet structure which can be deduced from the observation of several nonsequential $d_{\alpha N}$ contacts. These contacts, which correspond to the skewed arrows in Figure 7, have been indicated with short dashes (if no overlap with the long dashes would occur). In addition, these contacts are marked with dashed circles.

Q/E, and "other" long side chain residues. Many of these spin systems had either incomplete or unresolved cross-peak patterns in the TOCSY and/or COSY spectra. Unambiguous statements about their identity had to await sequential assignments.

Sequential Assignments. Sequential analysis of NMR data through the customary assignment strategy (Wüthrich, 1986) resulted in the nearly complete interpretation of the ^1H NMR spectrum. A schematic representation of the results obtained this way is given in Figure 4. One can observe that several sequential $d_{\alpha N}$ walks are obscured by overlap. In most cases, one can get around this problem by using other sequential contacts, i.e., $d_{\beta N}$ and/or $d_{\gamma N}$ connectivities. A brief outline of the sequential analysis will be given below; cases in which the assignments were less straightforward will be described in more detail. The sequential walk has been divided in separate steps based on the prevalence of $d_{\alpha N}$ or d_{NN} contacts and the involvement of these amino acid sequences in secondary structure elements (which will be defined more precisely at a later stage). We note that the sequential connectivities depicted in Figure 4 correspond to a H_2O-NOESY spectrum recorded with a rather short mixing time of 75 ms. At prolonged mixing times, more connectivities can be observed, and though very helpful in the actual analysis, the contribution of spin diffusion to the cross-peaks is significant. Therefore, to give a clearer picture of the relative intensities of the sequential contacts, we only present NOE data corresponding to this relatively short mixing time.

M1. Assignment of the N-terminal residue can in general not be based on sequential contacts. After completion of the sequential analysis, one long side chain pattern was found for which no sequential assignment has been made. By elimination, we have ascribed this pattern to M1.

Residues 2–6. The N-terminal part of IKe GVP can be assigned in a straightforward manner (Figure 5) by making a $d_{\alpha N}$ walk which was, in most cases, secured by sequential $d_{\beta N}$ connectivities. This stretch of amino acids contains several residues which had firmly been characterized in the residue-specific analysis (e.g., L2, T3, V4). Furthermore, T3 could directly be assigned in a sequence-specific manner because in the amino acid sequence of IKe GVP only two threonines are present of which one had already been assigned to T28 in an earlier 500-MHz study of this protein (de Jong et al., 1989a). These considerations make the assignment of other residues in this sequence rather unambiguous.

Residues 7–10. The residues in this sequence are likely to adopt a turnlike structure as is reflected in the strong d_{NN} connectivities that prevail in this sequential walk.

Residues 11–15. Starting with V11, this sequential walk is strongly dominated by $d_{\alpha N}$ contacts. The assignment of V13 has not been without difficulty. Both the $d_{\alpha N}$ and the $d_{\beta N}$ connectivities are severely overlapping with other connectivities. After completion of the sequential analysis, however, this residue could be identified without ambiguity by elimination.

Residues 16–29. The sequential assignment of this amino acid sequence has been described in detail in a 500-MHz study (de Jong et al., 1989a). We note in passing that the improved performance of the 600-MHz spectrometer enabled us to make additional assignments for some side chain resonances (Table I).

Residues 30–38. The β-sheet structure that has been described for residues 10–29 (de Jong et al., 1989a) seems to be extended into the sequence running from residues 30 to 38. This is reflected in a sequential walk dominated by $d_{\alpha N}$ contacts.

The sequential walk from E33 to A34 was rather cumbersome. The α resonance of E33 is close to the H_2O resonance which makes observation of intra- and interresidue $d_{\alpha N}$ contacts rather difficult. In addition, the intra- and interre-

Table I: Resonance Positions of Identified Spin Systems of IKe GVP at 298 K and pH 4.6[a]

	chemical shifts (ppm)					chemical shifts (ppm)			
	NαH	CαH	CβH	others		NαH	CαH	CβH	others
M1		3.79	2.02	$C_\gamma H$ 2.26, 2.48	L45	8.38	4.46	1.64	$C_\gamma H$ 1.34; $C_\delta H$ 0.68, 0.44
L2	9.12	3.86	1.88[b]	$C_\gamma H$ 1.23; C_δ 0.31, 0.45	F46	9.56	5.32	3.43, 2.97	$C_\zeta H$ 7.34, $C_\delta H$ 7.02, $C_\epsilon H$ 6.64
T3	7.72	4.48	4.31	$C_\gamma H$ 1.27	N47	8.65	5.33	2.09, 2.34	$N_\gamma H$ 7.42
V4	8.80	5.32	1.54	$C_\gamma H$ 0.78, 0.91	F48	8.12	5.01	3.10, 3.20	$C_\zeta H$ 6.82, $C_\epsilon H$ 6.69
E5	9.46	5.39	2.01	$C_\gamma H$ 2.30	N49	8.42	5.01	2.78	
I6	8.90	4.70	2.02	$C_\gamma H$ 0.95	L50	8.48	4.41	1.42	$C_\zeta H$ 1.42; $C_\delta H_3$ 0.80, 0.49
H7	9.95	5.05	3.65, 3.13	C^2H 8.37, C^4H 7.24	E51	8.81	4.47	2.13	$C_\gamma H$ 2.26
D8	9.07	4.37	2.76		D52	8.97	4.34	2.63	
S9	8.52	4.35	4.13, 3.93		G53	8.81	3.65, 4.09		
Q10	8.03	4.51	2.48[b]	2.96	Q54	7.90	4.25	2.18	$C_\gamma H$ 2.39
V11	7.54	3.93	2.34	$C_\gamma H$ 1.17, 0.99	Q55	8.73	4.37	1.76	$C_\gamma H$ 2.15, 2.39
S12	7.76	4.48	3.77		P56		4.16	1.87	$C_\delta H$ 3.65, 3.28
V13	8.40	4.36	1.94	$C_\gamma H$ 0.86	Y57	10.12	4.40	2.54, 2.87	$C_\delta H$ 7.33, $C_\epsilon H$ 6.86
K14	9.09	4.53	1.62	$C_\gamma H$ 1.25, 1.32; $C_\delta H$ 1.52, $C_\epsilon H$ 2.87	P58		4.54	2.42, 2.09	$C_\delta H$ 4.04, 3.92
E15	8.59	5.20	1.86, 1.76	$C_\gamma H$ 2.18, 1.99	A59	8.45	4.03	1.34	
R16	8.76	4.66	1.77, 1.63	$C_\gamma H$ 1.45, $C_\delta H$ 3.13	G60	8.84	4.31, 3.88		
S17	8.59	4.95	3.83, 3.74		K61	8.03	5.12	1.74	
G18	8.03	3.83, 3.74			Y62	9.53	5.26	2.32, 2.79	$C_\delta H$ 6.72, $C_\epsilon H$ 6.57
V19	8.37	4.29	1.84	$C_\gamma H$ 0.82, 0.74	R63	9.32	5.01	1.78	$C_\gamma H$ 1.58, 1.40
S20	8.61	4.45	4.09, 3.90		L64	8.46	4.32	1.53	$C_\gamma H$ 1.39; $C_\delta H$ 0.56, 0.65
Q21	9.01	4.19	2.17, 2.09	$C_\gamma H$ 2.49, 2.44	H65	8.93	4.52	2.88, 2.65	C^2H 7.84, C^4H 7.07
K22	8.25	4.23	1.81	$C_\gamma H$ 1.46, 1.42; $C_\delta H$ 1.68; $C_\epsilon H$ 2.99	P66		4.42	2.48[c]	$C_\gamma H$ 1.90, $C_\delta H$ 3.59
S23	7.94	4.55	4.01, 3.85		A67	11.35	4.64	1.54	
G24	8.14	4.13, 3.83			S68	8.81	4.64	3.32, 3.51	
K25	7.88	4.76	1.73	$C_\epsilon H$ 3.03	F69	7.83	5.07	2.81, 3.06	$C_\zeta H$ 7.20, $C_\delta H$ 6.99, $C_\epsilon H$ 6.83
P26		5.08	2.32, 1.92	$C_\gamma H$ 2.10, 2.01; $C_\delta H$ 4.00, 3.70	K70	9.11	4.74		
Y27	8.42	4.92	3.05, 2.95	$C_\delta H$ 6.92; $C_\epsilon H$ 6.54	I71	8.30	4.74	2.04	$C_\gamma H$ 1.11
T28	8.46	4.65	3.90	1.04	N72	9.43	4.75	3.67, 2.75	
I29	9.04	4.13	1.62	$C_\gamma H$ 1.31, 1.04; $C_\gamma H_3$ 0.82; $C_\delta H_3$ 0.70	N73	8.48	4.40	2.60, 2.52	
R30	8.71	4.73	1.40, 1.51	$C_\gamma H$ 1.63, 1.86; $C_\delta H$ 3.32, 3.22; $N_\epsilon H_2$ 7.28	F74	7.77	4.82	2.89, 3.50	$C_\zeta H$ 7.39, $C_\delta H$ 7.32, $C_\epsilon H$ 7.22
E31	8.83	5.14	1.74	$C_\gamma H$ 2.09, 1.97	G75	8.31	4.13, 3.73		
Q32	8.22	4.50			Q76	7.65	5.26	2.33, 2.04	$C_\gamma H$ 2.58, 2.44
E33	8.31	4.84			V77	9.10	4.50	1.63	$C_\gamma H$ 0.68, 0.81
A34	8.10	4.56	1.16		A78	9.10	4.76	1.40	
Y35	8.72	5.76	2.78, 2.95	$C_\delta H$ 6.91, $C_\epsilon H$ 6.63	V79	8.51	4.05	2.05	$C_\gamma H$ 0.67, 0.80
I36	9.95	5.33	1.34	$C_\gamma H$ 0.45	G80	9.07	3.52, 4.30		
D37	8.30	4.61	2.73, 2.93		R81	8.19	4.41	1.80	$C_\gamma H$ 1.65, 1.53; $C_\delta H$ 3.18, 3.14
L38	8.36	4.58	1.91	$C_\gamma H$ 1.64; $C_\delta H$ 0.58, 0.78	V82	8.23	3.41	1.80	$C_\gamma H$ 0.20, 0.95
G39	8.36	4.41, 3.80			L83	8.21	4.42	1.39[c]	
G40	8.70	3.80, 4.18			L84	7.15	4.91	0.37	$C_\gamma H$ 0.37; $C_\delta H$ −0.72, 0.11
V41	8.39	3.58	1.74	$C_\gamma H$ 0.48, 0.87	Q85	8.99	4.91	1.98	$C_\gamma H$ 2.10, 2.31
Y42	7.62	4.98	3.20, 2.69	$C_\delta H$ 7.16, $C_\epsilon H$ 6.81	S86	9.12	4.11	3.83	
P43		4.73	2.77, 1.88	$C_\delta H$ 3.81, 3.68	V87	7.79	4.15	1.90	$C_\gamma H$ 0.72, 0.83
A44	9.15	4.73	1.50		K88	7.85	4.16	1.64	$C_\gamma H$ 1.36, $C_\delta H$, 1.78, $C_\epsilon H$ 2.96

[a] All chemical shifts are measured relative to TSP. [b] A second β resonance might be present, but so far the COSY data do not allow discrimination of this resonance from a γ resonance. [c] NMR data obtained so far could not provide unambiguous evidence for the nondegeneracy of this resonance.

sidue $d_{\beta N}$ contacts were poorly defined. After elucidation of the secondary structure of the amino acid sequence comprising these residues, more evidence for these assignments was provided by nonsequential $d_{\alpha N}$ connectivities (vide infra).

Residues 39–42. In the 500-MHz study mentioned before, the sequential assignment of V41 and Y42 was described (de Jong et al., 1989a). The identification of G40 was rather straightforward as intense sequential $d_{\alpha N}$ and d_{NN} contacts were observed. The assignment of G39 was rather troublesome, however, because of overlap of the amide resonances of G39 and Leu-38 and the overlap of two α resonances of G40 and G39. After completion of the sequential walk, G39 could be assigned by elimination.

Residues 42–49. The sequential walk could be continued with a $d_{\alpha(\delta\delta')}$ contact between Y42 and P43. A strong $d_{\alpha N}$ contact between the latter residues forms the start for a straightforward $d_{\alpha N}$ walk from L44 to L49. We note that the assignment of F48 is at variance with a previous residue assignment (de Jong et al., 1987b). On the basis of the upfield position of the ring proton resonances, these were ascribed to a tyrosyl residue (denoted as Tyr-V). The unambiguous sequential assignment of these resonances to F48 necessarily implies that a set of more downfield aromatic ring protons that were previously ascribed to a phenylalanine residue [denoted as Phe-IV in de Jong et al. (1987)] must originate from a tyrosyl residue (Y57, vide infra).

Residues 50–59. Continuation of the sequential walk through residues 50–54 does not encounter any difficulty. Strong $d_{\alpha N}$ and d_{NN} connectivities are observed, accompanied by a $d_{\alpha_i N_{i+2}}$ contact between D52 and Q54. The sequence P56-Y57-P58 could be assigned after all tyrosyl residues, but one had been identified. The remaining tyrosyl residue (Y57) turned out to be the residue previously (de Jong et al., 1987b) assigned as Phe-IV (vide supra). The $d_{\alpha N}$ and $d_{\alpha(\delta\delta')}$ contacts to Y57 provide for the assignment of the α resonance of P58 and the δ resonances of P56, respectively.

Residues 60–71. Sequential assignments of the sequence encompassing residues 60–70 are, in all cases, based on sequential $d_{\alpha N}$ contacts. In one case, L64, the $d_{\alpha N}$ contact was overlapping with other connectivities. Observation of $d_{\beta N}$, $d_{\gamma N}$, and $d_{\delta N}$ contacts to H65 makes the sequential assignment of this residue rather safe. The identity of this histidyl residue could be simply established because the only other histidine was already assigned to H7 (vide supra). Like the other proline residues (vide supra), overlap problems in the aliphatic region limit the assignment of the spin system of P66 to its α and δδ′ resonances.

Residues 72–74. The sequential assignment of this sequence is seriously hampered by resonances overlapping with each other and with the H_2O resonance. Combination of data

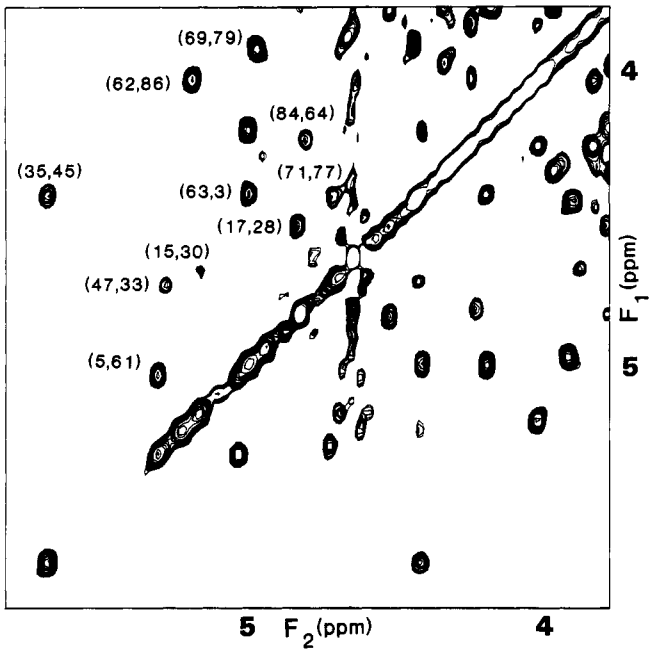

FIGURE 6: Detail of the CαH region of a 125-ms NOESY experiment performed on a 1.5 mM IKe GVP sample (pH 4.6, 303 K). Long-range $d_{\alpha\alpha}$ contacts which characterize the antiparallel β-sheet structures have been indicated by their ω_2,ω_1 coordinates.

obtained at different temperatures (298 and 303 K) and additional observation of sequential side chain to amide connectivities nevertheless permit a firm assignment of this stretch of residues.

Residues 75–88. Strong $d_{\alpha N}$ connectivities dominate the sequential walk in this amino acid sequence. An exception is formed by the segment R80-G81 in which a strong d_{NN} contact is observed, probably indicative of a turn or bend in a β-sheet structure. The complete aliphatic spin system of S86 is overlapping with the $d_{\alpha\alpha}$ connectivity of G24. This spin system can be assigned, however, by the straightforward sequential walk Q85-S86-V87 (Figure 5). The uniqueness of this stretch of residues in the amino acid sequence of IKe GVP allows the identification of the aliphatic resonances of S86 by their intra- and interresidue $d_{\alpha N}$ and $d_{\beta N}$ contacts with V87.

Secondary Structure Elucidation. The manifestation in NOESY spectra of the close proximity of particular backbone hydrogen atoms enables the elucidation of secondary structure elements in proteins (Wüthrich, 1986). In D_2O-NOESY spectra, several $d_{\alpha\alpha}$ contacts can be observed (Figure 6). The mere observation of these contacts is indicative of the presence of antiparallel β-sheet structure. Scrutiny of these and other NOE data resulted in elucidation of two β-loops which partly combine into a triple-stranded β-sheet (structure I), one β-loop (structure II) and one triple-stranded β-sheet (structure III). Figure 7 depicts the nonsequential $d_{\alpha N}$ and d_{NN} contacts which, together with $d_{\alpha\alpha}$ connectivities, define a regular antiparallel β-sheet structure for the amino acid sequence running from 13 to 50 (structure I). In addition, also numerous long-range contacts from side chain to backbone and to other side chain protons were observed. We note that the ambiguity in the assignment of Q32 and E33 (due to overlap problems) is removed by several nonsequential amide to α contacts. As is indicated in Figure 7, the β-ladder in the β-loop encompassing residues 31–49 comprises residues 37 and 43. The $d_{\alpha\alpha}$ contact between these residues is not visible due to the close proximity of both αH resonances to the H_2O signal. Because of overlap, a nonsequential $d_{\alpha N}$ contact between P43 and L38 that can be expected when these residues are involved in an extended β-ladder is not clearly discernible. Clearly distinguishable, however, are side chain to side chain contacts between D37 and P43. The observation of a nonsequential $d_{\alpha N}$ contact between D37 and L44 provides additional evidence for the involvement of D37 and P43 in the β-sheet structure. Structure II consists of a β-loop formed by residues 69–79. The identification of the connectivities of this sequence has been rather troublesome due to overlap problems. Both the sequential assignment and structure elucidation obtained from the spectra recorded at 298 K could, however, be reproduced using 2D spectra recorded at 303 K where, for this stretch of residues, overlap was less severe. The same self-consistency that secures both the sequential assignment and the secondary structure elucidation of structures I and II is present in structure III. This structure is composed of the N-terminal end (residues 2–6) and the C-terminal end (residues 84–87) between which the amino acid sequence running from residues

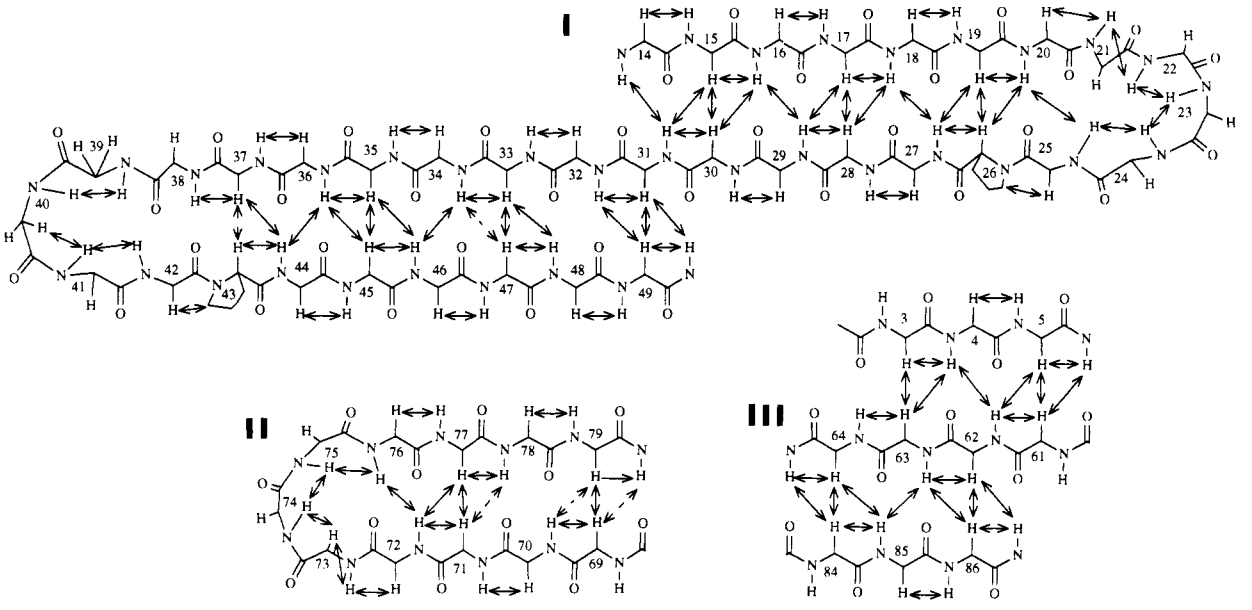

FIGURE 7: Schematic representation of the antiparallel β-sheet structures I–III (see text) in IKe GVP. Sequential (horizontal) and nonsequential (skewed) $d_{\alpha N}$ contacts and long-range d_{NN} and $d_{\alpha\alpha}$ contacts are indicated by arrows. The interrupted arrows indicate contacts whose observation is troubled by overlap.

61 to 65 is sandwiched. Its secondary structure elucidation is, in addition to the backbone to backbone connectivities depicted in Figure 7, secured by numerous side chain to side chain and side chain to backbone contacts.

The antiparallel β-sheet structures are reflected in the pattern of slowly exchanging amide protons in the amino acid sequence (Figure 6). Most amide resonances disappear within a few hours, and therefore relatively few slowly exchanging amides can be observed in a NOESY experiment. Most of them correspond to amide protons that are expected to be involved in hydrogen bonding as indicated by the β-sheet structure drawn in Figure 7. In the β-loops running from residues 15 to 30 and from residues 69 to 79, no slowly exchanging amide protons can be observed. This indicates an increased flexibility for these parts of the molecule, a feature also encountered in the homologous structural elements in Y41H GVP (Folkers et al., 1991).

Besides the three antiparallel β-sheet structures, other structure elements can be recognized. The observation of $d_{\alpha_i N_{i+2}}$ contacts between H7 and S9 and between D52 and Q54 accompanied by strong d_{NN} contacts in the sequences 7–10 and 52–54 indicates that these residues are involved in turn or half-turn structures. The incompleteness of data obtained so far (i.e., the lack of J coupling data and overlapping NOE connectivities) hampers further classification of these elements into any of the regular turn structures (Wüthrich, 1986).

Discussion

Comparison of IKe GVP and M13 GVP (Folkers et al., 1991) shows that their secondary structure is almost completely conserved. In Figure 8A, the $d_{\alpha\alpha}$ contacts observed in the NOESY spectra of IKe and M13 GVP have been depicted in an adjacency matrix. The stretches of $d_{\alpha\alpha}$ contacts running perpendicular to the diagonal are highly characteristic of β-sheet structures. Inspection of the adjacency matrix shows that the polypeptide chains of M13 and IKe GVP are almost completely identically folded in antiparallel β-sheet structures. Note that the insertion of Q21 in the sequence of IKe GVP does not distort the β-sheet structure in this region. When the other regions of the β-sheet structures of these proteins are studied in more detail, however, two minor differences can be observed. First, the β-ladder in the β-loop running from residues 13 to 31 in M13 GVP does not seem to extend beyond residues 14 and 31 in IKe GVP. This is indicated by the absence of a $d_{\alpha\alpha}$ contact in IKe between V13 and Q32 while a strong $d_{\alpha\alpha}$ contact exists between F13 and E31 in M13 GVP. Second, the β-ladder in the β-loop running from residues 31 to 50 comprises residues D37 and P43 as is indicated in Figure 7. In M13 GVP, however, the β-ladder does not involve the homologous residues D36 and P42. So far, the lack of detailed information about the three-dimensional solution structure of these proteins precludes any explanation of these observations.

The occurrence of IKe GVP as a dimer in solution imposes a special problem on its structure determination by NMR. As only one set of resonances is observed in the ^{1}H NMR spectrum, the respective monomers must be magnetically equivalent, and we expect the protein to have a 2-fold axis of symmetry similar to that proposed for the crystal structure of M13 GVP (Brayer & McPherson, 1983). If, for the moment, we disregard the loop sections drawn in Figure 7, these considerations do not have consequences for the β-sheet folding proposed above. However, the equivalence of the monomers in the NMR spectrum has repercussions for distinguishing β-sheets with strands originating from within the monomer or from the combination of the symmetry-related counterparts. We feel confident, however, that all contacts depicted in Figure 7 are intramonomer. This is based on the folding of the monomer chain in the aforementioned crystallographic structure proposed for M13 GVP. Also, the crystallographic data indicate a structure predominantly consisting of antiparallel β-sheet. In Figure 8B, the intramonomer $d_{\alpha\alpha}$ distances (shorter than 4 Å) deduced from the crystallographic structure of M13 GVP and the $d_{\alpha\alpha}$ connectivities observed in IKe GVP by NMR have been plotted in the form of an adjacency matrix. The overall symmetry observed in Figure 8B leads us to conclude that all $d_{\alpha\alpha}$ contacts as observed by NMR are intramonomer. At close range, some differences between crystallographic data for M13 GVP and NMR data for IKe GVP emerge from adjacency matrix 8B. The differences between the secondary structure of M13 GVP as recently determined by NMR and that derived from X-ray diffraction data have been discussed in detail (Folkers et al., 1991). We remark that the structural elements in IKe GVP that are at variance with the crystal structure are identical in the M13 and IKe GVP structures that were determined by NMR. These results support the conclusion that it is these secondary structure elements that are present in the solution structures. The observed anomaly between crystallographic and NMR data also bears consequences for a modeling study that has

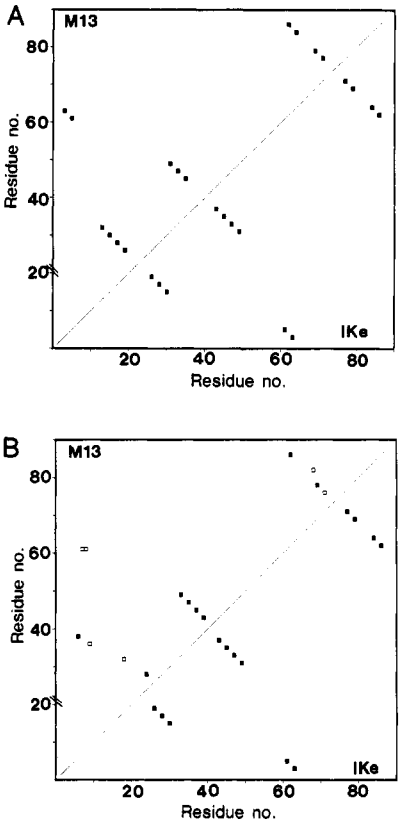

FIGURE 8: (A) Adjacency matrix showing, below the diagonal, the observed $d_{\alpha\alpha}$ contacts (closed squares) in a 100-ms NOESY spectrum of IKe GVP (1.5 mM, 298 K, pH 4.6). The closed squares above the diagonal indicate the $d_{\alpha\alpha}$ contacts that were observed in an NOESY experiment performed on M13 GVP [data taken from Folkers et al. (1991)]. Note that on the vertical axis the "extra" amino acid residue at position 21 of the amino acid sequence IKe GVP has been taken into account. (B) Adjacency matrix showing, above the diagonal, the intramonomer $d_{\alpha\alpha}$ contacts in the crystal structure of M13 GVP [no intermonomer $d_{\alpha\alpha}$ contacts (<4.0 Å) can be observed]. Data were taken from the Protein Data Bank. Closed and open squares correspond to distances smaller than 2.5 and 4.0 Å, respectively. Below the diagonal, $d_{\alpha\alpha}$ contacts are indicated that were observed in a 100-ms NOESY experiment performed on IKe GVP (1.5 mM, 298 K, pH 4.6). The diagonal symmetry indicates that the $d_{\alpha\alpha}$ contacts observed by NMR are also intramonomer.

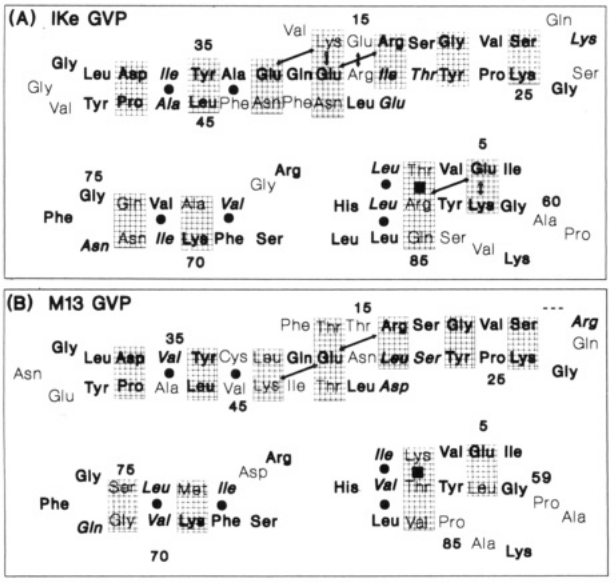

FIGURE 9: Comparison of the β-sheet structures derived for (A) IKe GVP and (B) M13 GVP. Residues that are conserved in IKe and M13 GVP are indicated in boldface. Residues for which the replacement was conservative with respect to their polar, apolar, or charged character are indicated in boldface italic. The boxes enclose residues that are pointing toward the reader. Pairs of hydrophobic/apolar residues pointing in the opposite direction and occupying the other side of the antiparallel aligned strands are indicated by filled circles. These residues are, most probably, pointing toward the interior of the protein. The squares refer to an interchange of positively charged residues K3 in M13 GVP and R63 in IKe GVP for threonine residues T62 in M13 GVP and T3 in IKe GVP.

been undertaken for IKe GVP using the crystallographic structure of M13 GVP as point of departure (Brayer, 1987).

The homology between M13 and IKe GVP becomes more apparent when the secondary structures of the proteins are considered in more detail (Figure 9). Apart from the conserved residues, there are many semiconserved residues at identical sites in the secondary structure. This is, for example, indicated by the presence of several conserved pairs of hydrophobic residues. The typical character of antiparallel β-sheet structures is reflected in the alternating occurrence of these hydrophobic patches which are most likely part of the interior of the protein. It is noted that most of the charged residues of the β-sheet structures are located at sides opposite of those occupied by the hydrophobic residues. Many of the positively charged residues, likely candidates for electrostatic interactions with ssDNA or intramolecular salt bridges, are conserved or semiconserved (R21 in M13 GVP and K22 in IKe GVP). An exception that seems to prove this rule can be found in residues R63 in IKe GVP and K3 in M13 GVP, which are both replaced by threonine residues. As indicated in Figure 9, the positively charged residues and the threonine residues are counterparts in the antiparallel β-sheet structure. The T3–R63 pair in IKe GVP and the K3–T62 pair in M13 GVP are both connected by $d_{\alpha\alpha}$ contacts, indicating that these positively charged residues are conserved in a spatially confined region. Also worth mentioning is that amino acid sequences which make up the turns formed by residues 21–24 and 39–42 end with a proline residue (P26 and P43, respectively). These proline residues most likely adopt a trans conformation as is indicated by the observation of sequential $d_{\alpha(\delta\delta')}$ contacts. Interestingly, these proline residues (and their trans conformation) are conserved in M13 GVP (P25 and P42, respectively). This indicates that these residues probably have a function in stabilizing the structures of the turns in these regions.

It is likely that the homology between the secondary structures of M13 and IKe GVP may be extrapolated to their tertiary structures. Experimental evidence is available that homologous residues are exposed to the solvent. Such information can be extracted from earlier work on the titration properties of histidines and from photo-CIDNP experiments. The histidyl residues now identified as H7 and H65 possess pK values (de Jong et al., 1987) indicative for exposure of the first to the solvent and burial within the interior of the second residue. Comparison with H64 in M13 GVP shows that this residue has a very low pK which indicates that also this residue is buried in the interior of the protein (IKe H7 is not conserved in M13 GVP so no statement can be made about this residue). Furthermore, photo-CIDNP experiments have shown that three (out of five) tyrosyl residues in IKe GVP are exposed to the solvent (de Jong et al., 1987). In previous work, sequential assignment of only two of these residues, Y27 and Y42, was possible. The sequential assignment of all tyrosyl residues available from the present work shows that the remaining third residue is Y35. From a comparison with data available for M13 GVP (Garssen et al., 1978), it appears that the photo-CIDNP effect observed for these tyrosyl residues is also visible in M13 GVP (Y26, Y34, and Y41). These data suggest that besides a large homology existing between the secondary structure of M13 and IKe GVP, also their spatial folding is similar. We remark that the observed solvent accessibilities of these residues are in agreement with the crystallographic data for M13 GVP. This indicates that, despite some local differences in the secondary structure elements, the overall folding of the molecules is the same in the crystal and in the solution structures.

Recently we showed that the β-loops encompassing residues 16–29 in IKe GVP and 13–32 in M13 GVP were in a similar way involved in ssDNA binding (van Duynhoven et al., 1990). The observed large structural homology in other domains of M13 and IKe GVP suggests that the ssDNA binding properties of IKe and M13 GVP can be based on the same protein–ssDNA interaction model.

ACKNOWLEDGMENTS

We thank Dr. L. Orbons for kindly providing the PASCAL program for processing P-COSY data. Paul Schoenmakers is acknowledged for his technical assistance. The NMR experiments were performed at the SON HF-NMR facility (Nijmegen, The Netherlands). The authors are indebted to its staff for keeping the 600-MHz spectrometer in excellent condition.

REFERENCES

Alma, N. C. M., Harmsen, B. J. M., Hilbers, C. W., van der Marel, G., & van Boom, J. H. (1981) *FEBS Lett. 135*, 15–20.

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.

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

Bax, A., Sklenar, V., Clore, G. M., & Gronenborn, A. M. (1987) *J. Am. Chem. Soc. 109*, 6511–6513.

Brayer, G. D. (1987) *J. Biomol. Struct. Dyn. 4*, 859–868.

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

de Jong, E. A. M., Harmsen, B. J. M., Konings, R. N. H., & Hilbers, C. W. (1987a) *Biochemistry 26*, 2039–2046.

de Jong, E. A. M., Konings, R. N. H., Harmsen, B. J. M., Prinse, C. W. J. M., & Hilbers, C. W. (1987b) *Eur. J. Biochem. 167*, 563–572.

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., 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.

Folkers, P. J. M., van Duynhoven, J. P. M., Harmsen, B. J. M., Konings, R. N. H., & Hilbers, C. W. (1991) *Eur. J. Biochem. 202*, 349–360.

Garssen, G. J., Kaptein, R., Schoenmakers, J. G. G., & Hilbers, C. W. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5281–5285.

Gray, C. W. (1989) *J. Mol. Biol. 208*, 57–64.

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

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

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

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

Marion, D., & Bax, A. (1988a) *J. Magn. Reson. 79*, 352–356.

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

Model, P., & Russel, M. (1988) *The Bacteriophages* (Calendar, R., Ed.) pp 2099–2107, Plenum, New York.

Model, P., McGill, C., Mazur, B., & Fulford, W. D. (1982) *Cell 29*, 329–335.

Peeters, B. P. H., Konings, R. N. H., & Schoenmakers, J. G. G. (1983) *J. Mol. Biol. 169*, 197–215.

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

Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.

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

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Chapters 7–10, Wiley & Sons, New York.

Yen, T. S. B., & Webster, R. E. (1982) *Cell 29*, 337–345.

Zaman, G. J. R., Schoenmakers, J. G. G., & Konings, R. N. H. (1990) *Eur. J. Biochem. 189*, 119–124.

Zuiderweg, E. R. P., Hallenga, K., & Olejnickzak, E. T. (1986) *J. Magn. Reson. 70*, 336–343.