NMR Studies
of the Single-Stranded DNA Binding Proteins
Encoded by the Filamentous Bacteriophages
M13 and IKe

John van Duynhoven

NMR Studies

of the Single-Stranded DNA Binding Proteins

Encoded by the Filamentous Bacteriophages

M13 and IKe

NMR Studies

of the Single-Stranded DNA Binding Proteins
Encoded by the Filamentous Bacteriophages M13 and IKe

een wetenschappelijke proeve op het gebied van de

Natuurwetenschappen

in het bijzonder de chemie.

Proefschrift,

ter verkrijging van de graad van doctor

aan de Katholieke Universiteit Nijmegen,

volgens het besluit van het College van Decanen

in het openbaar te verdedigen op

woensdag 9 september 1992

des namiddags te 1.30 precies

door

Johannes Petrus Martinus van Duynhoven

geboren op 19 juli 1962 te Langenboom.

Promotores: Prof. Dr. C.W. Hilbers
               Prof. Dr. R.N.H. Konings

Aan mijn Ouders.

Dankwoord

Contents

Abbreviations

CD	circular dichroism
CIDNP	chemically induced dynamic nuclear polarization
dsDNA	double-stranded DNA
DOTP	N^1,N^4,N^7,N^{10},-(tetrakis-phosphonomethyl)-tetrazacyclodedecane
DQF-COSY	double quantum filtered correlated spectroscopy
GVP	gene V protein
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	nuclear Overhauser enhancement spectroscopy
RF	replicative form
ssDNA	single-stranded DNA
TEMPO	4-hydroxy-1-oxyl-2,2,6,6-tetramethyl piperidine
TOCSY	total correlation spectroscopy
TPPI	time-proportional phase incrementation
TrNOESY	transferred NOESY
wt	wild-type

CHAPTER 1
General Introduction

PROTEIN-DNA INTERACTIONS

The central role of DNA in the genetic process has been well established. At all levels of DNA function, such as replication, gene expression and regulation, proteins play an indispensable role. Fundamental understanding of these processes requires insight in the mechanism of DNA-protein interaction at the sub-molecular level. Until the last decade, elucidation of the 3-dimensional structure of macromolecules was the monopoly of X-ray crystallography. The development of high resolution multi-dimensional NMR techniques (Ernst et al., 1987), however, paved the way for the determination of the structure of proteins in solution up to 20 kDa in molecular mass (Wüthrich, 1986; 1989). The information gained from NMR is in many ways complementary to that obtained from X–ray crystallography. One of the main advantages of NMR over crystallographic methods is that it creates the possibility to observe macromolecules in solution. This allows the study of the structure, dynamics and intermolecular recognition of biomolecules under physiological (non-crystalline) conditions.

Protein-dsDNA interactions. So far most of our knowledge on DNA-protein interactions pertains to proteins that bind to the DNA double-helix in a sequence-specific manner, e.g. repressors, transcription factors or hormone receptors (for reviews see Freemont et al., 1991; Churchill et al., 1991; Saenger, 1984; Schwabe & Rhodes, 1991). Although an elaborate description of this type of interaction is beyond the scope of this thesis, however, some of its aspects are instructive for the understanding of the ssDNA-protein interaction and a brief outline will be given below. X-ray as well as NMR studies have resulted in models for several complexes of dsDNA binding proteins with their target sequences. Though their structural context differs considerably over a wide variety of dsDNA binding proteins, the following recurring structural motifs can be recognized:

● Helix-turn-helix: This element consists of two α-helical structures which are connected by a tight β-turn. In the DNA-protein complex, one helix, termed the recognition helix, lies in the major groove of the DNA double helix and is involved in the recognition of the target sequence.

● <u>Zinc finger:</u> Two types of zinc fingers have been discovered. In the first one, commonly denoted as C_2-H_2, a two-stranded antiparallel β-sheet and an α-helix are held together by a zinc atom which is coordinated (in a tetrahedral manner) to the sulphur atoms of two cysteine residues in the β-sheet and the nitrogen atoms of two histidine residues in the helix. In the second type of zinc finger, designated as C_2-C_2, an α-helix is anchored to the preceding N-terminal segment by tetrahedral coordination of four cysteine sulphur atoms with a zinc atom.

● <u>Leucine zipper:</u> This is a dimer consisting of two α-helices each preceded by basic N-terminal regions. The dimerization domain is formed by the interdigitization of stretches of leucine residues lying on the same side of the helical structures (the "zipper"). The N-terminal ends recognize specific sequence targets in the DNA; the sole function of the helices is to provide a dimerization interface.

● <u>β-ribbon:</u> A planar structural element, mainly consisting of antiparallel β-sheet, which provides a suitable framework for residues involved in the specific recognition of dsDNA target sequences.

Although these elements vary significantly with respect to the structure and the mechanism of DNA recognition two common features can be distinguished. In all structural motifs positively charged amino acid residues are present that are involved in electrostatic interactions with the negatively charged phosphate-backbones of the DNA duplex. These interactions are also responsible for the non-specific interaction of dsDNA binding proteins with non-cognate sequences. The major interaction for sequence specificity is thought to be the hydrogen bonding between amino acid residues and functional groups in the major groove of the DNA duplex (von Hippel & Berg, 1989).

The classification of the sequence-specific dsDNA binding proteins in terms of structural motifs can be held up as a model for other types of DNA binding proteins. In the next sections the current knowledge of the mechanism by which proteins interact with ssDNA is summarized. It will be shown that for this type of interaction the recognition of structural motifs is still in its onset.

Protein-ssDNA interactions. For several retroviral proteins the application of a zinc-finger motif for sequence-specific anchoring of single-stranded RNA to the cell membrane ("budding") has been reported (for a review see Summers, 1991). Based on amino acid sequence homology the same interaction model has been proposed for an eukaryotic sequence-specific ssDNA binding protein involved in gene repression (Rajavashisht et al. 1989).

Most of our knowledge regarding the single-stranded DNA binding proteins, however, pertains to their non-specific interaction. These proteins play indispensable roles whenever unwinding of the DNA double helix and subsequent stabilisation of the displaced DNA strand is required (as in DNA replication, recombination and repair). A number of eukaryotic non-specific ssDNA binding proteins has been described (reviewed by Falashi et al., 1980) but these have been characterized in biochemical terms only. For the prokaryotic (*Escherichia coli*) ssDNA binding proteins SSB (Sancar et al., 1981; Lohman et al., 1988) and RecA (Sancar et al., 1980) the ssDNA binding characteristics have been described very well in thermodynamic terms. At the same level also the helix-destabilizing proteins encoded by the phage T4 (gene-32 protein, Newport et al., 1981; Lohman et al., 1988) and the filamentous *Pseudomonas* phage Pf1 (Plyte & Kneale, 1991) have been studied. Compared to these viral systems the genomes of the M13 and IKe phages code for relatively few gene products and especially the ssDNA binding proteins can be isolated in large quantities. For this reason, and more importantly, because of their relatively small size, they lend themselves for high resolution structural studies. Moreover, many of the ssDNA binding proteins have shown to play an essential and sequence-specific function in gene regulation. These aspects will be discussed in the next section which describes the most important characteristics of the filamentous phages.

THE FILAMENTOUS BACTERIOPHAGES

General aspects. The filamentous bacteriophages form a class of non-lytic, single-stranded DNA viruses (Inoviridiae) that share morphological and structural features (for a review see Model & Russel, 1988). The thread-like phage particles contain a circular, covalently closed, single-stranded DNA strand molecule encapsulated in a tube-like protein coat. The genome of the filamentous phages of subclass I is packed into the protein tube as a right-handed helix

with the bases pointing towards the helix axis (Day, 1988). The *Escherichia coli* phages M13, f1, fd and IKe are its best known members. The phages M13, fd and f1, often collectively referred to as Ff, have genomes with a high degree of homology (>98%) (Model & Russel, 1988). The amino acid sequences of their gene V encoded ssDNA binding proteins are identical. Although their genetic organization is almost identical, the identity between the genomes of the evolutionary distantly related phages Ff and IKe is less pronounced (55%) (Peeters et al., 1983).

In bacteriophages of subclass II, the bases of the viral DNA strand point towards the protein coat. Also with respect to their genetic organization they differ from the Ff and IKe phages. Subclass II comprises the *Pseudomonas* phages Pf3 and Pf1 and the *Xanthomonas* phage Xf.

The replication cycle of the phages of subclass I, and that of M13 in particular, has been studied best and is summarized below. The homology in nucleotide sequence and genetic organization of the phages of this subclass indicates that most of the replication characteristics of M13 also pertain to those of the others.

The M13 replication cycle. Since its first isolation from the sewage of München in 1963 (Hofschneider & Preuss, 1963), the filamentous bacteriophage M13 has been the subject of many fundamental and applied studies. It is considered as an important model system for the elucidation of the fundamentals of genetic processes and its DNA has shown to be an extremely versatile expedient in all kinds of studies. The characteristics of its life-cycle have been investigated elaborately (for a review see Model & Russel, 1988) and are only briefly outlined below.

After binding of the filamentous phage particle to the tip of a conjugative pilus of the host bacterium, the viral DNA strand (vsDNA) is injected into the cell. The polarity of the single-stranded vsDNA is identical to that of the phage encoded mRNAs. The replication of the viral strand can be subdivided into three stages: (1) conversion of vsDNA into the double-stranded replicative form, (2) multiplication of the replicative form via rolling circle replication, and (3) synthesis of progeny vsDNA.

During the first stage, host enzymes convert the vsDNA into a circular duplex, henceforth referred to as RF DNA. Transcription of the RF DNA template results in the

synthesis of ten phage encoded proteins. One of these, the gene II protein, is a site specific topoisomerase that nicks the viral strand in RF DNA, thereby initiating the second stage of viral DNA replication according to the rolling circle DNA replication mechanism (originally proposed by Gilbert & Dressler, 1968).

As long as the concentration of the phage encoded ssDNA binding protein (gene V protein, GVP) is low the expelled vsDNA is converted into RF DNA by the process described. However, when the GVP has reached a certain threshold concentration, it starts to bind to the displaced viral strand thereby preventing its conversion into RF. The requirement of a threshold concentration for the formation of the nucleoprotein assemblies is related to the cooperative binding characteristics of GVP (Alma et al., 1983b). Via sequestering of the viral strand an increase in the number of RF molecules per cell is prevented (Stage 3). The switch thus maintains the number of RF molecules at a homeostatic constant level and prohibits the accumulation of toxic concentrations of phage products.

In addition to its indispensable function in the regulation of viral strand synthesis, GVP also fulfils an important function in the regulation, at the level of translation, of the proteins encoded by the genes I, II, III, V and X (Zaman, 1991; Zaman et al., 1991b). Both genetic and molecular biology studies have demonstrated the specific binding of GVP to unique mRNA target sequences. The concerted action of DNA sequestering and gene regulation have proposed to be the integral parts of a fine-tuned regulatory circuit in which GVP plays a central role. In this thesis we will focus exclusively on the phage M13 and IKe encoded GVPs as model systems for the elucidation of the mechanisms by which single-stranded DNA binding proteins recognize and bind to ssDNA.

THE GENE V PROTEINS

Physico-chemical characteristics. In solution the gene V proteins predominantly occur as dimers of 19 kDa mass, but at millimolar concentrations also higher aggregates are formed (Oey & Knippers, 1972; Pretorius et al. 1975; Pörschke & Rauh, 1983; Bulsink et al., 1986). The binding of M13 and IKe GVP to ssDNA has been monitored by a wide variety of spectroscopic techniques. Upon binding to ssDNA, significant effects on the spectral properties of the GVP tyrosyl chromophores are observed by optical techniques, such as UV

(Day, 1973; van Amerongen et al. 1990), fluorescence (Alma et al., 1983b; Pretorius et al., 1975) and circular dichroism (Day, 1973; Kansky et al. 1986). Also considerable shifts of one pair of tyrosyl resonance positions in the NMR spectrum can be observed (Alma et al., 1981a,b; de Jong et al., 1987b; King & Coleman, 1987). From these data it was inferred that one tyrosyl residue is involved in a hydrophobic (stacking) interaction with the ssDNA.

From a quantitative interpretation of the data derived from these and other experiments (reviewed by Kansy et al., 1986), two types of ssDNA binding have emerged which will be referred to as the oligonucleotide and the polynucleotide binding mode. The main differences between the two are the stoichiometry, cooperativity and salt dependency of the binding (Alma et al., 1982; Kansy et al., 1986; Bulsink et al., 1988b). In the polynucleotide binding mode, four nucleotides bind per protein monomer (n=4) with a high cooperativity, i.e. binding of GVP adjacent to an occupied site is stronger than binding to an isolated site (Alma et al., 1983b; de Jong et al., 1987a). The ability of GVPs to participate in cooperative protein-protein interactions is most probably related to their tendency to form aggregates (de Jong et al., 1987b; Folkers et al., 1991a). The oligonucleotide binding mode is characterized by a different stoichiometry (n=3) and a binding cooperativity that is at least two orders of magnitude smaller than in the polynucleotide binding mode. Both types of binding are sensitive to the salt strength of the solution, which is a strong indication that, besides hydrophobic interactions, also electrostatic interactions play an important role. The polynucleotide binding mode is dominant in binding of GVP to DNA lattices exceeding 20 nucleotides. At very low salt concentration, however, also an n = 3 stoichiometry can be observed in binding experiments to polynucleotide lattices (Kansy et al., 1986).

Several kinetic studies (Pörschke & Rauh, 1983; Shimamoto & Utiyama, 1983; Bulsink et al., 1988a) clearly established the existence of fast rearrangements of the distribution of M13 GVP molecules bound to polynucleotides. These rearrangements are facilitated by sliding (translocation) of GVP molecules along the DNA lattice which may be functional in the rapid formation of nucleoprotein assemblies. Similar translocation mechanisms have been proposed for the sequence-specific dsDNA binding proteins where they may facilitate the search for the appropriate sequence targets (Saenger et al., 1983).

Structural aspects. Figure 1 shows the amino acid sequences of M13 and IKe GVP (87 and 88 residues respectively). The high degree of sequence identity (40%, Peeters et al., 1983) is obvious. The double underlinings refer to residues that are found to be conserved in the amino acid sequences of other ssDNA binding proteins (SSB, RecA and the gene-32 encoded protein of phage T4 (Prasad & Chiu, 1987). This suggests that these proteins employ a common denominator in their interaction with ssDNA.

```
IKe  1 Met Leu Thr Val Glu Ile His Asp Ser Gln Val Ser Val Lys Glu 15
M13  1 Met Ile Lys Val Glu Ile Lys Pro Ser Gln Ala Gln Phe Thr Thr 14

    16 Arg Ser Gly Val Ser Gln Lys Ser Gly Lys Pro Tyr Thr Ile Arg 30
    16 Arg Ser Gly Val Ser --- Arg Gln Gly Lys Pro Tyr Ser Leu Asn 29

    31 Glu Gln Glu Ala Tyr Ile Asp Leu Gly Gly Val Tyr Pro Ala Leu 45
    30 Glu Gln Leu Cys Tyr Val Asp Leu Gly Asn Glu Tyr Pro Val Leu 44

    46 Phe Asn Phe Asn Leu Glu Asp Gly Gln Gln Pro Tyr Pro Ala Gly 60
    45 Val Lys Ile Thr Leu Asp Glu Gly Gln Pro Ala Tyr Ala Pro Gly 59

    60 Lys Tyr Arg Leu His Pro Ala Ser Phe Lys Ile Asn Asn Phe Gly 75
    59 Leu Tyr Thr Val His Leu Ser Ser Phe Lys Val Gly Gln Phe Gly 74

    71 Gln Val Ala Val Gly Arg Val Leu Leu Glu Ser Val Lys 88 IKe
    70 Ser Leu Met Ile Asp Arg Leu Arg Leu Val Pro Ala Lys 87 M13
```

Figure 1. Comparison of the amino acid sequences of M13 (Cuypers et al., 1974) and IKe (Peeters et al., 1983) GVP. Optimum alignment of the amino acid sequences is achieved when Gln 21 in IKe GVP is considered as an insertion. Conserved residues are underlined. Double underlined residues are conserved or conservatively replaced in SSB, RecA, SSF and G32P (see text).

Electron microscopic (Gray et al., 1982b; Gray, 1989) and neutron scattering studies (Gray et al., 1982a) have revealed that, upon complexation with ssDNA, both M13 and IKe GVP form rod-like nucleoprotein assemblies in which the viral DNA adopts a left-handed helical structure. X-ray diffraction studies on M13 GVP have resulted in a dimeric GVP structure with 2.3 Å resolution (Brayer & McPherson, 1983). In this structure a hydrophobic β-barrel can be recognized as well as a solvent exposed triple-stranded β-sheet (Figure 2). In the monomer unit three β-loop structures can be discerned which have been termed the DNA binding loop, the complex loop and the dyad loop.

Though the GVPs of M13 and IKe are small compared to other ssDNA binding pro- teins, their occurrence as 19 kDa dimers is near the limit of what is considered as suitable for
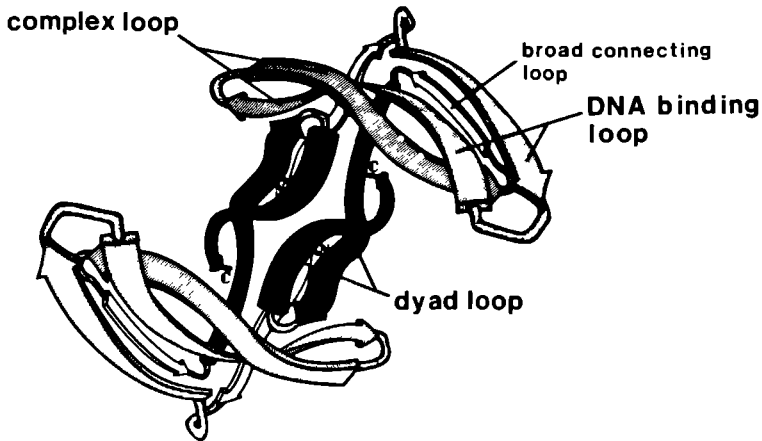
Figure 2. Low resolution representation of the crystal structure of the M13 GVP dimer. The three major β-loops are labeled and shaded differently. The C2 (dyad) axis, relating the two subunits, is perpendicular to the plane of drawing. N and C denote the N- and C-termini respectively. (Reprinted with permission from Brayer & McPherson, 1983).

structural studies by NMR. The efficiency of magnetic relaxation processes in systems with this mass results in a considerable broadening and overlap of resonances in NMR spectra. Moreover, the strong tendency of the GVPs to form aggregates, and the concomitant increase in particle size, makes the problem even worse. For this reason, NMR studies aimed at the structure elucidation of the GVPs have for a long time lagged behind similar studies performed on other proteins.

NMR studies involving oligonucleotides have been successful in revealing the involvement of aromatic and positively charged residues in M13 and IKe GVP in the recognition of ssDNA (Alma et al., 1981a,b; de Jong et al., 1987b; King & Coleman, 1987; 1988). The further exploration of the ssDNA-GVP interaction, however, was seriously hampered by the aforementioned aggregation effect. This has been overcome by labelling oligonucleotides with paramagnetic groups (spin labels). Even upon addition of small aliquots of these agents to the GVPs, the resonances belonging to the amino acid residues in close proximity to the bound oligonucleotide are broadened significantly. These effects can be visualized by difference spectroscopy. Particularly the difference representation of 2D-spectra (resulting from the subtraction of 2D-spectra recorded in the presence and absence of the

paramagnetic ligand) are very useful for spectral simplification and mapping of ssDNA binding domains (de Jong et al., 1988). A meaningful interpretation of the linebroadening effects, however, requires a nearly complete assignment of the NMR spectra and structure elucidation. At the start of the studies presented in this thesis only a single β-loop structure in IKe GVP had been identified (de Jong et al., 1989a). Its structural elucidation was of large significance, however, because the conformation of this β-loop structure differed from that proposed for the homologous segment in M13 GVP on the basis of X-ray diffraction studies. The spectral perturbation effects induced by spin-labeled oligonucleotides indicated that in IKe GVP this structure is intimately involved in the interaction with ssDNA.
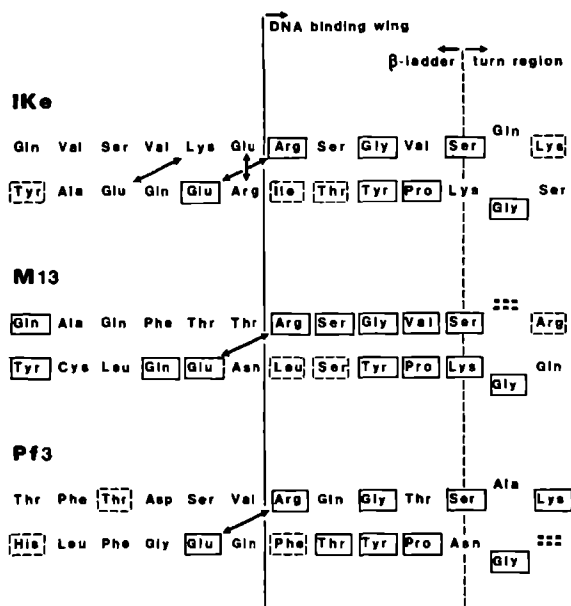
Figure 3. Schematic representation of the β-loop structure of the amino acid segments running from residue 14 to 31 in IKe GVP (a), from residue 14 to 30 in M13 GVP (b) and from residue 10 to 26 in Pf3 DBP (c). The structures for M13 GVP and Pf3 DBP have been drawn in analogy with the structure for IKe GVP. The continuous boxes mark conserved residues while the broken boxes indicate conservative replacements. The dashes in the turn region of Pf3 DBP indicate deleted residues. The arrows refer to putative salt-bridges.

Striking patterns of amino acid residue conservation (Figure 3) in the GVPs and in the ssDNA-binding protein encoded by phage Pf3 (vide supra) have led to the proposal that this

loop structure is a structural motif exploited by ssDNA binding proteins in their binding to ssDNA (de Jong et al., 1989b). Comparison of the ssDNA binding domains of IKe and M13 GVP in terms of common structural elements is one of the main ideas that are explored in this thesis.

OUTLINE OF THESIS

In Chapter 2 the theoretical as well as practical aspects of the so-called transferred NOE effect and the transfer of coherence in a TOCSY experiment are discussed. Particularly those features of NMR theory are treated which are relevant for a quantitative interpretation of these effects.

In pursuit of the experimental verification, by NMR, of the aforementioned β-loop motif in M13 GVP, severe experimental difficulties were encountered. Chapter 3 describes how these have been overcome by the discovery of a M13 gene V mutant protein with properties that were far more favourable for NMR measurements. In Chapter 4 it is demonstrated that the β-loop motif is indeed present in M13 GVP. Chapter 5 describes the complete sequential assignment of the ^{1}H NMR spectrum and the subsequent secondary structure elucidation of IKe GVP. A considerable degree of similarity with the secondary structure of M13 GVP was found. This study reinforced aspects of the M13 GVP secondary structure in solution that were different in the crystal structure. Subsequently, the ssDNA binding domains of M13 and IKe GVP were (further) explored with the aid of a spin-labeled oligonucleotide and paramagnetic lanthanide DOTP complexes (Chapter 6). In Chapter 7 the conformation and dynamics of oligonucleotides in their complexes with M13 GVP are described.

CHAPTER 2

1. Quantitative Evaluation of Two-Dimensional NMR Experiments: TrNOESY

INTRODUCTION

The transferred NOE effect has become an important tool for the determination of the structures of ligands bound to macromolecules. So far, the effect has successfully been applied to a diversity of biologically relevant interactions, e.g. between proteins and oligonucleotides, enzymes and substrates, antibodies and antigens, hormones and receptors, thrombin and peptides, etc. (the most recent contributions in this field have been listed in Campbell & Sykes, 1991). In a typical transferred NOE experiment the effect of cross-relaxation between nuclei of the bound ligand is transferred to the resonances of the free ligand by chemical exchange. The transferred NOE effect was at first observed by Balaram et al. (1972), and treated theoretically by Clore & Gronenborn (1982), for the steady state NOE experiment (ssNOE). Subsequently, the transferred NOE effect was exploited in truncated (driven) NOE experiments (TOE) (Clore & Gronenbron, 1983). In both experiments a single spin is irradiated and the resulting effects on the rest of the 1D-spectrum are measured. Every 1D-experiment, however, provides only information on a limited part of the relaxation network of a macromolecular structure. Furthermore, selective irradiation of spins in such systems is often impeded by resonance overlap.

In recent years a growing number of applications of the transferred NOE effect in NOESY experiments (TrNOESY) has appeared (for a review see Campbell & Sykes, 1991). Many of the limitations of the 1D-NOE experiments (vide supra) are overcome by this two-dimensional technique. The present treatment aims at a theoretical description of the effect of chemical exchange on NOE intensities, occurring in TrNOESY experiments, in terms of a simple matrix formalism. First, the effect of chemical exchange on transverse and longitudinal motion of the magnetization is described. Subsequently, the behaviour of magnetization in the NOESY experiment is discussed.

TWO-SITE CHEMICAL EXCHANGE

We consider the complexation of a low molecular weight ligand, L, with a high molecular weight molecule (a protein), P:

(1)
$$L + P \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} LP$$

By accounting for the free protein concentration, [P], in the association rate, $k_{FB} = k_1[P]$, we may represent the chemical exchange between the free (F) and bound (B) forms of the spin system of L (with n spins and the corresponding magnetizations m_1, m_2.....m_n) in a vector representation:

(2)
$$M_F = \begin{pmatrix} m_{1F} \\ m_{2F} \\ m_{3F} \\ . \\ m_{nF} \end{pmatrix} \underset{k_{BF}}{\overset{k_{FB}}{\rightleftharpoons}} \begin{pmatrix} m_{1B} \\ m_{2B} \\ m_{3B} \\ . \\ m_{nB} \end{pmatrix} = M_B$$

These vectors do not include the magnetizations of the protein P. Hence, we assume that in the complex LP, the spin systems of protein P and ligand L can be considered as separate spinpools and we will only discuss the intramolecular transferred NOE effect.

The fractions of the bound and free ligand forms (p_B and p_F respectively) are given by:

(3)
$$p_F = \frac{k_{BF}}{k_{FB} + k_{BF}}; \quad p_B = \frac{k_{FB}}{k_{FB} + k_{BF}}$$

In the absence of exchange the motions of the (complex) transverse magnetization vector M_F^+ and the (real) longitudinal magnetization vector M_F^z of the free form, F, are described by the spectral matrix Ω_F and longitudinal relaxation matrix L_F respectively (Ernst et al., 1987):

$$(4) \qquad \frac{d}{dt}M_F'(t) = \Omega_F M_F'(t); \quad \Omega_F = \begin{pmatrix} i\omega_{1F}-R_{1F} & 0 & & 0 \\ 0 & i\omega_{2F}-R_{2F} & \cdot & 0 \\ & & \cdot & \\ \cdot & & \cdot & \\ 0 & 0 & \cdot & i\omega_{nF}-R_{nF} \end{pmatrix}$$

$$(5) \qquad \frac{d}{d\tau}M_F^z(\tau) = L_F(M_F^z-M_F^0)(\tau); \quad L_F = \begin{pmatrix} \rho_{1F} & \sigma_{12} & \cdot & \sigma_{1n} \\ \sigma_{21} & \rho_{2F} & \cdot & \cdot \\ & & \cdot & \\ & & \cdot & \\ \sigma_{n1} & & \cdot & \rho_{nF} \end{pmatrix}$$

Here $\omega_{i,F}$ is the chemical shift and $R_{i,F}$ the transverse relaxation rate of spin i; $\rho_{i,F}$ is the spin-lattice relaxation rate of spin i and $\sigma_{i,j}$ the cross-relaxation rate between spin i and j. In an analogous manner a spectral matrix Ω_B and longitudinal relaxation matrix L_B can be defined for spin system B in the bound state.

If we consider the transverse and longitudinal magnetization vectors of spin system B and F as subvectors of the total magnetization vectors, i.e.:

$$(6) \qquad M' = \begin{pmatrix} M_F^+ \\ M_B^+ \end{pmatrix} \qquad M^z = \begin{pmatrix} M_F^z \\ M_B^z \end{pmatrix}$$

then the equations of motion for transverse and longitudinal magnetizations can be written as:

$$(7) \qquad \frac{d}{dt}\begin{pmatrix} M_F' \\ M_B^+ \end{pmatrix}(t) = \Omega_{tot}\begin{pmatrix} M_F^+ \\ M_B^+ \end{pmatrix}(t); \qquad \Omega_{tot} = \begin{pmatrix} \Omega_F-K_{FB} & K_{BF} \\ K_{FB} & \Omega_B K_{BF} \end{pmatrix}$$

In this representation all matrix elements are submatrices; K_{BF} and K_{FB} are the corresponding exchange rates multiplied with unity matrices. M_F^0 and M_B^0 are the longitudinal equilibrium magnetizations of the bound and free ligand forms respectively.

(8)
$$\frac{d}{d\tau}\begin{pmatrix} M_F^z \\ M_B^z \end{pmatrix}(\tau) = L_{tot}\begin{pmatrix} M_F^z - M_F^0 \\ M_B^z - M_B^0 \end{pmatrix}(\tau); \qquad L_{tot} = \begin{pmatrix} L_F - K_{FB} & K_{BF} \\ K_{FB} & L_B - K_{BF} \end{pmatrix}$$

Via the diagonalization of Ω_{tot}, the equation of motion for the transverse magnetization can be integrated which yields the expression for a simple 1D-experiment:

(9)
$$M'(t) = Xe^{(X^{-1}\Omega_{tot}X)t}X^{-1}M'(0)$$

Here X is the eigenbase in which Ω_{tot} takes the diagonal form. The absorbtion mode 1D-spectrum of this system, $S(\omega)$, is given by the Fourier transform ($\mathscr{F}$) of the observable transverse magnetizations $m^+(t)$:

(10)
$$S(\omega) = Re\mathscr{F} \Sigma m_i^+(t)$$

The resonance positions and linewidths in the spectrum are given by the real and imaginary parts of the diagonal elements of the matrix product $X^{-1}\Omega_{tot}X$; intensities and phases can be obtained by performing the matrix calculus in Equation 9.

The NOE effects as they appear in NOESY can be described with the aid of the matrix notation (Macura & Ernst, 1980; Ernst et al., 1987):

(11)
$$M'(t_1,\tau,t_2) = Xe^{(X^{-1}\Omega_{tot}X)t_1}X^{-1}Ye^{(Y^{-1}L_{tot}Y)\tau}Y^{-1}Xe^{(X^{-1}\Omega_{tot}X)t_2}X^{-1}M^0$$

Here Y is the eigenbase in which L_{tot} takes the diagonal form and M^0 is the equilibrium magnetization. The frequency domain NOESY spectrum of this system, $S(\omega_2,\omega_1)$, is given by the (double) Fourier transform ($\mathscr{F}\mathscr{F}$) of the observable transverse magnetizations $m^+(t_1,\tau,t_2)$:

(12)
$$S(\omega_1,\omega_2) = Re \ \mathscr{F} \ \mathscr{F} \ \Sigma m_i^+(t_1,\tau,t_2)$$

The positions and widths of the NOE crosspeaks in the ω_1 and ω_2 dimension correspond with the diagonal elements of the matrix products $(X^{-1}\Omega_{tot}X)t_1$ and $(X^{-1}\Omega_{tot}X)t_2$, respectively. The

corresponding intensities can be obtained by performing the matrix calculus given in Equation 11. In the next section, it will be shown that several exchange regimes exist for which this matrix equation simplifies to conceptually simple expressions.

Transverse magnetization. As was mentioned previously, numerical calculation of linewidths and resonances positions requires the diagonalization of the spectral matrix Ω_{tot}. We may apply a perturbation-like approach by separating the chemical shift and exchange contribution, Ω and K respectively:

$$(13) \qquad \Omega_{tot} = \Omega + K; \quad \Omega = \begin{pmatrix} \Omega_F & 0 \\ 0 & \Omega_B \end{pmatrix} \quad K = \begin{pmatrix} -K_{FB} & K_{BF} \\ K_{FB} & -K_{BF} \end{pmatrix}$$

In the fast exchange regime defined with respect to the chemical shift scale, the matrix K can be considered as a perturbation of Ω (and vice versa in the case of slow exchange). By simple determinant calculus it can be shown that the matrix

$$(14) \qquad X^{-1}\Omega_{tot}X = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

takes the diagonal identity when the following condition pertains:

$$(15) \qquad BC << A-D$$

<u>Fast exchange</u> The exchange contribution, K, to the total spectral and longitudinal relaxation matrices (Ω_{tot} and L_{tot} respectively) can be diagonalized by the transformation Z (I is the unity matrix):

$$(16) \qquad Z = Z^{-1} = \frac{1}{\sqrt{k_{FB}(k_{FB}+k_{BF})}} \begin{pmatrix} -K_{FB} & K_{BF} \\ K_{FB} & K_{FB} \end{pmatrix}$$

Thus:

$$(17) \qquad Z^{-1}KZ \; = \; -(k_{FB}+k_{BF})\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$$

We may apply a perturbation-like approach by evaluating the spectral matrices Ω_B and Ω_F in the eigenbase Z of the kinetic matrix K:

$$(18) \quad Z^{-1}\Omega_{tot}Z \; = \; -(k_{FB}+k_{BF})\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} + (k_{FB}+k_{BF})^{-1}\begin{pmatrix} k_{BF}\Omega_B-k_{FB}\Omega_F & k_{BF}(\Omega_B-\Omega_F) \\ k_{FB}(\Omega_B-\Omega_F) & k_{FB}\Omega_B+k_{BF}\Omega_F \end{pmatrix}$$

By application of Condition 15, we arrive at the definition for the fast exchange regime of the chemical shift scale:

$$(19) \qquad k_{FB}-k_{BF} \; > \; > \; |\Omega_F-\Omega_B|_i$$

(only the diagonal elements i of the difference matrix $\Omega_B-\Omega_F$ have to be considered). The time-dependent term of the equation of motion now simplifies to:

$$(20) \quad e^{X^{-1}\Omega_{tot}Xt} \; \sim \; e^{Z^{-1}\Omega_{tot}Zt} \; \sim \; \begin{pmatrix} e^{(-K_{FB}-K_{BF}+p_F\Omega_B-p_B\Omega_F)t} & 0 \\ 0 & e^{(p_F\Omega_F-p_B\Omega_B)t} \end{pmatrix}$$

When we further take into consideration that in chemical equilibrium:

$$(21) \qquad p_F M_B^+(0) \; = \; p_B M_F^+(0); \quad XM(0) \; \propto \; \begin{pmatrix} 0 \\ M_B^+(0)+M_F^+(0) \end{pmatrix}$$

we arrive at a simple expression for the 1D-experiment:

$$(22) \qquad (M_F^+-M_B^+)(t) \; \sim \; e^{\bar{\Omega}t}(M_F^++M_B^+)(0) \qquad \bar{\Omega} \; = \; p_F\Omega_F + p_B\Omega_B$$

Thus, the equation of motion of the transverse magnetization is ruled by the weighted average of the spectral matrices of the bound and free form. In the NMR spectrum (vide supra) we will therefore observe mean Larmor precession frequencies and mean linewidths (the imaginary and real elements respectively of the mean spectral matrix). For a two-spin system the exchange behaviour of transverse magnetization has also been described analytically (Ernst et al., 1987). From these expressions an analogous expression of fast exchange behaviour can be derived.

Slow exchange The perturbation-like approach, employed to obtain a simple matrix expression for the equation of motion for the fast exchange limit on the chemical shift scale, can also be applied to the slow exchange limit. Now we consider the kinetic matrix K as the perturbation and we evaluate Ω_{tot} in the eigenbase of Ω_{tot}-K, which is the unity matrix I. By simple matrix calculus it can be shown that, if the condition

(23) $$k_{FB}k_{BF} << |\Omega_F - \Omega_B|_i$$

pertains, we can make the simplification:

(24) $$e^{(X^{-1}\Omega_{tot}X)t} \sim \begin{pmatrix} e^{(\Omega_F - K_{FB})t} & 0 \\ 0 & e^{(\Omega_B - K_{BF})t} \end{pmatrix}$$

Thus, we arrive at the trivial expression:

(25) $$M_F'(t) = e^{(\Omega_F - K_{FB})t} M_F'(0)$$

For spin system B we arrive at an analogous equation. Thus, in the slow exchange regime, only the lineshapes (corresponding to the real parts of the matrices Ω-K) are noticeably affected by the transport of magnetization from one site to another. The fast and slow exchange expressions match the extreme limits of analytical solutions derived for a two spin-system in two-site chemical exchange (Ernst et al., 1987). In the next section it will be shown that the matrix formalism outlined above can also be applied to the longitudinal relaxation in the presence of chemical exchange.

Longitudinal Relaxation.

Fast exchange The longitudinal relaxation matrix in the presence of exchange may be treated analogously to the transverse magnetization. The definition of the fast exchange regime, however, now pertains to the elements of the longitudinal matrices. Therefore, application of our perturbation-like approach (vide supra) leads to the simplification:

(26)
$$e^{Y^{-1}LYt} \sim \begin{pmatrix} e^{(-K_{FB}-K_{BF}+p_F L_B+p_B L_F)t} & 0 \\ 0 & e^{\overline{L}t} \end{pmatrix}$$

when:

(27)
$$k_{FB}+k_{BF} >> |L_F-L_B|_{ij}$$

Comparison with the definition of fast exchange on the chemical shift scale shows that we now have to take into account both diagonal and off-diagonal elements of the difference matrix L_F–L_B. This relation shall from now on be referred to as the fast exchange limit on the longitudinal relaxation scale. Recently, analytical expressions were obtained for the spin-lattice and cross-relaxation rates in a two-spin system which was subjected to two-site chemical exchange (Lippens et al., 1992). From these expressions a similar definition of the fast exchange regime on the longitudinal relaxation scale can be derived.

Slow exchange On the longitudinal relaxation scale we may also define a slow exchange regime:

(28)
$$k_{BF} \, k_{FB} << (L_F-L_B)_{ij}$$

where the time-dependence of the longitudinal magnetization simplifies to:

(29)
$$e^{Y^{-1}L_{tot}Yt} \sim \begin{pmatrix} e^{(L_F-K_{FB})t} & 0 \\ 0 & e^{(L_B-K_{BF})t} \end{pmatrix}$$

For complexes of macromolecules of the size of M13 and IKe GVP (20 kD) the magnitude of the matrix elements in L_B and L_F is of the order of 1-10 s^{-1}. For most complexes of such

molecules with small ligands, however, we encounter association and dissociation rates that are at least an order of magnitude larger. We can therefore assume that for such reactions the exchange is fast on the longitudinal relaxation scale. Unlike the transverse magnetization which is directly observable in NMR experiments the cross-relaxation of the longitudinal magnetization can only be visualized in NOE-experiments.

THE TrNOESY EXPERIMENT

In the NOESY experiment, two time-scales are of importance, i.e. the chemical shift scale and the longitudinal relaxation scale. When we consider the matrix representation for the NOESY experiment (Equation 11) in the exchange regimes discussed in the previous section, we arrive at conceptually simple matrix representations.

Fast exchange on both time-scales Under the condition:

(30)
$$k_{BF} - k_{FB} >> |\Omega_F - \Omega_B|_i \ , \quad |L_F - L_B|_{ij}$$

the diagonalizing transformations X and Y simplify to:

(31)
$$X, Y \sim Z = \frac{1}{\sqrt{k_{FB}(k_{FB} + k_{BF})}} \begin{pmatrix} -K_{FB} & K_{BF} \\ K_{FB} & K_{FB} \end{pmatrix}$$

When we take into consideration that:

(32)
$$p_F M_B^0 = p_B M_F^0 \ ; \quad X M^0 \sim \begin{pmatrix} 0 \\ M_B^0 + M_F^0 \end{pmatrix}$$

we arrive at:

(33)
$$(M_F^+ + M_B^+)(t_1, \tau, t_2) = e^{\bar{\Omega} t_2} e^{\bar{L}\tau} e^{\bar{\Omega} t_1} (M_F^0 - M_B^0)$$

Sum magnetization vectors are observed with elements which evolve in the acquisition times t_1 and t_2 with mean Larmor frequencies and linewidths. The transfer of magnetization during the mixing time is governed by a mean longitudinal relaxation matrix.

Slow exchange on both time-scales The appearance of transferred NOE effects requires an exchange rate that is sufficiently fast on the longitudinal relaxation scale. When exchange is slow on both scales:

(34) $$k_{BF} \; k_{FB} << |\Omega_F - \Omega_B|_t \; , \quad |L_F - L_B|_{ij}$$

the diagonalizing transformations X and Y can be replaced by unity matrices (vide supra) and the expressions for the NOE effects of spin system F simplify to:

(35) $$M_F^i(t_1, \tau, t_2) = e^{(\Omega_F - K_{FB})t_2} \; e^{(L_F - K_{FB})\tau} \; e^{(\Omega_F - K_{FB})t_1} \; M_F^0$$

(for spin system B similar expressions can be derived). Obviously, the magnetization vectors of the free and bound spin system evolve independently during the acquisition times (t_1, t_2) and mixing time τ and no transferred NOE effect can be observed.

Exchange that is slow on the chemical shift scale and fast on the longitudinal relaxation scale This is a situation that is often encountered when studying the complexation of biomolecules with strong binding agents by high-resolution ^{1}H NMR. In this case we write:

(36) $$|L_F - L_B|_{ij} << k_{BF} + k_{FB}; \quad k_{BF} \; k_{FB} << \Omega_F - \Omega_B|_t$$

(37) $$X \sim I; \quad Y \sim Z = \frac{1}{\sqrt{k_{FB}(k_{FB} + k_{BF})}} \begin{pmatrix} -K_{FB} & K_{BF} \\ K_{FB} & K_{FB} \end{pmatrix}$$

This results in:

(38) $$e^{X^{-1}\Omega_{tot}Xt} \sim \begin{pmatrix} e^{(\Omega_F - K_{FB})t} & 0 \\ 0 & e^{(\Omega_B - K_{BF})t} \end{pmatrix}$$

(39)
$$e^{Y\,{}^1LYt} \sim \begin{pmatrix} e^{(-K_{FB}-K_{BF}+p_FL_B-p_BL_F)t} & 0 \\ 0 & e^{\bar{L}t} \end{pmatrix}$$

When the condition

(40)
$$k_{BF}+k_{FB} > > |L_{F,ij}| \ , \ \ |L_{B,ij}|$$

applies, the upper term in Expression 39 will decay rapidly and by elaboration of Expression 11 we arrive at:

(41)
$$\begin{pmatrix} M'_F \\ M'_B \end{pmatrix}(t_1,\tau,t_2) = \begin{pmatrix} e^{(\Omega_F-K_{FB})t_2} & 0 \\ 0 & e^{(\Omega_B-K_{BF})t_2} \end{pmatrix}$$
$$\begin{pmatrix} p_F e^{\bar{L}} & p_F e^{\bar{L}} \\ p_B e^{\bar{L}} & p_B e^{\bar{L}} \end{pmatrix}$$
$$\begin{pmatrix} e^{(\Omega_F-K_{FB})t_1} & 0 \\ 0 & e^{(\Omega_B-K_{BF})t_1} \end{pmatrix} \begin{pmatrix} M^0_F \\ M^0_B \end{pmatrix}$$

From this expression it can be inferred that, though the bound and free forms can be observed separately in the two-dimensional spectrum, the sets of intramolecular connectivities between the resonances of the free and the bound spin system are both propagated by the same average longitudinal relaxation process. In addition, the set of crosspeaks *between* bound and free ligand forms (the off-diagonal matrix elements in Equation 41) is also described by the same average longitudinal relaxation matrix. It is noted that the requirement (Condition 40) for the simplification to an average relaxation matrix is more stringent than for the regime where exchange is fast on both time-scales.

The exchange regimes discussed above cover most of the physically relevant situations. It has been mentioned that the time-scales of transverse motion and longitudinal relaxation differ by at least an order of magnitude. Therefore, the case in which exchange is

fast on the chemical shift scale but slow on the longitudinal relaxation scale bears little physical relevance and is not discussed here.

TrNOESY for a two-spin system in a two-site exchange model. To gain further insight in the manner in which the intramolecular transferred NOEs depend on the main experimental parameters of the TrNOESY experiment, in particular on the fraction p_B of the bound form, B, and the mixing time τ, we now turn to a two-spin system where straightforward analytical solutions for the crosspeak intensities can be obtained (Macura & Ernst, 1980). In a typical transferred NOE experiment the cross-relaxation in the bound form B (mostly in the slow tumbling limit) is far more efficient than in the free form, F. It is therefore allowed to neglect cross-relaxation in the free form. Thus, we deal with the following matrices:

$$(42) \qquad L_B = \begin{pmatrix} \rho_B + \rho_L & \sigma_B \\ \sigma_B & \rho_B + \rho_L \end{pmatrix} \qquad L_F \approx \begin{pmatrix} \rho_F & 0 \\ 0 & \rho_F \end{pmatrix}$$

The diagonal terms ρ_L in L_B account for external leakage pathways in the bound form.

The cross and diagonal peaks of the TrNOESY spectrum in the fast and slow exchange limits of the longitudinal relaxation scale are described by the matrices NF and NS respectively:

$$(43) \qquad NF = e^{(p_F L_F + p_B L_B)\tau}$$

$$(44) \qquad NS = p_B e^{L_B \tau}$$

The analytical solutions for the crosspeaks $NF_{12}(\tau)$ and $NS_{12}(\tau_m)$ in this two-spin system are given by:

$$(45) \qquad NF_{12}(\tau) = \frac{1}{2}[1 - e^{-2p_B \sigma_B \tau}]e^{[p_F \rho_F + p_B(\rho_B - \sigma_B + \rho_L)]\tau}$$

(46)
$$NS_{12}(\tau) = \frac{1}{2} p_B [1-e^{-2\sigma_B\tau}] e^{(\rho_B+\sigma_B+\rho_L)\tau}$$

NOE build-up curves calculated for the fast and the slow exchange regime on the longitudinal relaxation scale are presented in Figure 1A(I) and 1B(I) respectively. The relaxation properties of the free and bound form that were used in these simulations are summarized in Table I.

Table I

Summary of relaxation rates of the bound and free forms (see text) that were used in the simulations represented in Figure 1.

$[s^{-1}]$	Free	Bound
σ	0	-4.5
ρ	+0.5	+4.5
ρ_L	--	+0.5

Curves 1A(II) and 1B(II) are the build-up curves represented with scaled mixing time axes ($p_B\tau$). This representation shows that the effect of p_B on the appearance of the fast exchange limit NOESY spectrum can be separated in two effects. In first order, the fraction p_B scales down the mixing time τ: in the scaled fast exchange curves (Figure 1A(II)) all initial build-up rates are approximately equal. In second order, we observe that, at decreasing p_B, the scaled mixing time at which the transferred magnetization starts to decay shifts to smaller values. This effect can be ascribed to the spin-lattice relaxation properties of the free form which become dominant at small p_B values (this result can also easily be deduced from Equation 45). In contrast, in the slow exchange limit on the longitudinal relaxation scale, the build-up curves have congruent shapes and the amount of transferred magnetization at any mixing time (unscaled) is proportional with p_B. A main advantage of measuring NOEs in the fast exchange limit is that relatively large amounts of magnetizations can be transferred from one spin to another without the requirement, like in the slow exchange limit, of large fractions of bound material present. From these observations we infer that, in the fast exchange limit on the longitudinal relaxation scale, TrNOESY experiments can best be performed at small
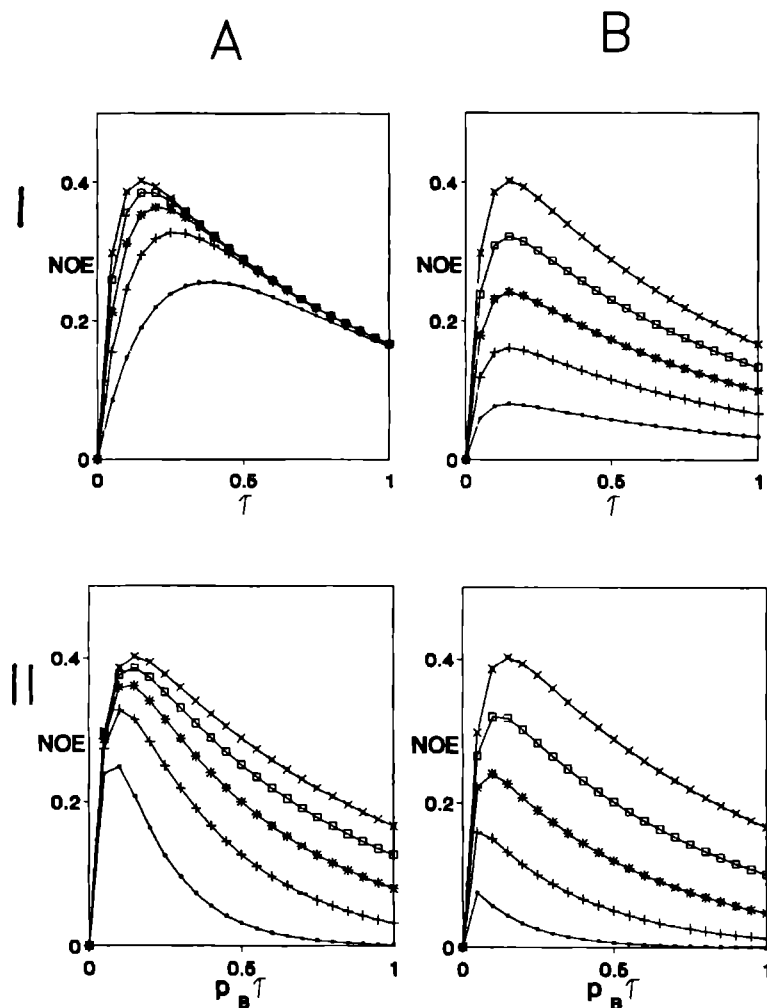
Figure 1. *NOESY build-up curves of a two-spin system (A), in the fast, and (B), in the slow exchange regime of the longitudinal relaxation scale (two sites). The curves were calculated for $p_B = 0.20$ (...); 0.40 (+++); 0.60 (***); 0.80 (□□□); 1.00 (xxx). Other spin system characteristics are described in Table I. The curves are represented (I) with normal mixing time axes (τ_m) and (II) mixing time axes scaled with p_B ($p_B\tau$).*

fractions of the bound (slowly tumbling) material present. At small p_B values, in the initial rate regime, less signal will be observed, but mixing times can be prolonged to longer values before magnetization will start to decay. When only small fractions of bound material are

present also the linebroadening effects due to exchange and particle size increase are still minimal which is an important prerequisite for two-dimensional experiments.

In Chapter 7 the theoretical considerations elaborated in this Chapter will find a practical application. In particular the simple matrix formalism described in the previous sections will prove useful for the quantitative interpretation of TrNOESY spectra of oligonucleotides bound to M13 GVP.

CHAPTER 2

2. Quantitative Evaluation of Two-Dimensional NMR Experiments: TOCSY

Application to Deoxyribofuranose Ring Conformational Analysis[*]

INTRODUCTION

TOCSY experiments are routinely employed in NMR studies of proteins and nucleic acids. The common denominator of all TOCSY (or HOHAHA) experiments is an average mixing Hamiltonian from which all chemical shift differences are removed as much as possible without affecting the mutual spin-spin couplings. In the ideal situation, often termed isotropic mixing, efficient net transfer of coherence is achieved, within a scalar-coupled network of spins. The connectivity patterns generated in two- (Bax & Davis, 1985) and three-dimensional (van Mierlo et al., 1990; Vuister et al., 1990; Mooren et al., 1991) TOCSY experiments are extremely useful in the spectral assignment of protein and nucleic acid ^{1}H NMR spectra. TOCSY experiments have two main advantages over COSY and relayed COSY (RCT) experiments (Ernst et al., 1987), which also exploit the scalar coupling. First, the relay of coherence achieved during the TOCSY mixing times is far more efficient than by means of the relayed COSY (RCT) method. Secondly, in TOCSY experiments net coherence is transferred, which is often desirable because it results in in-phase multiplet patterns. The phase-sensitive COSY experiment produces anti-phase multiplets which tend to self-cancelation at increasing linewidths. For this reason, the latter experiment is less suitable for high molecular weight systems with their efficient relaxation pathways and concomitant large intrinsic linewidths. A main advantage of COSY is, however, the straightforward extraction of accurate J-coupling data. It is well recognized that J-coupling constants can be related to torsion angles, which are important parameters for structure refinement. A well-known example is the conformational analysis of the sugar ring in nucleic acids. In the concept of pseudorotation, the sugar (deoxy)ribose conformation can be described completely by two

---

parameters, the phase angle of the pseudorotation, P, and the pucker amplitude, Φ_m (Altona & Sundaralingham, 1972). Commonly, the ring conformation is analyzed in terms of a fast equilibrium between the two energetically most favourable conformers, which are denoted as N-type (C3'-endo, $-18^{\circ}<P_N<18^{\circ}$, $\Phi_m=35\pm5^{\circ}$) and S-type (C2'-endo, $140<P_S<180^{\circ}$, $\Phi_m=35\pm5^{\circ}$) puckers.

In this Chapter the cyclic dinucleotide cd(CpGp) is employed as a model system to evaluate the feasibility of extracting quantitative structural information (e.g the percentage of N and S conformers) from TOCSY data. The relatively small size of this molecule allowed the determination of both the chemical shifts and the J-coupling constants from 1D-spectra (Mooren et al., 1992). By analysis of the J-coupling data the mean sugar ring conformation could be established. We will show that this mean sugar ring conformation can also be obtained by matching experimental crosspeak intensities from a TOCSY (MLEV17) experiment with simulated coherence transfer curves.

MATERIALS AND METHODS

Experimental. The cyclic dinucleotide cd(CpGp) was synthesized by an improved phosphotriester method (Vroom, 1988). TOCSY spectra were recorded at 400 MHz and 298 K on a Bruker AM400 spectrometer with mixing times of 20, 37, 50, 66, 81, 96 and 111 ms, using a MLEV17 mixing pulse (Bax & Davis, 1985). The carrier frequency was placed at the HDO resonance and the strength of the radiofrequency field was 10.9 kHz. TPPI was used for signal accumulation in the t_1 dimension (Marion & Wüthrich, 1983). Continuous irradiation was applied to suppress the HDO resonance; the irradiation frequency was phase coherent with the carrier frequency (Zuiderweg et al., 1986). All spectra were acquired with 512 points in the t_1 and 1024 points in the t_2 dimension. The spectral widths in f_1 and f_2 were both set to 5000 Hz. After zero-filling in the t_1 dimension and weighting with shifted squared cosine functions the data were Fourier transformed in the phase-sensitive mode. Processing of the NMR data, including determination of the peak volumes in TOCSY spectra, was done with standard Bruker software installed on a Bruker X32 workstation.

Calculation of coherence transfer during the MLEV17 mixing sequence. The MLEV17 pulse sequence brings about nearly isotropic mixing conditions, for magnetization aligned along the x-axis, by permutation of two composite pulses A and B: $ABBA$-$AABB$-$BAAB$-$BBAA$-$180°$. A is a composite $180°$ rotation about the x-axis and B is the inverse of A. The last $180°$ pulse removes a major part of pulse and mismatch artifacts that may build up during the preceding pulse train (Bax & Davis, 1985).

The rather complicated behaviour of coherence transfer in such multiple pulse cycles is best described by numerical density matrix calculations. During any multi-pulse experiment the time-dependent Hamiltonian $\hat{H}(t)$ is, in the rotating frame representation, piecewise constant in p successive time intervals:

(1) $$\hat{H}(t) = \hat{H}_i \qquad t_i < t < t_{i+1} \qquad i = 1, p$$

In the time-independent Hamiltonians $\hat{H}_i$ an invariable contribution from the intrinsic spin system $(\hat{H}_{intr})$ and a variable contribution from the radiofrequency pulses $(\hat{H}_{i,pulse})$ can be distinguished:

(2) $$\hat{H}_i = \hat{H}_{intr} - \hat{H}_{i,pulse}$$

(3) $$\hat{H}_{intr} = \sum_i^n \omega_i \hat{I}_{i,z} + \sum_{i}^{n}\sum_{j>i}^{n} J_{ij}\hat{I}_i\hat{I}_j$$

(4) $$\hat{H}_{i,pulse} = \pm\omega_{rf}\sum_i^n \hat{I}_{ix,iy}$$

In $\hat{H}_{intr}$ we recognize Zeeman contributions $(\omega_i\hat{I}_{i,z})$, also denoted as chemical shift, frequency offset or (with respect to TOCSY experiments) as mismatch terms, and the scalar-coupling contributions $J_{ij}\hat{I}_i\hat{I}_j$.

For periodic multiple pulse sequences, the time-evolution of the density matrix can be described by a single unitary transformation, U_{av}, which corresponds with an average Hamiltonian, $\hat{H}_{av}$ (Ernst et al., 1987):

(5)
$$U_{av}(t_c) = e^{-i\hat{H}_{av}t_c} = e^{-i\hat{H}_p t_p} \ldots\ldots e^{-i\hat{H}_i t_i} \ldots\ldots e^{-i\hat{H}_1 t_1}$$

(6)
$$t_c = \sum_i^p t_i$$

It is the incomplete removal of the Zeeman effects from the average (effective) Hamiltonian that is responsible for departures from the coherence transfer behaviour dictated by ideal isotropic mixing conditions. For the numerical calculation of U_{av}, diagonalization of the matrix representations of the Hamiltonians $\hat{H}_i$ is required:

(7)
$$U_i = e^{-i\hat{H}_i t_i} = T_i e^{-i(T_i^{-1}\hat{H}_i T_i)t_i} T_i^{-1}$$

where T_i is the corresponding diagonalization transformation.

For *stroboscopic* observation in synchronism with the *periodic* multiple pulse sequence (every t_c):

(8)
$$\sigma(nt_c) = U_{av}{}^n \sigma(0) U_{av}{}^{-n}$$

The expectation value $\langle \hat{I}_{k,\alpha} \rangle$ of any component α (α=x,y,z) of the nuclear spin angular momentum of spin k can be obtained by calculating the trace of the product of the matrix representations of $\hat{I}_{k,\alpha}$ and $\sigma(t)$:

(9)
$$\langle I_{k,\alpha} \rangle(t) = Trace(I_{k,\alpha} \cdot \sigma(t))$$

To simulate the time-evolution of the intensities in a so-called "TOCSY ladder", where the diagonal intensity corresponds with spin l, the initial density matrix, $\sigma(0)$, is set to $\hat{I}_{l,x}$ and the coherence transfers to the x-components of the other spins, k, are monitored.

The incorporation of chemical shift (offset) effects in the numerical calculations, precludes the a priori assumption of equivalent spins. Therefore no group (composite) spin simplifications can be made (Cavanagh, 1990). In the numerical simulations of the time behaviour of coherence the calculation of the average Hamiltonian is relatively most time-consuming. An inspection of the MLEV17 sequence shows that only $\pm$x and $\pm$y phases are

applied. Therefore only two diagonalizations of Hamiltonian matrices need to be performed to calculate all unitary transformations U_i (Equation 7) and to construct the average unitary transformation U_{av} (Equation 5) for a given MLEV17 set-up (field strength, scalar-coupling network, chemical shift terms). The subsequent calculation of the time evolution of a nuclear spin angular momentum requires only simple matrix multiplications (Equation 8) which can be performed with considerably much greater speed and efficiency than the diagonalization steps. The simulation program was written in FORTRAN and run on a CONVEX C1 computer.

RESULTS AND DISCUSSION

Back-calculation of intensities in experimental TOCSY spectra. The transfer of coherence, occurring during the MLEV17 mixing pulse, between the H5 and H6 spins (of the cytosine ring in cd(CpGp), is presented in Figure 1. In this representation the loss of magnetization due to relaxation effects is taken out of consideration by scaling the sum of transferred and non-transferred magnetization to unity. Under the assumption of ideal isotropic mixing conditions (i.e. assuming that only the scalar J-coupling $^3J_{H5-H6}$ contributes to the mixing Hamiltonian), the periodicity of the coherence transfer between the H5 and H6 spin can be described analytically by simple cosine and sine functions with a period $1/2^3J_{H5-H6}$ (dashed curves in Figure 1). Clearly, the experimental data depart from ideal isotropic mixing behaviour. A main cause is the incomplete elimination of offset terms in the effective (average) Hamiltonian operative during the MLEV17 multi-pulse sequence (Waugh, 1987). When Zeeman (chemical shift) terms are accounted for in the numerical simulation (the drawn line in Figure 1) we indeed observe a better fit to the experimental data. There is, however, a small systematic deviation between the simulated and experimental coherence transfer curves. Most likely this effect can be attributed to inhomogeneities in the applied radiofrequency fields (Bax, 1984).

The experimental and simulated data obtained for the time evolution of the coherence transfer from the H1' spin of the guanine spin system to the other members of the scalar coupled network are shown in Figure 2A and 2B, respectively. The experimental intensities in these H1' TOCSY ladders were scaled in a manner that preserved the total magnetization
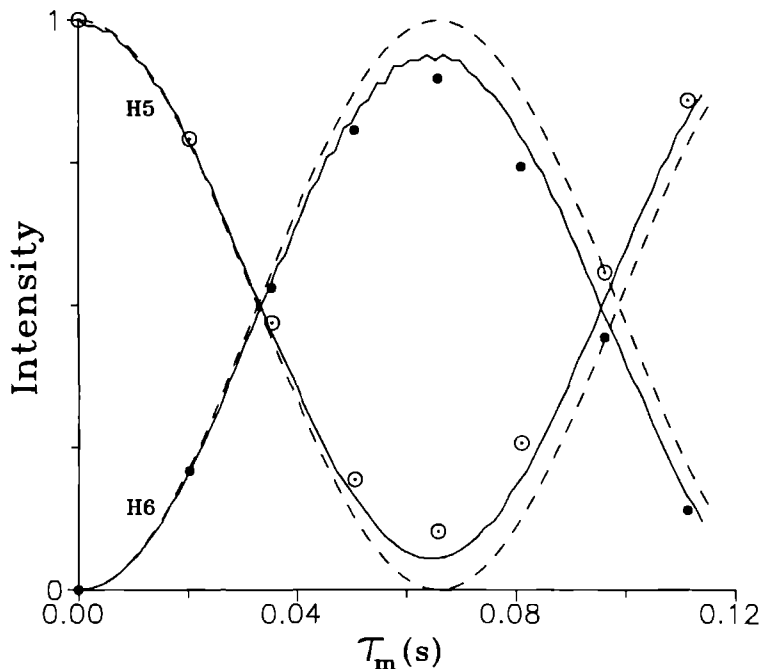
Figure 1. *Experimental points and calculated curves for the coherence transfer between the H5 and H6 spin of the cytosine ring spin system in cd(CpGp)* $(^3J_{H5\text{-}H6}= 7.6\ Hz)$. *The dashed line represents the behaviour of a two-spin system under ideal isotropic mixing conditions i.e. no offset effects and coherence transfer with a periodicity* $1/2^3J_{H5\text{-}H6}$. *The drawn line corresponds with a numerical simulations which also takes the difference in chemical shift of the H5 and H6 resonance into account. Chemical shift data were taken from (Mooren et al., 1992). The symbols correspond to the normalized (see text) experimental H5-H6 TOCSY crosspeaks.*

(to unity). The correlation diagram in Figure 2C illustrates that the incorporation of Zeeman terms in the numerical calculations results in a close match between experimental and simulated TOCSY data.

Manifestation of the N⇌S equilibrium in TOCSY spectra. The neat reproduction of experimental TOCSY data by the numerical simulations urged us to investigate the feasibility of establishing the mean sugar ring conformations in the cyclic dinucleotide model system. We first considered the two energetically most favourable sugar conformers, i.e. the N-puckered and the S-puckered forms. In Figure 3 the curves are presented for coherence

Figure 2.

(A) Experimental build-up curves for the
intensities of the TOCSY connectivities
between the H1'-resonance and the
other resonances (H2', H2", H3, H4',
H5', H5") of the sugar ring of the
guanine nucleotide in cd(CpGp). For
this H1' TOCSY ladder the experimental
TOCSY intensities were scaled in a
manner that preserved the total
magnetization (see text).
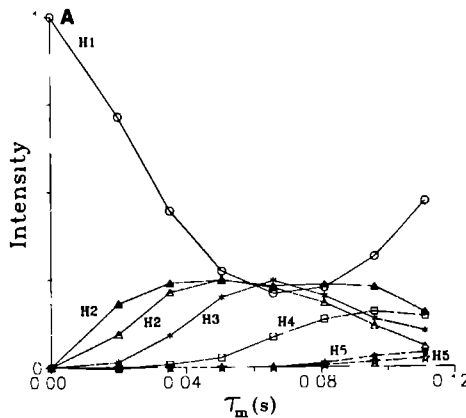
(B) Simulated build-up curves for the
TOCSY intensities mentioned in (A)
using the J-coupling and chemical shift
data taken from Mooren et al., 1992.
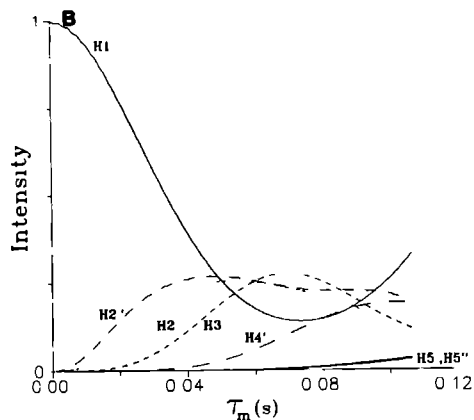
(C) Plot of the experimental TOCSY
intensities (from Figure A) versus the
simulated values (from Figure B) for the
H1' connectivities of the guanine (o) and
cytosine (*) sugar ring spin systems in
cd(CpGp).
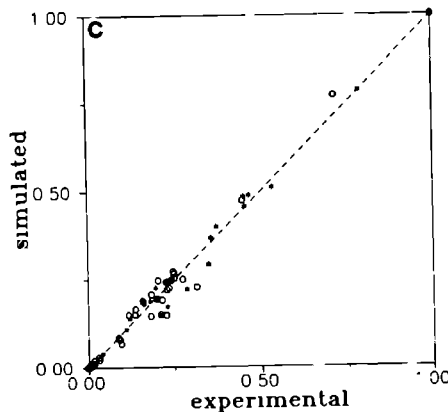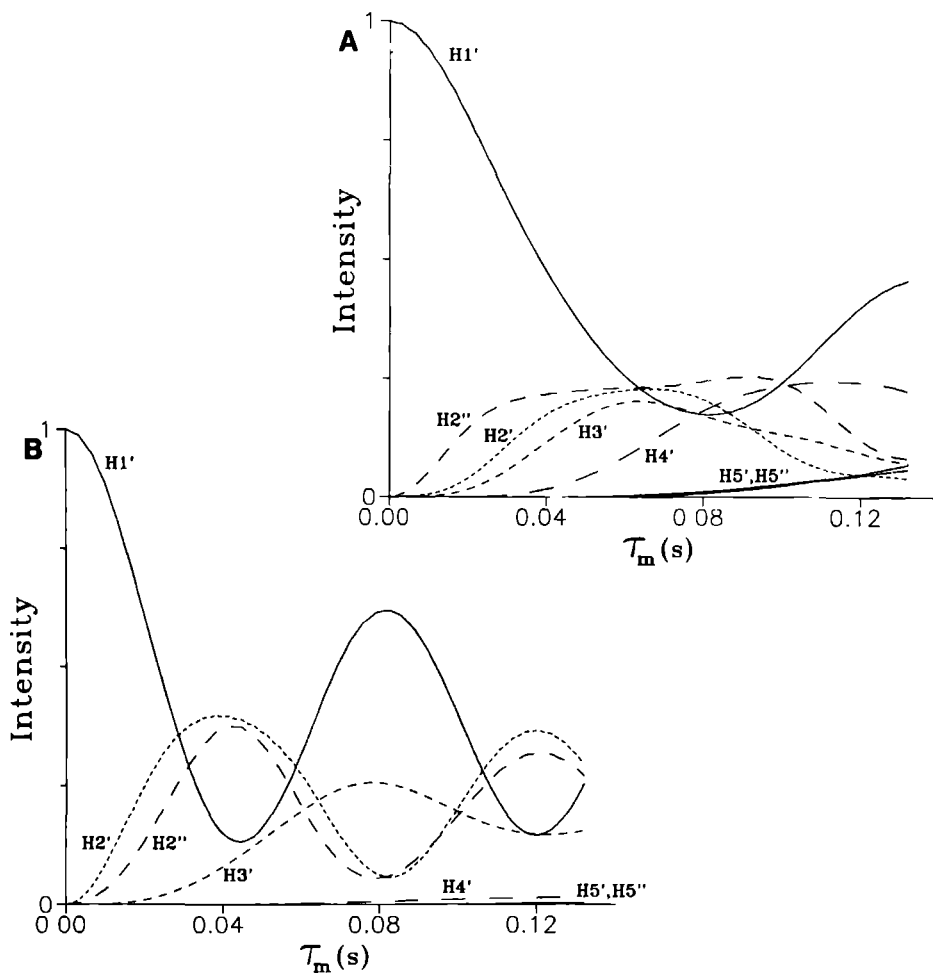
Figure 3. Simulated curves for the coherence transfer from the H1' spin to other spins of a deoxyribose ring spin system in (A) a N-pucker conformation ($P_N=9^0$, $\Phi_m=35^0$) and (B) a S-pucker conformation ($P_S=162^0$, $\Phi_m=35^0$). The offset (chemical shift) terms used in these numerical calculations corresponded with the cytidine spins system in cd(CpGp) (Mooren et al., 1992). The sets of J-couplings corresponding to the N- and S-conformers were calculated according to Rinkel & Altona (1987).

transfer from the H1' spin to other members of the scalar coupled network for conformers with typical N and S pseudorotation parameters ($P_N=9^0$, $\Phi_m=35^0$ and $P_S=162^0$, $\Phi_m=35^0$ respectively). The quite different appearance of the coherence transfer curves in Figure 3A and 3B shows that the different conformations of the sugar rings are manifested by H1'

TOCSY ladders with considerably different intensities. Variation of the P_N and P_S pseudorotation angles within a range of approximately 10^0-20^0 does not significantly change the coherence transfer characteristics (data not shown).

Subsequently, we assume the existence of a fast equilibrium between the N- and S-conformers. This implies that the network of effective J-couplings is the weighted average of the sets of J-couplings of the pure N and S conformers. Figures 4A-C show (for several MLEV17 mixing times) simulations of the coherence transfer as a function of the fraction N conformer. In every diagram, three curves are presented, corresponding with a inter-conversion between an N- ($P_N=9^0$, $\Phi_m=35^0$) and three S-type conformations ($P_S=117^0$; $P_S=162^0$; $P_S=189^0$; $\Phi_m=35^0$) respectively. It is clear that the intensities of the TOCSY connectivities between H1' and the other sugar ring spins vary significantly with the fraction N conformer, X_N. To a lesser extent, the TOCSY intensities also depended on the pseudorotation angle P_S. As expected (vide supra), the effect of relative small variations in P_S, e.g. $162^0 \rightarrow 189^0$, is much smaller than the effect of X_N. Only larger variations, e.g. $P_S=162^0 \rightarrow 117^0$, have a noticeable influence on the TOCSY intensities, but only at prolonged mixing times.

Assessing the mean ring conformation by iterative minimization. The simulations presented in the previous section suggest that the fraction X_N can be established by compa-rison of the simulated intensities $I_i^{sim}(\tau_k, X_N)$ with the experimental values. To investigate this, again fast exchange is assumed between S and N conformers with pseudorotation parameters of ($P_S=9^0$, $\Phi_m=35^0$) and ($P_N=162^0$, $\Phi_m=35^0$) respectively. For every subset, k, of N_k experimental TOCSY intensities, $I_i^{exp}(\tau_k)$, a merit function, $\chi_k^{2r}(X_N)$, (Bevington, 1969) is defined:

$$(10) \qquad \chi_k^{2r}(X_N) = \sum_{i-1}^{N_k} \frac{[I_i^{exp}(\tau_k)-I_i^{sim}(\tau_k, X_N)]^2}{\sigma_i^2}$$

Every subset includes the intensities of a H1' ladder, measured at mixing times τ_k; σ_i is the (estimated) standard deviation of intensity i. Also a merit function χ_{all}^{2r} can be defined, which takes the intensities of a series of 7 experiments (mixing times 20, 37, 50, 66, 81, 96 and 110
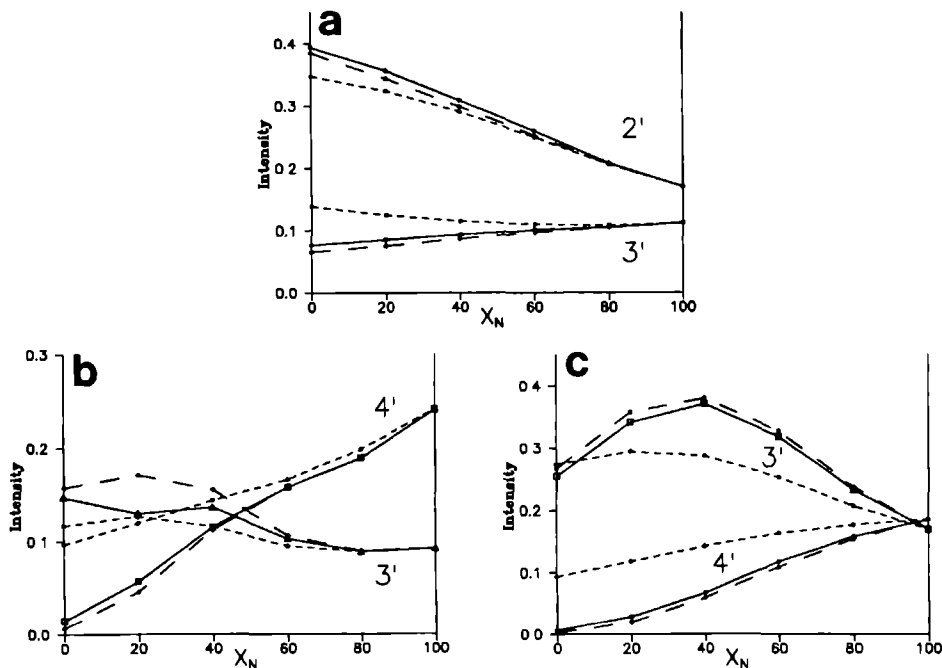
Figure 4. Simulated TOCSY intensities between the H1' and other sugar resonances, as a function of the fraction N-type conformer (at MLEV17 mixing times of (A) 40 ms, (B) 80 ms and (C) 120 ms) for a sugar ring in fast equilibrium between an N-type conformer ($P_N=9^0$, $\Phi_m=35^0$) and S-type conformer ($P_S=117^0$ (---); $P_S=162^0$ (straight); $P_S=189^0$ (———)). The set of effective (mean) scalar J-coupling constants was calculated by weighting the sets of J-couplings for the pure N- and S-conformers. The set of average chemical shifts of the guanosine and cytidine sugar ring spin systems were taken as representative for deoxyribofuranose chemical shifts.

ms) into account:

(11)
$$\chi_{all}^{2r}(X_N) = \sum_{k=1}^{7} \chi_k^{2r}(X_N)$$

The Levenberg-Marquardt algorithm (Press et al., 1989; Bevington, 1969) provides an elegant method for the minimization of χ^{2r} functions, for models which depend nonlinearly on an

unknown parameter, in our case X_N. Results of the minimizations of $\chi^{2r}(X_N)$ as a function of X_N are summarized in Table I.

TABLE I

Comparison of the fraction of N conformer, X_N, for the cytosine and guanosine sugar ring spin system in cd(CpGp), as determined by minimization of the merit functions χ_k^{2r}, and the total merit function χ_{all}^{2r}. N_k is the number of TOCSY intensities included in the calculations. The 95% confidence limits of X_N are given in brackets. For comparison, the result of the J-coupling analysis of the 1D-spectrum via the graphical method (Rinkel & Altona, 1987) is presented.

	Cytidine		Guanosine	
Subset	X_N	N_k	X_N	N_k
50 ms	0.77 (0.10)	4	0.79 (0.06)	3
65 ms	0.93 (0.08)	4	0.83 (0.12)	3
80 ms	0.93 (0.28)	4	0.83 (0.13)	3
All	0.88 (0.08)	28	0.79 (0.04)	21
J-coupling	0.86-0.91	--	0.75-0.80	--

These values agree nicely with those values obtained directly from J-coupling data via the customary method (Mooren et al., 1992). The correlation diagram in Figure 5 shows the nice correspondence between experimental data and the simulations that gave the best fit in the minimization of the total merit function, $\chi_{all}^{2r}(X_N)$.

In the current calculations only intensities of connectivities between the H1' spin and other members of the sugar ring spin system have been included. For other "TOCSY ladders" similar results can be obtained. It is our experience, however, that in larger nucleic acid structures these TOCSY ladders show a larger degree of overlap and will therefore be of less significance with respect to future sugar ring conformational analysis in such systems.

The present results prove the feasibility of extracting valuable quantitative structural information from TOCSY spectra. Instead of going through the intermediate of J-couplings and assessing torsion angles via Karplus relations the relevant structural parameter, X_N, was

Figure 5. Plot of the experimental intensities of the TOCSY connectivities between the H1' spin and other spins of the cytosine (*) and guanine (o) sugar ring spin system in cd(CpGp) versus the calculated values for the simulations that gave the best fits in the optimization procedure (guanine: $X_N=0.75$; cytosine: $X_N=0.86$).
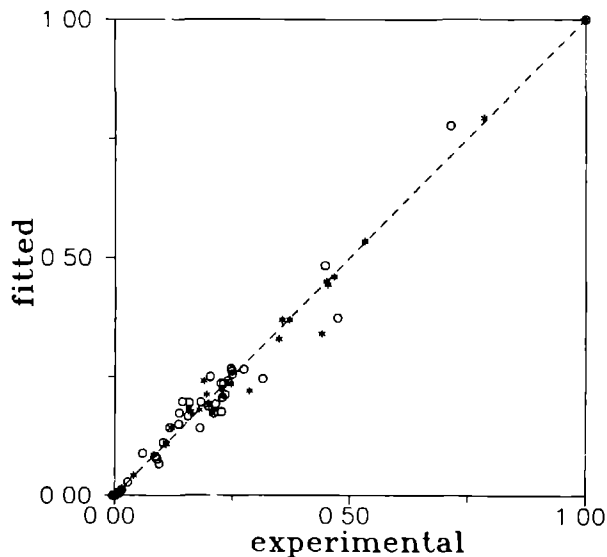
determined directly from the spectral data. This approach bears relevance for structure refinements of large macromolecular systems where accurate J-coupling data can no longer be obtained from COSY experiments but where TOCSY still yields good quality spectra.

CHAPTER 3

Characterization of Wild-type and Mutant M13 Gene V Proteins by Means of ^{1}H NMR[*]

ABSTRACT

Recording of good quality NMR spectra of the single-stranded DNA binding protein gene V of the bacteriophage M13 is hindered by a specific protein aggregation effect. Conditions are described for which NMR spectra of the protein can best be recorded. The aromatic part of the spectrum is reinvestigated by means of two-dimensional total correlation spectroscopy. Sequence-specific assignments are obtained for all of the aromatic amino acid residues with the help of a series of single-site mutant proteins. The solution properties of the mutants of the aromatic amino acid residues have been fully investigated. It is shown that, for these proteins, either none or only local changes occur compared to the wild-type molecule. Spin-labeled oligonucleotide binding studies of wild-type and mutant gene V proteins indicate that Tyrosine 26 and Phenylalanine 73 are the only aromatic residues involved in binding to short stretches of ssDNA. The degree of aggregation of wild-type gene V protein is dependent on both the total protein and salt concentration. The data obtained suggest the occurrence of specific protein-protein interactions between dimeric gene V protein molecules in which the tyrosine residue at position 41 is involved. This hypothesis is further strengthened by the observation that the solubility of tyrosine 41 mutants of gene V protein is significantly higher than that of the wild-type protein. The discovery of the so-called "solubility" mutants of M13 gene V protein finally has made it possible to study the solution structure of gene V protein and its interaction with ssDNA by means of 2D NMR.

INTRODUCTION

One of the essential genes in the life-cycle of filamentous bacteriophages, such as Ff (M13, fd, f1) and IKe, is gene V. The product of this gene is a single-stranded DNA binding protein (GVP) which is indispensable for the production of progeny viral strand DNA (Model & Russel, 1988). Late after infection, the protein binds strongly and cooperatively to ssDNA and initiates in this way the switch of double stranded DNA synthesis to the production of progeny viral strands. Since its first isolation, the protein has been subjected to various physico-chemical studies and is considered an important model for protein-ssDNA

interactions. The protein consists of 87 amino acids, and exists in solution predominantly as a dimer (molecular mass 19.4 kDa). It is the only ssDNA binding protein whose crystal structure is currently known (Brayer & McPherson, 1983). The refined model of the protein reveals that the individual monomers are entirely composed of ß-structure and that they are closely associated about a twofold axis. On the basis of the crystal structure, a model for the complex of ssDNA and GVP has been proposed (Brayer & McPherson, 1984). In this interaction model, two nucleic acid strands bind to one dimer in an antiparallel fashion with stoichiometries of 5 nucleotides per protein monomer. The proposed protein ssDNA interactions involve electrostatic binding of the nucleic acid phosphodiester bond by residues Arg 16, Arg 21, Arg 80 and Lys 46 and stacking of the nucleic acid bases with the side chains of Tyr 26, Tyr 34 and Tyr 41 of one monomer and Phe 73 of the opposite monomer. In the model, the observed cooperativity in the binding of ssDNA to GVP has been explained in terms of an interaction between GVP dimers located on well defined domains of the individual protein molecules.

However, in previous papers physico-chemical observations contrasting the proposed ssDNA GVP interaction model as well as of features of the crystal structure have been reported (Gray et al., 1984; King & Coleman, 1987; 1988). One of the fundamental differences is the chosen stoichiometry of 5 nucleotides per monomer in the interaction model despite the fact that stoichiometries of 3 nucleotides per monomer for oligonucleotide protein complexes (n=3 binding mode) and 4 nucleotides per monomer for polynucleotide protein complexes (n=4 binding mode) have been reported (Alma et al., 1981a; Bulsink et al., 1986).

NMR studies have suggested the involvement of only two aromatic amino acid residues in the binding to ssDNA (King & Coleman; 1987). The assignments of these residues has been based on NOESY spectra recorded in D_2O solution combined with the knowledge of the crystal structure. However, ^1H NMR assignments which are based on crystallographic data may easily lead to errors. In order to obtain more information on the possible DNA binding domain and to unequivocally establish the roles of the individual aromatic amino acid residues in the complex formation with ssDNA, we started with the characterization of missense GVP mutants which have been constructed by means of saturation and site-directed mutagenesis techniques (Stassen et al., 1992).

This Chapter is the first in a series of NMR studies on M13 GVP. It describes general spectral and structural features of both wild-type GVP and a variety of GVP mutants. Special attention is paid to the GVP mutant proteins in which the aromatic residues are substituted for other residues. Solution properties of wild-type and mutant proteins such as stability and solubility are investigated. Optimal conditions for NMR experiments are established and the DNA binding domains of the proteins are explored by means of spin-labeled oligonucleotides. The results are compared with those obtained in earlier studies.

MATERIALS AND METHODS

Protein isolation and preparation. The procedure used for the construction and expression of a library of M13 GVP mutants is described by Stassen et al. (1992). The level of expression of the mutant proteins can be divided into two categories. The low producers (wild-type, GVP Y41H and Y41F) reach a level of expression which is comparable to that in M13 infected cells and where the protein is present in the soluble fraction of the cells. The high producers (GVP Y26C, Y34H, Y61H and F73L) reach, for reasons still unknown, a level of expression which is at least ten times higher. These proteins accumulate in the form of electron dense particles called "inclusion bodies". The soluble proteins were isolated as described previously (Garssen et al., 1977). The gel filtration step was, however, replaced by chromatography on a MonoS cation-exchange FPLC column (Pharmacia). After dialysis against 1 mM cacodylate buffer (pH 6.9) containing 50 mM NaCl, the purified GVP was lyophilized and stored at -20 ^{0}C.

A modified procedure was developed to isolate insoluble mutant proteins (unpublished results). In short, after sonification and subsequent centrifugation of the cells, the pellet, which contained the insoluble mutant protein was resuspended in 6M guanidiniumchloride/0.02 M Tris.HCl pH 7.8/ 5 mM EDTA/ 1 mM ß-mercaptoethanol/ 0.05 M NaCl/ 10% w/v glycerol buffer during 4-15 hours at 4 ^{0}C. Then, the protein was renatured by means of dialysis in the same buffer without the guanidiniumchloride. After subsequent centrifugation, the soluble protein fraction was loaded on a DNA cellulose affinity column. From this point, the normal procedure for GVP isolation was reinstated.

After dialysis against 1 mM cacodylate buffer (pH 6.9) containing 50 mM NaCl, all purified proteins were lyophilised and stored at -20 ^{0}C. The GVP mutant Y41H contains a histidine residue instead of a tyrosine residue at position 41. The other mutants are named in a similar manner.

The protein monomer concentrations were determined by extinction measurements at 276 nm (pH 6.9) using the absorption coefficients of 6400 $M^{-1}cm^{-1}$ for GVP Y26C, 5150 $M^{-1}cm^{-1}$ for GVP Y34H, 5680 $M^{-1}cm^{-1}$ for GVP Y41F, GVP Y41H and GVP Y61H, 7100 $M^{-1}cm^{-1}$ for wild-type GVP and 8000 $M^{-1}cm^{-1}$ for GVP F73L. The absorption coefficients of the mutant proteins were determined relative to the wild-type GVP by means of a Bio-Rad Protein Assay. Samples of wild-type (wt) and mutant GVP were measured in D_2O at very low ionic strength with a concentration ranging from approximately 0.5 - 4.0 mM dependent on the sample quality. The pH of the samples which ranged from pH 5.1 to 7.5 was adjusted with diluted DCl.

Spin-labeled oligonucleotides. The spin-labeled trinucleotide, three adenyl residues, to which the spin label, 4-hydroxy-1-oxyl-2,2,6,6-tetramethyl piperidine (TEMPO), was covalently attached via phospho-diester bonds at both the 5' and 3' end 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 "*" 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. 37500 $M^{-1}cm^{-1}$. The binding experiments were carried out by adding small amounts of a concentrated solution of the spin-labeled oligonucleotide to the protein solutions. 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.

NMR measurements. ^1H NMR experiments were performed at 400 MHz on a Bruker AM400 spectrometer interfaced to an ASPECT 3000 computer.
MLEV17 TOCSY experiments (Bax & Davis, 1985; Griesinger et al., 1988) were conducted at 298 K unless otherwise stated, with mixing times around 30 ms. The residual water resonance was suppressed using DANTE (Morris & Freeman, 1978). 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.

RESULTS

Optimization of solution conditions for ¹H NMR analysis of M13 wt GVP. Conditions under which reasonable one-dimensional ¹H NMR spectra of M13 wild-type (wt) GVP can be recorded were established previously (Alma et al., 1981a). Typically, the NMR samples contained 50 mM NaCl, 1 mM NaCacodylate, pH 7.0, and up to 1 mM GVP. However, we have discovered that it is possible to further optimize the solution conditions.

Wild-type M13 GVP is particularly sensitive to alterations in protein concentration, salt concentration and pH. In Figure 1, changes which occur as a function of salt concentration are shown for the aromatic region of the ¹H NMR spectrum of M13 wt GVP. The peaks belonging to a single Tyr residue which had been designated as Tyr IV by Alma et al. (1981a), shift upfield when the salt concentration is increased. This effect is accompanied by a broadening of all spectral lines, indicative of an increase in aggregation. The increase of the protein concentration results in spectral alterations similar to those observed when the salt concentration is increased (data shown in de Jong et al., 1987). Lowering the pH down to approximately 5.1 reverses the aggregation effect to some extent, while a further decrease results in rapid denaturation of the protein. Once denatured, it is not possible to renature the protein by increasing the pH. Hence, NMR measurements of M13 wt GVP can be performed best at very low ionic strength and at pH 5.1. Under these conditions, it was possible to study samples with concentrations up to 1.5 mM M13 wt GVP without inducing too significant aggregation and with reasonably narrow lines. The temperature range at which experiments can be conducted with these relatively concentrated samples is however very limited. Typically, NMR experiments can be performed at a temperature around 298 K.

Analysis of the aromatic region of the ¹H NMR spectrum. The aromatic part of the ¹H NMR spectrum of M13 wt GVP has been investigated thoroughly by Alma et al. (1981a). M13 wt GVP contains one histidine, three phenylalanine and five tyrosine residues. The individual spin systems of all aromatic residues have been identified. The unique histidine
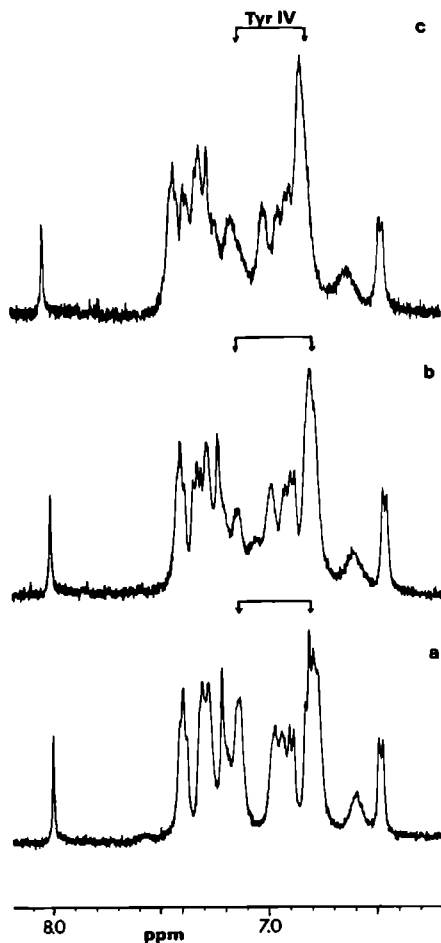
Figure 1. Salt concentration dependency of the aromatic region of the M13 wt GVP spectrum. The arrows indicate the resonance positions of the Tyr IV spin system. The sample contained 0.5 mM GVP, pH 7.5. Salt concentration: (a) low ionic strength (b) 25 mM NaCl. (c) 100 mM NaCl.

residue in GVP at position 64 could be sequence-specifically assigned. In Figure 2, a ^{1}H NMR spectrum of the aromatic part of M13 wt GVP is presented in which the assignments of the individual protons to residue type are indicated by roman numbers. The results were obtained by means of photo CIDNP experiments (Garssen et al., 1978), selective decoupling experiments and 1D NOE-difference measurements (Alma et al., 1981a). Nowadays, all coupled resonances in the aromatic region can be identified using two dimensional Double

Figure 2. The aromatic part of the ^1H NMR spectrum of 0.8 mM M13 wt GVP (pH 5.1). The assignments of the individual protons to residue type are indicated by roman numbers.
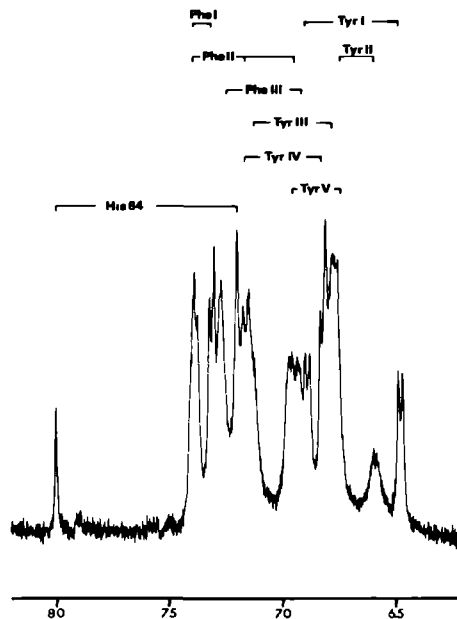
Quantum Filtered COSY (Rance et al., 1983) and total correlation (TOCSY) spectra (Griesinger et al., 1988). An example of a TOCSY spectrum of M13 wt GVP is given in Figure 3. Picture (A) displays a protein with a low degree of aggregation as can be judged from the position of the Tyr IV spin system. In this situation, the Tyr IV crosspeak is downfield shifted and is positioned left from the crosspeak connecting the resonances which make up the Tyr III spin system. At the insets in Figure 3, a part of the aromatic region of two other TOCSY spectra of M13 wt GVP is displayed which reflect the influence of an increased aggregation. The crosspeak belonging to the Tyr IV spin system is, in spectrum B, right on top of the crosspeak of spin system Tyr III and in spectrum C even further upfield shifted. In the latter case, the resonances of spin system Tyr V are also slightly shifted.

Tyr II gives rise to broadened resonances indicative of intermediate exchange between the two possible chemical shift positions for the pairs of Hδ and Hε protons of the ring (Wüthrich, 1986). The remaining four tyrosines generate relatively narrow resonances characteristic of an aromatic ring flipping rapidly about the Cβ-Cγ bond such that the Hδ1 and

Hδ2 resonances and the Hϵ1 and Hϵ2 resonances are superimposed. Similarly, the Phe resonances exhibit a degenerate spin pattern due to rapid ring flips.
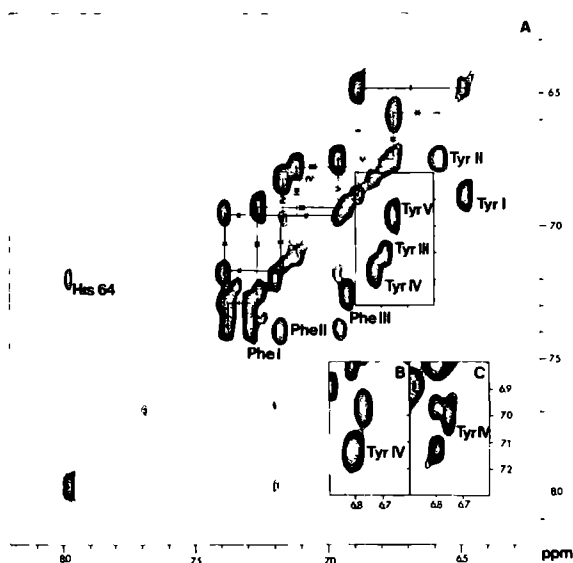
Figure 3. *(a) 30 ms TOCSY spectrum of the aromatic region of 0.5 mM M13 wt GVP at low ionic strength (pH 7.1). The individual spin systems are indicated by connecting arrows. The assignments to residue type are taken from Alma et al. (1981a) (b) TOCSY spectrum of a part of the aromatic region of wt GVP (1 mM pH 7.0) at low ionic strength. (c) TOCSY spectrum of a part of the aromatic region of wt GVP at higher ionic strength (1 mM, pH 7.5), displaying upfield shifted Tyr IV resonances.*

Characterization of mutants of the aromatic residues of M13 GVP. Missense mutants were constructed for four out of five tyrosines and one out of three phenylalanines. The level of GVP synthesis in the cells varies significantly with the mutants investigated. The level of expression of the wild-type protein and of the GVP mutants Y41F and Y41H is comparable to that of gene V protein in phage M13 infected cells. These recombinant proteins are present in the cytoplasmic fraction and isolated as described previously (Garssen et al., 1977). The expression level of the GVP mutants Y26C, Y34H, Y61H and F73L is, for reasons still unknown, at least ten times higher and their encoded protein accumulate as inclusion bodies in the cell. The proteins were isolated from these particles using a denaturation step with
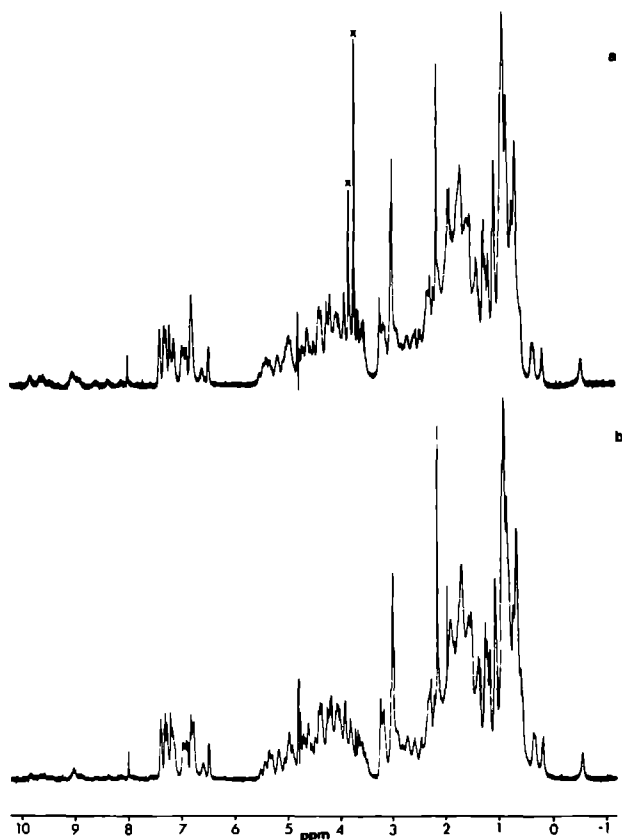
Figure 4. ^{1}*H NMR spectra of M13 wt GVP. (a) wt GVP which was subjected to a denaturation step with guanidiniumchloride (0.5 mM pH 5.2). The impurities which were introduced are indicated by an x. (b) wt GVP taken from the same batch as a control (0.5 mM pH 5.2)*

guanidiniumchloride (unpublished results). A significant amount of mutant protein F73L could however also be isolated from the soluble fraction of the cell extract.

To test whether the renaturation procedure resulted into a proper folding of the polypeptide, ^{1}H NMR spectra of wild-type M13 GVP isolated from the cytoplasmatic fraction and wild-type GVP that was subjected to the denaturation and renaturation procedure and derived from the same protein batch were compared (Figure 4). It can be concluded that these spectra are very similar, thus indicating that the renaturation procedure developed results in a proper refolding of the denatured protein. Close examination of the aromatic regions of the

spectra, however, revealed a small difference in aggregation state between both samples, with the control wild-type sample, being less aggregated.

Optimal samples of all mutant proteins were obtained at very low ionic strength, a property which is similar to wt GVP. The differences in quality of the mutant GVP samples did however vary significantly. For example, GVP mutant Y61H is very unstable. Samples of Y61H could only contain up to 0.5 mM protein and experiments could only be conducted at a maximum temperature of 295 K. Other experiments such as salt titrations and pH titrations could not be performed as yet, for this protein.

The mutant proteins GVP Y26C and F73L behave very similar in solution and have a sample quality that is comparable to wt GVP. Samples of GVP F73L isolated both from the soluble fraction and the insoluble fractions were also compared. The ^{1}H NMR spectra of both samples were almost identical with a small difference in aggregation state in favour of the sample which was isolated from the soluble fraction.

The proteins GVP Y34H, GVP Y41F and GVP Y41H clearly exhibit different properties than wt GVP. The ^{1}H NMR spectra of these samples clearly indicate that the aggregation is much less severe than in wt GVP. The solution properties of these samples were therefore further studied by means of salt titrations, pH titrations and concentration dependency tests. Good quality ^{1}H NMR spectra of GVP Y34H were obtained up to a protein concentration of 3 mM. However, samples of GVP Y34H were not as stable as wt GVP which made it more difficult to further test its solution properties. The pK of the histidine introduced at position 34 is lower than 5.0. An exact determination could not be made because of rapid protein denaturation below pH 5.

At low ionic strength, the GVP mutants Y41H and Y41F gave rise to ^{1}H NMR spectra with relatively sharp resonances up to a concentration of approximately 5 mM (pH 5.1). At these concentrations, the proteins are stable for at least a few days at a temperature of 298 K. Increase of the temperature by only a few degrees results in rapid denaturation. Both samples are less sensitive to changes in salt and protein concentration than samples of wt GVP. In Figure 5, the aromatic spectrum of the sample Y41H is shown at two different protein concentrations. The spectral alterations which can be observed are not as prominent as in the case of wt GVP and mainly involve a line broadening effect going to a higher concentration. The line widths of the individual resonances still remain relatively small at a
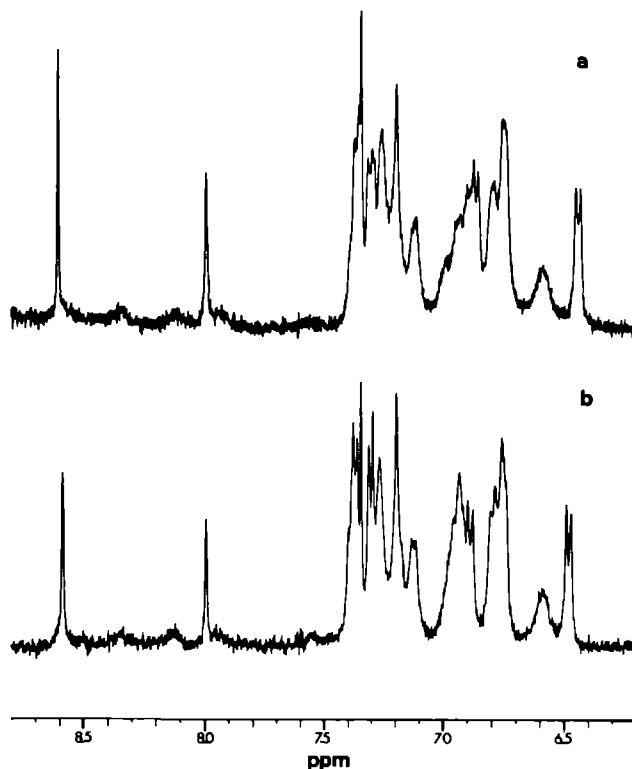
Figure 5. *Concentration dependency of the aromatic region of the ^1H NMR spectrum of M13 Y41H GVP (pH 5.2). M13 Y41H GVP concentrations (a) 0.5 mM (b) 4 mM*

concentration of 4 mM. The peaks belonging to the spin system Tyr IV in the spectra of wt GVP are missing. The histidine introduced at position 41 in the Y41H mutant has a pK of 6.5, indicative of a residue which is exposed to the solvent.

Sequence-specific assignment of the aromatic residues. Sequence-specific assignments for the aromatic region of wt GVP were obtained by making a comparison of the TOCSY spectra of all aromatic mutant samples. The applicability of this approach is limited to the case where the spectra of wild-type and mutant protein are very similar. In Figure 6, the aromatic regions of the TOCSY spectra of wt GVP and of six mutant proteins are displayed. It can be seen that apart from the absence of a few resonances most of the aromatic spectra of mutant GVP are
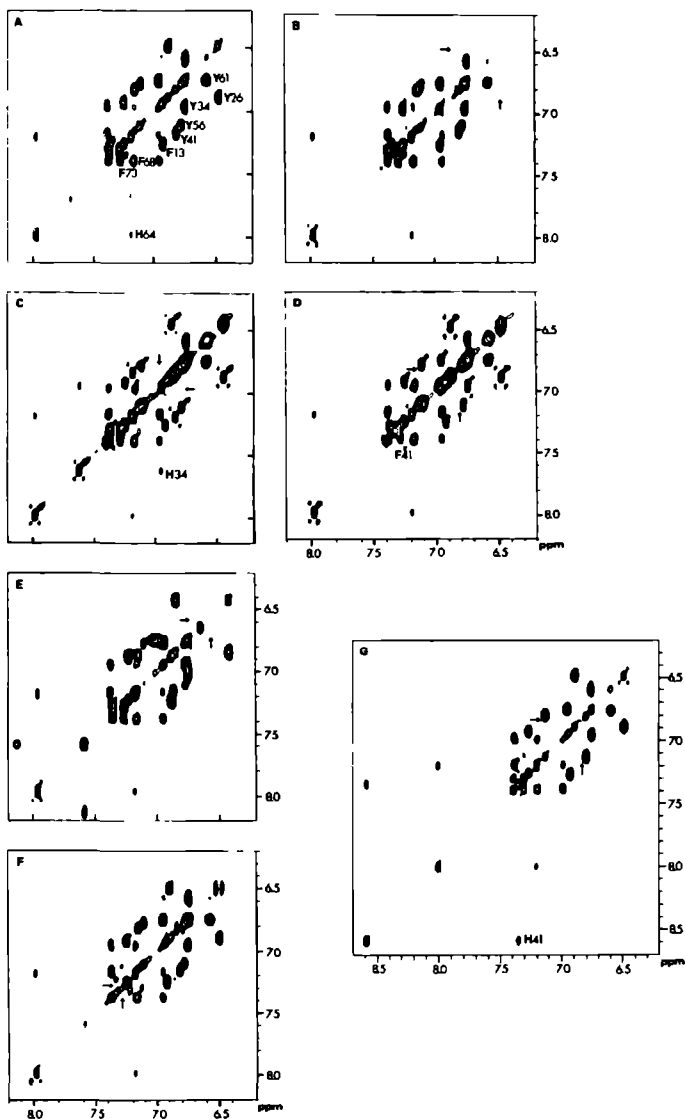
Figure 6. Aromatic regions of TOCSY spectra of M13 wt GVP and a variety of aromatic GVP mutants. All spectra were recorded at 298 K unless stated otherwise. The mixing times used were approximately 30 ms. Sequence-specific assignments are indicated in the figure near (one of) the crosspeaks connecting the resonances of the spin system (see figure 3). (a) 0.5 mM wt GVP (pH 7.1); (b) 1.0 mM GVP Y26C (pH 7.0); (c) 3 mM GVP Y34H (pH 5.2); (d) 4 mM GVP Y41F (pH 5.2); (e) 0.5 mM GVP Y61H at 295 K (pH 7); (f) 0.5 mM GVP F73L (pH 5.2); (g) 1 mM GVP Y41H (pH 5.1)

almost identical to that of wt GVP. Some resonances undergo small shifts but could nevertheless easily be identified. In Table I, a summary of the assignments obtained for wt GVP is presented. The distinction between δ and ε resonances which in part already has been made previously (Garssen et al., 1980; Alma et al., 1981a; de Jong et al., 1987), was reevaluated by means of a NOESY (D_2O) spectrum of wt GVP which showed NOE crosspeaks between the Hβ proton resonances and the Hδ ring proton resonances (data not shown). In the spectrum of the GVP mutant Y26C, the cross peaks connecting the resonances at 6.48 ppm and 6.88 ppm which earlier had been designated as Tyr I in wt GVP, are clearly missing (Figure 6B). The rest of the aromatic spectrum is hardly different from that of wt GVP and therefore Tyr I can be ascribed to Tyrosine 26. Similarly, in the spectrum of GVP Y34H, the Tyr V spin system is missing (Figure 6C). The spectrum displays a downfield shifted Tyr IV, indicative of a decreased level of aggregation. This observation is in agreement with the improved spectral quality of GVP Y34H in comparison to wt GVP.

The tendency to aggregate is even less in the GVP mutants Y41F and Y41H, in which the Tyr IV spin system is missing. Now it can be concluded unambiguously that the residue which accounts most for the aggregation effect in wt GVP is the tyrosine residue at position 41 (Figures 6D & 6G).

The aromatic spectrum of the mutant GVP Y61H clearly lacks the broad resonances at 6.58 ppm and 6.75 ppm in wt GVP (Figure 6E). Tyrosine 61 makes up spin system Tyr II and the only remaining tyrosine residue at position 56 can now by assigned to spin system Tyr III by elimination. Due to the availability of only one phenylalanine mutant, a complete assignment of all phenylalanine residues using the site specific mutants could not be made. The spectrum of the GVP mutant F73L can be superimposed on the spectrum of wt GVP except for the missing spin system Phe I (Figure 6F).

The assignment of the phenylalanine residues could however be completed using the sequence-specific resonance assignment procedure, which could be applied to the GVP mutant Y41H (described in the next Chapter). The reported sequential assignment of a β-loop of this molecule, encompassing residues 13 through 31, included the aromatic residues phenylalanine 13 and tyrosine 26. Because of the similarities between wt GVP and mutant GVP Y41H, the sequential assignments of GVP Y41H were used to complete the assignment of the aromatic resonances in wt GVP. Indeed, the assignment of tyrosine 26 in GVP Y41H agrees with the

Table I

*Resonance assignments for the aromatic region of M13 wt GVP at very low ionic strength,
pH 7.1, 298 K. Chemical shifts are expressed relative to 2,2-dimethyl-2-silapentane-5-
sulfonate (DSS). The residue specific assignments according to Alma et al. (1981a) appear
in parentheses*

Residue	chemical shift		
	δ	ε	ζ
Tyr 26 (Tyr I)	6.88	6.48	-
Tyr 34 (Tyr V)	6.96	6.75	-
Tyr 41 (Tyr IV)	7.16	6.82	-
Tyr 56 (Tyr III)	7.12	6.78	-
Tyr 61 (Tyr II)	6.58	6.75	-
Phe 13 (Phe III)	6.92	7.25	7.25
Phe 68 (Phe II)	7.39	7.17	6.95
Phe 73 (Phe I)	7.31	7.39	7.31
	$\delta 2$	$\varepsilon 1$	
His 64	7.19	7.97	

assignment based on mutant studies. Phenylalanine 13 had not been assigned before and can
now be attributed to the Phe III spin system. Consequently, spin system Phe II can be
assigned to Phe 68 by elimination.

Binding of wild-type and mutant gene V proteins to spin-labeled oligonucleotides.
Binding of wt GVP to oligonucleotides causes upfield shifts of specific aromatic resonances.
Significant shifts have been observed for one phenylalanine and two tyrosine residues (de
Jong et al., 1987; King & Coleman, 1987). However, for one of the tyrosine residues the shift
does not only depend on the degree of saturation of the protein with ssDNA but also on the
total protein concentration (de Jong et al., 1987). Therefore, it has been concluded that only
two aromatic residues are involved in binding to ssDNA (de Jong et al., 1987; King &
Coleman, 1987). These can now definitely be identified as tyrosine 26 and phenylalanine 73.
Apart from binding normal oligonucleotides to GVP, it is also possible to probe the ssDNA
binding domain and its surroundings with spin-labeled oligonucleotides (de Jong et al.,
1989b). The spin-labeled ligand bound to a macromolecule selectively broadens the

resonances which are in close vicinity of the spin label. Difference spectra computed from spectra obtained with and without spin-labeled ligand, will thus display the resonances originating from the residues which are situated in the region of the binding domain.

The changes observed in the aromatic part of the spectrum upon binding of spin-labeled oligonucleotides to M13 wt GVP are congruent with the changes observed in binding of non-modified oligonucleotides. This is shown in Figure 7, in which a comparison is made of the aromatic spectra of M13 wt GVP in the absence and presence of small amounts of *d(A)3*. The aromatic resonances of Phe 73 and Tyr 26 are erased from the spectrum. The difference spectrum also reveales some shifts that are caused by the change in sample conditions upon addition of the spin-labeled DNA. Very prominent in this respect is the shift of spin system Tyr IV which is now assigned to Tyr 41. This effect can be reduced by diluting the sample which further indicates that Tyr 41 does not form a part of the DNA binding domain. Similar experiments were performed on mutant GVP samples. A few examples are shown in Figure 7. In GVP Y34H, Tyr 26 and Phe 73 show up in the difference spectra. Residue Tyr 41 shifts as well but to a lesser extent than for the wt GVP. The binding domain seems also to be unaffected in the GVP mutants Y41H and Y41F. Despite the absence of the Tyr 41 residue and the reduced aggregation, it can be seen that the shift effects did not vanish: apart from small shifts of the introduced residue (histidine/phenylalanine) at position 41, other shifts could be observed, such as those of Tyr 34 and Phe 13. These residues were identified in 2D-difference spectra (data not shown). The mutants F73L and Y26C still bind to the spin-labeled oligonucleotide despite the absence of one of the two important aromatics. In the difference spectra of mutant GVP F73L the tyrosine residue at position 26 can be seen. The difference spectra of the GVP Y26C mutant shows the Phe 73 residue. These data suggest that the major part of the DNA binding domains of these mutants is still intact.

DISCUSSION

Optimized conditions have been obtained for measurement of the ^{1}H NMR spectrum of M13 wt GVP. These conditions should also be used for measurements of several mutant GVP samples we have at our disposition. From the results presented, it can be concluded that
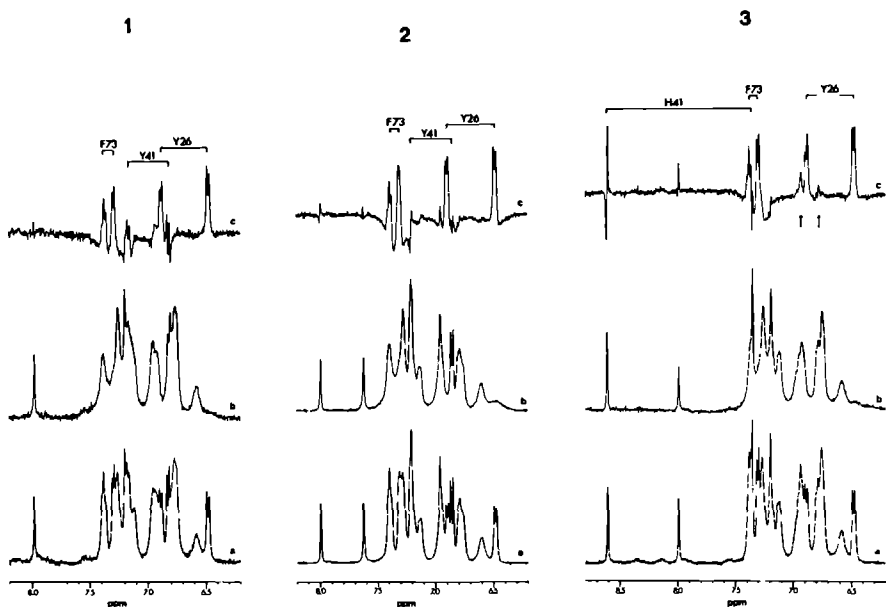
*Figure 7. Examples of spectra of the aromatic region of wt GVP (1), GVP Y34H (2) and GVP Y41H (3), recorded in the absence and presence of the spin-labeled oligonucleotide *d(A)3*. The small arrows in the GVP Y41H spectrum (3) indicate the presence of small shift effects of respectively Phe 13 (left arrow) and Tyr 34 (right arrow).*
*(a) is the spectrum in the absence of spin-labeled oligonucleotide; (b) is the spectrum after addition of 1/25 equivalent *d(A)3*(1), of 1/45 equivalent *d(A)3* (2) and 1/12 equivalent *d(A)3* (3), respectively; (c) is the difference between (a) and (b).*

NMR studies on wild-type and mutant M13 GVP can best be conducted at very low ionic strength and at pH 5.1.

M13 wt GVP has a high tendency to aggregate. The aggregation effect could be evaluated by considering the aromatic region of the ^{1}H NMR spectrum of wt GVP under different solution conditions. Upon increase of the protein and/or salt concentration two effects occur. The first effect is a broadening of all spectral lines. Secondly, the peaks belonging to a single tyrosine residue, which has now been assigned to Tyr 41 shift upfield. In this respect, the behaviour of M13 wt GVP is very similar to that of the previously studied IKe GVP. IKe GVP also contains a tyrosine spin system which is very sensitive to changes in solution conditions (de Jong et al., 1987). This particular tyrosine spin system in IKe GVP has been assigned to Tyr 42, by means of sequential assignment procedures (de Jong et al.,

1989a). Because of the homology which exists between both proteins and the similarity in behaviour of this specific residue, it had already been suggested that this residue in M13 wt GVP is Tyr 41. Using missense mutants we have been able to show that this is indeed the case.

The analysis of the spin patterns in the aromatic region of the 1H NMR spectrum of M13 wt GVP by means of TOCSY spectra is in agreement with the analysis made by Alma et al. (1981a). However, it differs slightly from the data presented by King & Coleman (1987). Their NMR samples contained between 1 and 1.8 mM wt GVP in the presence of 25 mM phosphate buffer pH 7.6. NOESY experiments were conducted at temperatures ranging from 298-303 K. The major difference between their and our data is the Tyr IV spin pattern which in their spectra exhibits four different chemical shifts positions for the protons of the ring, indicating that the residue is immobilized, while in our case, it appears to be rapidly flipping around the Cβ-Cγ bond. This indicates that at their conditions, aggregation of the protein has occurred which led to the immobilisation of the tyrosine ring.

The solution properties of the GVP mutants we have studied differ significantly from one another. One factor affecting the quality of the protein is the protein isolation procedure used. For a few mutant clones, the expressed proteins accumulate in inclusion bodies in the cells and were isolated via a denaturation step. It could be shown that although the proteins isolated via this procedure did have a correct fold, the level of aggregation was clearly increased relative to the proteins which were isolated from the cytoplasmic fraction of the cell extract.

However, the solubility of the various mutants cannot be classified according to the isolation procedure used. GVP mutant Y61H which was isolated from inclusion bodies of the *Escherichia coli* cells is hardly soluble at concentrations at which NMR spectra can be recorded. The spectrum of the aromatic region contains an extra crosspeak in comparison to wt GVP. A possible explanation is that the Tyr 41 is slowly flipping indicative of the severe aggregation and low solubility. The increased aggregation of this particular mutant in comparison to wt GVP cannot be explained as yet.

The other set of mutants can be divided into two categories. The first category comprises mutants, such as GVP F73L and GVP Y26C which have properties which are similar to wt GVP. The second category consists of the mutants which have improved

solubility characteristics. The mutants GVP Y34H, GVP Y41H and GVP Y41F fall into the latter category. Increase of protein concentration and/or salt concentrations in wt GVP and mutants GVP Y26C and GVP F73L results in broadening of all spectral lines combined with the prominent shift of the Tyr 41 resonances. The specific shift of the tyrosine 41 residue can also be seen upon addition of normal oligonucleotides as well as spin-labeled oligonucleotides to wt and these mutant proteins. It is very likely that the tyrosine 41 residue plays a crucial role in the specific recognition of GVP dimers. On the basis of a model study, Brayer & McPherson have proposed that hydrophobic interactions between the protein surfaces of two dimers may form the physical basis for the observed cooperativity in the binding to ssDNA. The Tyr 41 residue is part of one of the suggested interaction sites. From these considerations, it can be suggested that the specific aggregation has to be explained on the basis of a set of interactions between dimers responsible for the cooperativity in the binding to ssDNA. Substitution of the crucial tyrosine 41 residue resulted in changes in the solubility characteristics of the protein. The strong interaction between the GVP dimers is clearly diminished in the GVP mutants Y41H and Y41F. This result further supports the hypothesis that the tyrosine 41 residue plays a dominant role in the interdimer interactions. From the DNA binding experiments and the TOCSY spectra of wt GVP at different conditions in combination with 2D NMR experiments (data not shown) it can be seen that for example the tyrosine residue at position 34 undergoes a minor shift under a change in solution conditions. This implies that this residue also might be part of the possible dimer-dimer interaction surface. However, it can also be explained as an intradimer effect in which the tyrosine 41 residue affects the tyrosine 34 residue.

The solubility and the spectral appearance of the GVP mutant Y34H has improved in comparison to the wt GVP. In line with the considerations made above this can be either an intradimeric or an interdimeric effect. The reduced stability of the Y34H mutant cannot be explained as yet.

The differences in solution properties could be explained in terms of a specific residue by virtue of the complete sequence-specific assignment of all aromatic residues. Previously, sequence-specific assignments of wt GVP have been obtained for all aromatic residues on the basis of NOESY spectra and the crystallographic data of wt GVP (King & Coleman, 1987). Our results differ from those reported earlier with respect to the assignments of tyrosine 34

and tyrosine 41. It is noted that the assignment of tyrosine 41 had already been corrected before on the basis of a mutant protein sample of which to our knowledge details never have been published (King & Coleman, 1988). This indeed indicates the necessity for unambiguous assignments which is difficult when only founded on a crystallographically determined structure. The next Chapters will reveal that parts of the secondary structure of M13 GVP indeed differ from the structure postulated on the basis of the X-ray crystallographic studies. This again emphasized the potential danger of making assignments which are solely based on crystallographic data.

Although the availability of aromatic mutant samples was limited to only 5 out of a total of 8 unassigned aromatic residues, a complete assignment has been made. For mutant GVP Y41H of which the solubility characteristics are significantly better than those of wt GVP, it is now possible to perform 2D NMR measurements in H_2O solution, a prerequisite for application of the sequence-specific resonance assignment approach (described in the next Chapter). Such measurements are very hard to perform on wt GVP because the aggregation is too severe to obtain spectra with a sufficiently suppressed H_2O resonance and an interpretable spectral appearance. The sequence-specific assignments, described for GVP Y41H in the next Chapter, included the phenylalanine residue at position 13 which was enough to complete the assignment of the aromatic spectrum.

The results of binding experiments of the spin-labeled oligonucleotides to wt GVP and to aromatic mutant proteins are in line with results of NMR studies reported previously on M13 wt GVP and the GVP of the evolutionary related phage IKe (King & Coleman, 1987; de Jong et al., 1987). From the binding experiments, it can be concluded that only two aromatic residues namely tyrosine 26 and phenylalanine 73 are involved in the binding to oligonucleotides. These data are not in agreement with model building proposals of Brayer & McPherson which were based on their reported crystal structure. In their proposals, it is suggested that the aromatics tyrosine 34 and tyrosine 41 are also involved in binding to ssDNA. Our data and those previously published by King & Coleman (1987) clearly suggest that these residues are not involved in the binding to oligonucleotides. NMR binding studies with longer ssDNA fragments which may be more relevant with respect to complex formation under cooperative binding conditions have not altered that view (King & Coleman, 1988). The tyrosine 41 residue instead plays an important role in interdimer interactions which may be

relevant to cooperativity in binding of GVP to ssDNA which we have discussed above.

Recently, a new model has been proposed for the complex of GVP with ssDNA (Hutchinson et al., 1990). In the model, Tyr 26, Phe 73 and Tyr 34 interact with nucleotide bases. It had been suggested that Tyr 34 interacts with nucleic acid in the n=4 binding mode but not in the n=3 binding mode. Thus far, our studies with spin-labeled oligonucleotides indicate that in the case of oligonucleotide binding, Tyr 34 is neither part of nor close to the DNA binding domain.

CHAPTER 4

Structure of the DNA Binding Wing of the Gene V Encoded Single-Stranded DNA Binding Protein of the Filamentous Bacteriophage M13[*]

ABSTRACT

The structure in solution of a ß-loop in mutant Y41H of the single-stranded DNA binding protein encoded by gene V of the filamentous phage M13 has been elucidated using 2-dimensional ^{1}H nuclear magnetic resonance techniques. Furthermore, these studies enabled us to demonstrate that an identical structural element is present in wild-type gene V protein and that this element intimately is involved in the binding of gene V protein to single-stranded DNA. It is shown that the structure of the DNA binding wing deviates from that proposed for the same amino acid sequence on the basis of X-ray diffraction data. The structure is, however, identical to that of the DNA binding wing present in the single-stranded DNA binding protein encoded by the genome of the evolutionary distantly related filamentous phage IKe. The latter observations support our current view that in the binding of these proteins to single-stranded DNA a common structural motif is involved.

INTRODUCTION

In the life-cycle of the filamentous phages, such as Ff (M13,f1, fd), IKe or Pf3, a single-stranded DNA binding protein encoded by gene V (GVP) plays an indispensable role in the DNA replication process (for a review, see Model & Russel, 1988). Recently, by using 2D-NMR techniques, the structure of part of the DNA binding domain of IKe GVP has been elucidated (de Jong et al., 1989a). This structure consists of a ß-loop and comprises residues 16 through 29 of the 88 amino acid residues long polypeptide. It is noted in passing that in solution these proteins occur as dimers (20 kD) and their structure determination by NMR forms a considerable challenge. Alignment of the amino acid sequences of the ssDNA binding proteins encoded by IKe, M13 and Pf3 has demonstrated that the GVP of the latter two contains a stretch of amino acid residues that is highly homologous to that of the aforementioned ß-loop sequence in IKe GVP (de Jong et al., 1989b; Figure 1). This homology

becomes even more apparent when these amino acid sequences are also folded into a ß-loop
and the residues which are situated at the DNA binding face are compared.
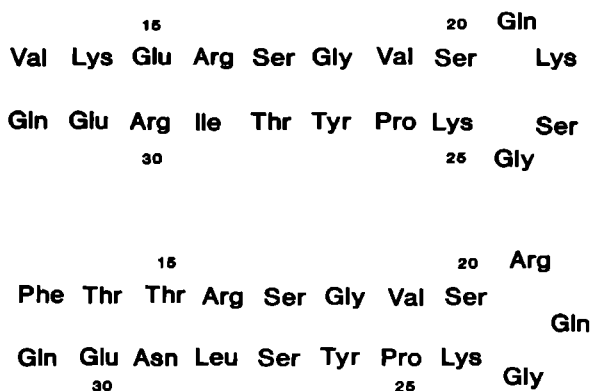
Figure 1. Comparison of the β-loop structures of (a) IKe GVP (residue 13 to 32) and (b) M13 GVP (residue 13 to 31).

Comparison of the hypothetical ß-loop structure in M13 GVP with the structure postulated for the same region on the basis of X-ray crystallographic studies (Brayer & McPherson, 1983) revealed that they differ in two respects. First, the proposed ß-loop structure is more regular than the ß-loop in the crystal structure and second, the residues involved in the formation of the loops are shifted by four amino acids with respect to one another.

In this Chapter, it will be demonstrated that the ß-loop structure, as proposed on the basis of the observed similarities with IKe GVP, is indeed present in both mutant (Y41H) and wild-type M13 GVP. This observation highly supports the thesis that in this class of DNA binding proteins a common structural motif is present for recognition of and binding to ssDNA.

MATERIALS AND METHODS

Wild-type and mutant (Y41H) GVP was prepared and isolated as described previously (Garssen et al., 1977). The gel filtration step was, however, replaced by chromatography on a MonoS cation-exchange FPLC-column (Pharmacia). After dialysis against 1 mM cacodylate buffer (pH 6.9) containing 50 mM NaCl the purified GVP was lyophilised and stored at −20°C. The final concentration of mutant GVP in the NMR-samples was about 1-4 mM , with no buffer components and only a few millimolar of NaCl. The pH was adjusted to pH 5.2 with diluted DCl.

The procedures used for the construction and expression of a library of M13 GVP mutants will be described elsewhere (Stassen et al., 1992). The GVP mutant Y41H, whose solution structure has been investigated in the 2D-NMR studies described, contained a histidine instead of a tyrosine residue at position 41 of wild-type M13 GVP.

The NMR experiments were performed at 600 MHz on a Bruker AM 600 spectrometer interfaced to an ASPECT 3000 computer. Clean TOCSY (MLEV17) (Bax & Davis, 1988; Griesinger et al., 1988), DQF-COSY (Rance et al., 1983) and NOESY (Jeener et al., 1979) experiments were conducted at 298 K. The TOCSY spectra were recorded with mixing times ranging from 24 to 40 ms and the NOESY spectra were recorded with mixing times of 100 and 175 ms. In 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 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/pt). Additional baseline corrections were performed using polynomial functions of standard Bruker software installed on a Bruker X32 computer.

Figure 2. The sequentially assigned residues of Y41H GVP with a summary of dαN, dβN, dNN, dα-(δδ)' contacts.

RESULTS AND DISCUSSION

2D-NMR studies of wild-type M13 GVP are rather difficult to perform because the protein is not soluble in millimolar concentrations. This is particularly harmful for spectra recorded of H_2O solutions because it impairs sufficient suppression of the H_2O resonance. Fortunately, mutation of the tyrosine residue at position 41 of wild-type M13 GVP (the protein itself is 87 amino acids long) into a histidyl residue (Y41H) caused a significant improvement in the solubility characteristics of this protein. As a result measurements of good quality NOESY and TOCSY spectra of H_2O-samples became possible, thus allowing reliable sequential analysis of the protein spectrum. The interpretation of the 2D-NMR data was carried out using standard procedures (Wüthrich, 1986) and as a result an unambiguous sequential assignment could be performed for residues 13 through 31. The sequential connectivities observed for this part of the molecule are summarized in Figure 2, the resonance positions of the amino acids spin systems are listed in Table 1. Frequently, a straightforward sequential analysis procedure is interrupted by a proline residue; this is also true in our situation, namely interruption by proline residue P25. Fortunately, the assignment could be continued by the observation of a sequential dαN connectivity between P25 and Y26 and a dα-(δ,δ') connectivity between K24 and P25. The latter observation is diagnostic for the presence of a normal trans conformation for P25.

From the available NOESY data it can furthermore be concluded that the sequence encompassing residues 13 through 31 consists of a ß-loop composed of an antiparallel ß-ladder (residues 13 to 20 and 24 to 31) and a turn (residues 21 to 23). The ß-ladder is defined by strong sequential dαN contacts and three long range dαα contacts. Because of overlap of α-proton resonances a fourth dαα contact between S17 and S27 could not be observed. Further evidence for the existence of the ß-loop is provided by numerous long range non-sequential interresidue dαN and dNN contacts and by long range NOEs between the aromatic protons of F13 and Y26 and side chain protons of the ß-ladder. A summary of these long range NOEs is presented in Figure 3. The presence of the turn is reflected in the sequential dNN walk from residue 20 to 24. A strong dNN contact between S20 and K24, together with a non-sequential dαN connectivity between N-S20 and α-P25, indicates the presence of a hydrogen bond between the amide of S20 and the carbonyl group of K24. This suggests that the ß-ladder structure includes residue 20 and 24.
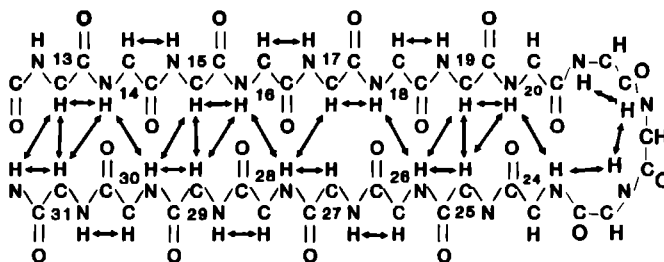
Figure 3. Schematic representation of the structure of the β-loop encompassing residues 13 through 31. Sequential and non-sequential dαN contacts and long range dNN and dαα contacts are indicated by arrows. They are characteristic of an antiparallel β-sheet. Within the turn only sequential connectivities were observed.

As pointed out in the beginning of this section, because of its lower solubility, we have not been able to make sequential assignments for wild-type M13 GVP. The spectra obtained for wild-type GVP solutions in D_2O are however of sufficient quality to permit a detailed comparison with those obtained for the Y41H mutant. From these comparisons it could unambiguously be concluded that the cross peaks arising from the ß-loop in the Y41H mutant are present in the spectra of the wild-type M13 GVP at almost identical positions.

The shifts observed for different peak positions are in the order of the small shifts one expects from slight variations in sample conditions.

Scrutiny of the NOE data shows that the long range NOE connectivities are, with the same intensity ratios, also present in the NOESY spectrum of the wild-type M13 GVP (Figure 4). These results thus demonstrate that in the wild-type GVP the same ß-loop structure is present as in the Y41H mutant. This conclusion can even be extended to the overall conformations/structures of the two protein molecules. For instance, most connectivities between other α-proton resonances have almost identical positions in both the mutant and the wild-type GVP spectra or, in two cases, are shifted only to a minor extent (< 0.10 ppm). The same holds true for the major part of the NOE connectivities of the aromatic ring proton resonances. Thus, the results suggest that substitution of a histidyl residue for a tyrosine residue at amino acid position 41 of wild-type M13 GVP has only a minor effect upon the overall protein structure. Indeed, in an earlier study the resonance of the 3,5-Tyr protons, which now can be assigned to Y41 because of its absence in the Y41H spectrum, showed a strong photo CIDNP effect (Garssen et al., 1978). This indicates that Y41 is situated at the outer surface of the protein. In line with this observation a pH-titration of the mutant GVP protein Y41H demonstrates that the pKa value of the histidyl residue is about 6.5. This value is characteristic of histidines exposed to the solvent. Thus, substitution of the tyrosine residue for the histidyl residue, which both are situated at the protein surface is expected to cause only minimal changes, if at all, in the overall protein structure.

Previous NMR studies have demonstrated that in the binding of M13 GVP to ssDNA both aromatic as well as charged residues are involved (Alma et al., 1981b; Dick et al., 1988). With the aid of NMR-studies on other GVP mutants (data not shown) and the above mentioned sequential assignments some of these residues can now be identified as Y26 and R21, respectively. Other NMR studies, both with spin-labeled and non-labeled oligonucleotides, have furthermore indicated that also the amino acid residues R16, G18, S20 (Folkers et al., 1992), K24 (Dick et al., 1988) and L28 (King & Coleman, 1987) form part of the DNA binding domain. Comparison of these data with the results of the NMR studies performed on IKe GVP (de Jong et al., 1989b) shows that the same (or similar) amino acids are also involved in the binding of the latter protein to ssDNA.

Table 1

Resonance positions of protons in amino acid residues 13 to 31 of mutant (Y41H) M13 GVP at 298 K and pH 5.2. Chemical shifts are quoted relative to DSS.

Chemical shifts (ppm)

Residue	$N^{\alpha}H$	$C^{\alpha}H$	$C^{\beta}H$	Others
F13	7.83	5.40	2.70, 3.26	$C^{\delta}H$ 6.93, $C^{\varepsilon}H$ 7.27, $C^{\zeta}H$ 7.27
T14	9.21	4.73	4.23	$C^{\gamma}H$ 1.28
T15	8.87	5.35	3.92	$C^{\gamma}H$ 1.23
R16	9.36	4.75	1.80, 1.88	$C^{\gamma}H$ 1.60
S17	8.56	5.01	3.81	
G18	7.66	3.54 4.07		
V19	8.41	4.57	1.89	$C^{\gamma}H$ 0.88
S20	9.15	4.70	4.38, 4.08	
R21	9.13	4.26	1.93	$C^{\gamma}H$ 1.76, $C^{\delta}H$ 3.26
Q22	7.93	4.42	1.93, 2.34	$C^{\gamma}H$ 2.50, 2.39
G23	8.19	3.64 4.20		
K24	7.67	4.92	1.93, 1.77	
P25		5.04	1.95, 2.34	$C^{\gamma}H$ 2.15, 2.04, $C^{\delta}H$ 4.09, 3.78
Y26	8.37	4.94	2.98	$C^{\delta}H$ 6.89, $C^{\varepsilon}H$ 6.48
S27	8.39	4.97	3.69, 3.63	
L28	8.94	4.65	1.60	$C^{\gamma}H$ 1.60, $C^{\delta}H$ 0.90, 0.85
N29	9.06	5.47	2.48, 1.62	
E30	8.93	5.54	1.99[a]	
Q31	9.31	4.90	2.03[b]	

a   *A second β-resonance might be present but so far the COSY data don't allow discrimination of this resonance from a γ-resonance.*

b   *NMR data obtained so far could not provide unambiguous evidence for the non-degeneracy of this resonance.*

Putting all results together we are forced to conclude that the observed structure of the β-loop in wild-type as well as mutant (Y41H) M13 GVP is analogous to the β-loop structure present in IKe GVP. The structure of this DNA binding wing deviates, however, significantly from the structure proposed for this stretch of amino acids on the basis of X-ray crystallographic studies of M13 GVP (Brayer & McPherson, 1983). Comparison of these structures reveals that the β-loop structure found in solution is more regular than the one found in the crystal. Furthermore in the solution structure of the DNA binding wing the β-sheet is shifted four amino acid residues with respect to that of the crystal structure: specifically, in the crystal
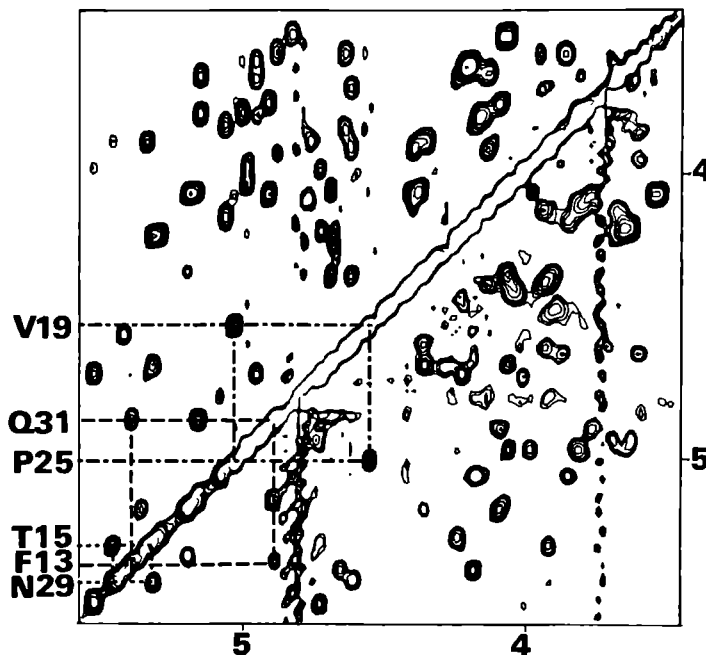
Figure 4. *600 MHz NOESY spectra of mutant (Y41H) GVP (upper diagonal) and wild-type GVP (lower diagonal). Mixing times were 100 ms. Samples contained 4 and 1 mM GVP, respectively. Long range dαα-contacts which define the β-loop (see Figure 3) have been indicated with dashed lines.*

structure residue Y26 is positioned at the tip of a ß-loop structure, while in the solution structure this residue forms part of the antiparallel ß-sheet (Figure 3). In a previous paper already anomalous observations towards crystallographic data have been reported (Gray et al., 1984). In this Chapter however, the NMR results provide unambiguous evidence that the conformation of M13 GVP in solution differs significantly from that in the crystal.

CHAPTER 5

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

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 GVP. The major part of this secondary structure is present as antiparallel ß-sheet, i.e. as two ß-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.

INTRODUCTION

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 f1) infects only cells that contain plasmids of the Inc F incompatibility groups while for host cell penetration by 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) 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 that GVP encoded by

phage Ff also fulfils, 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 on 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 (previous Chapter; 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 ß-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 non evolutionary related ssDNA binding proteins into similar ß-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) 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 basis of crystallographic studies (Brayer & McPherson, 1983). This Chapter reports the nearly complete assignment of the ^{1}H 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 a 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 ^{2}HCl, 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 ms 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 semi selective jump-return (Plateau & Gueron, 1982) pulse sequence. In the TOCSY(H_2O) measurements both continuous irradiation and a semi selective 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 baseline 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 baseline corrections were performed using polynomial functions of standard Bruker software installed on a Bruker X32 computer.

RESULTS

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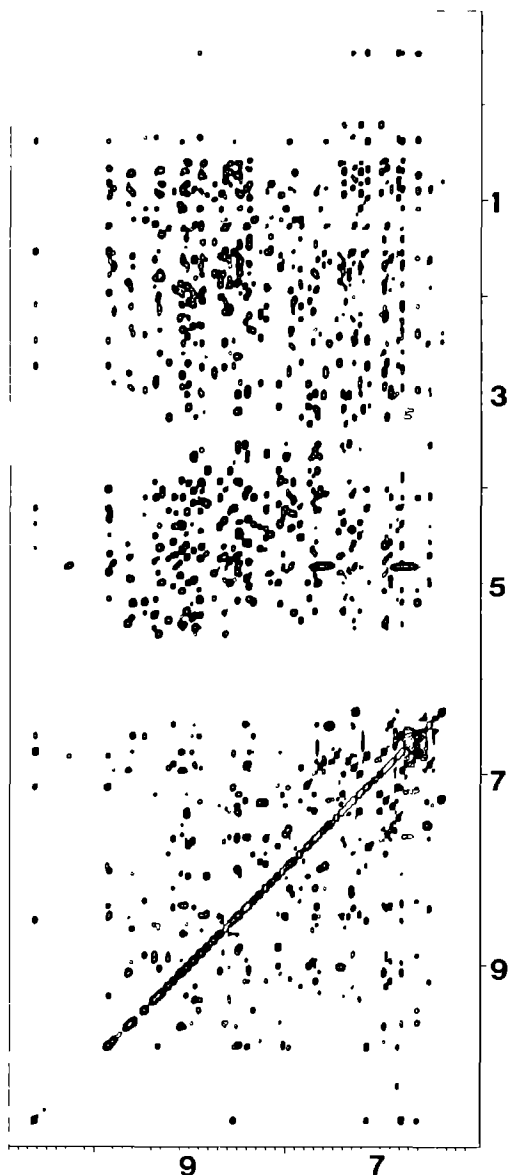
Figure 1. 600 MHz NOESY spectrum of IKe GVP dissolved in 90% H_2O/10% D_2O at a concentration of 2.5 mM and a pH of 4.7. The mixing time was 150 ms and the temperature 298 K. The H_2O-resonance was suppressed by a jump-return pulse sequence.

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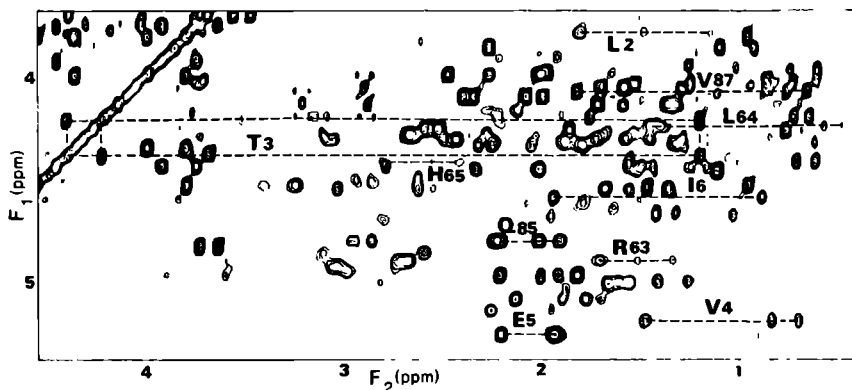
Figure 2. 600 MHz clean-TOCSY (MLEV17) spectrum of IKe GVP (1.5 mM, at pH 4.6, 298 K) dissolved in 99.8% D_2O. The mixing time was 30 ms. For clarity, only the spin systems belonging to residues in the triple stranded β-sheet denoted as III (see text) are indicated.

The amide-resonances, however, suffer from significant more loss of intensity during the spinlock time. Consequently, shorter mixing times have to be applied when working with H_2O-samples (20 and 30 ms at 298 K 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 crosspeaks significantly reduce their intensity and overlap of the crosspeaks hamper 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 crosspeak patterns the poor resolution of the COSY spectra precludes in many cases discrimination between ß- and γ-resonances. Because of the poor appearance of most fingerprint (Nα) connectivities in H_2O-COSY spectra, residue specific assignments for most amide resonances were obtained by combining the spinlock 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 assignment of five valines (out of nine) and four serines (out of
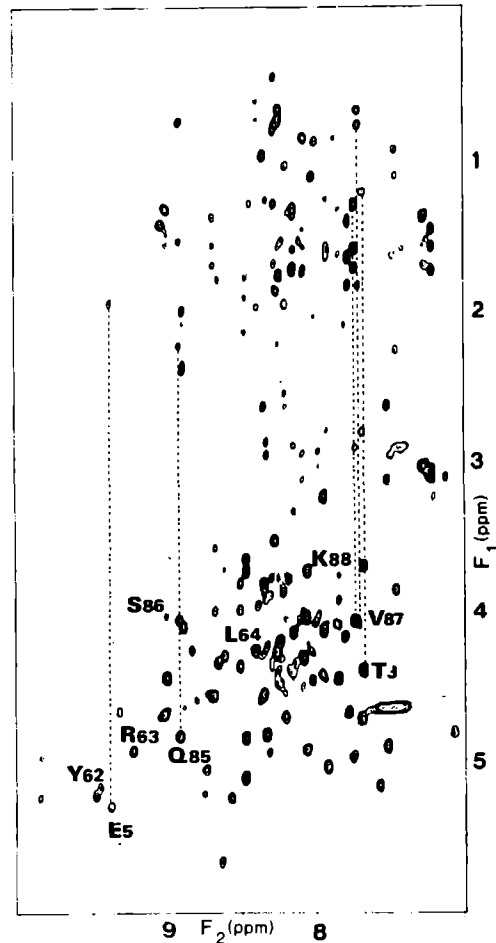
Figure 3. 600 MHz clean-TOCSY (MLEV17) spectrum of IKe GVP (1.5 mM, pH 4.6, 303 K) dissolved in 90% H₂O/10% D₂O. The spin systems in the triple stranded β-sheet structure (denoted as III in the text) are indicated to illustrate the quality of the TOCSY spectrum at these conditions.

seven) were reported. The quality of the 600 MHz data in the present study allows the assignment of the α- and ß-resonances of all aromatic residues by combining NOE data with AMX patterns that could be recognised 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.

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, Q/E and "other" long side chain residues. Many of these spin systems had either incomplete or unresolved crosspeak 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 ^{1}H NMR spectrum. A schematic representation of the results obtained this way is given in Figure 4. One can observe that several sequential dαN walks are obscured by overlap. In most cases, one can get around this problem by using other sequential contacts, i.e. dßN- and/or dγN-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 dαN- or dNN-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 crosspeaks 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.
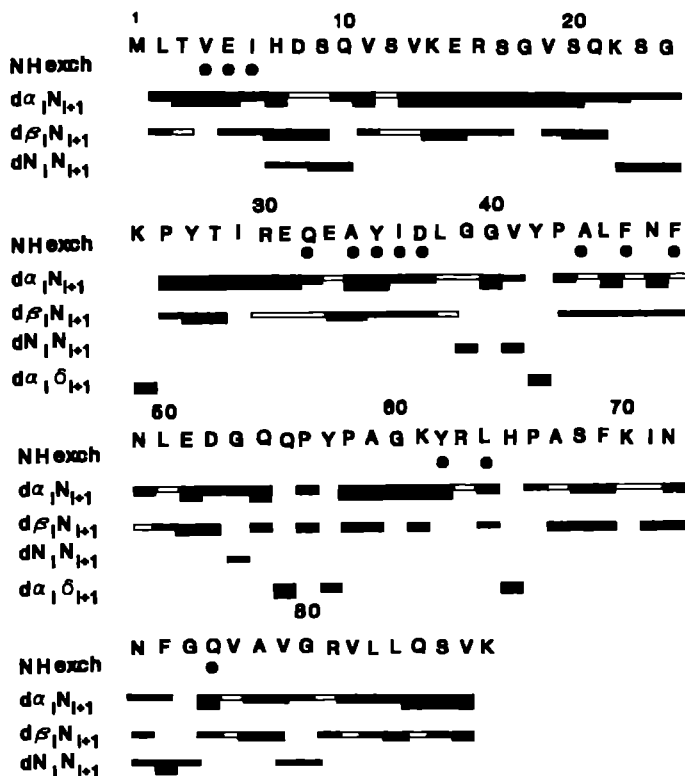
Figure 4. Sequentially assigned residues of IKe GVP together with a summary of dαN-, dβN-, dNN- and dα(δδ')-contacts. The indicated NOE intensities correspond with a mixing time of 75 ms for the exchangeable protons and with a 100 ms mixing time for the non-exchangeable protons. Open bars indicate sequential contacts whose assignment and/or intensity determination was obscured by overlap with other resonances (see text). The filled dots indicate slowly exchanging amide protons.

<u>M1</u>: 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 had been made. By elimination we have ascribed this pattern to M1. <u>Residues 2-6:</u> The N-terminal part of IKe GVP can be assigned in a straightforward manner (Figure 5) by making a dαN-walk which was, in most cases, secured by sequential dβN-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 turn-like structure as is reflected in the strong dNN-connectivities that prevail in this sequential walk.

Residues 11-15: Starting with V11 this sequential walk is strongly dominated by dαN-contacts. The assignment of V13 has not been without difficulty. Both the dαN- and the dßN-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 ß-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 dαN-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 dαN-contacts rather difficult. In addition, the intra- and interresidue dßN-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 non-sequential dαN-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 dαN- and dNN-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.

Residue 42-49: The sequential walk could be continued with a dα(δδ')-contact between Y42 and P43. A strong dαN-contact between the latter residues forms the start for a
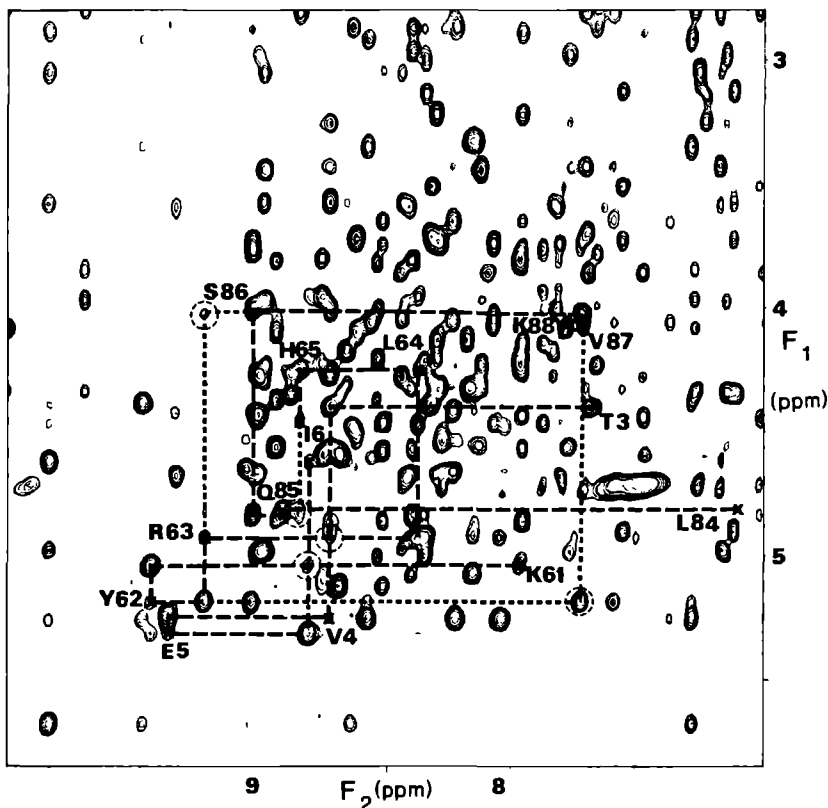
Figure 5. Detail of a 125 ms NOESY experiment showing the Nα-region. The experiment was performed on an IKe GVP sample (2.5 mM, pH 4.6, 298 K, 90% $H_2O/10\%\ D_2O$). The sequential walks for three consecutive sequences (3-6, 61-65, 84-88) are indicated by long dashes. These three strands form a triple stranded β-sheet structure which can be deduced from the observation of several non-sequential dαN-contacts. These contacts, which correspond to the skewed arrows in Figure 7 have been indicated with short dashes (as far no overlap with the long dashes would occur). In addition these contacts are marked with dashed circles.

straightforward dαN-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 dαN- and dNN-connectivities are observed, accompanied by a $d\alpha_i N_{i+2}$-contact between D52 and Q54. The sequence P56-Y57-58 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). A dαN- and a dα(δδ')-contact 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 to 70 are, in all cases, based on sequential dαN-contacts. In one case, L64, the dαN-contact was overlapping with other connectivities. Observation of dβN-, dγN- and dδN-contacts to H65 make 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 obtained at different temperatures (298 K 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 dαN-connectivities dominate the sequential walk in this amino acid sequence. An exception is formed by the segment R80-G81 in which a strong dNN-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 dαα-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 dαN- and dβN-contacts with V87.

Table I

Resonance positions of identified spin systems of IKe GVP at 298 K and pH 4.6. All chemical shifts are measured relative to TSP.

Chemical shifts (ppm)

	$N^{\alpha}H$	$C^{\alpha}H$	$C^{\beta}H$	Others
M1		3.79	2.02	$C^{\gamma}H$ 2.26, 2.48
L2	9.12	3.86	1.88^{a}	$C^{\gamma}H$ 1.23, C^{δ} 0.31 0.45
T3	7.72	4.48	4.31	$C^{\gamma}H$ 1.27
V4	8.8O	5.32	1.54	$C^{\gamma}H$ 0.78 0.91
E5	9.46	5.39	2.01	$C^{\gamma}H$ 2.30
I6	8.90	4.70	2.02	$C^{\gamma}H$ 0.95
H7	9.95	5.05	3.65, 3.13	$C^{2}H$ 8.37, $C^{4}H$ 7.24
D8	9.07	4.37	2.76	
S9	8.52	4.35	4.13, 3.93	
Q10	8.03	4.51	2.48^{a}	$C^{\gamma}H$ 2.96
V11	7.54	3.93	2.34	$C^{\gamma}H$ 1.17 0.99
S12	7.76	4.48	3.77	
V13	8.40	4.36	1.94	$C^{\gamma}H$ 0.86
K14	9.09	4.53	1.62	$C^{\gamma}H$ 1.25, 1.32, $C^{\delta}H$ 1.52, $C^{\varepsilon}H$ 2.87
E15	8.59	5.20	1.86, 1.76	$C^{\gamma}H$ 2.18, 1.99
R16	8.76	4.66	1.77, 1.63	$C^{\gamma}H$ 1.45, $C^{\delta}H$ 3.13
S17	8.59	4.95	3.83, 3.74	
G18	8.03	3.83, 3.74		
V19	8.37	4.29	1.84	$C^{\gamma}H$ 0.82 0.74
S20	8.61	4.45	4.09, 3.90	
Q21	9.01	4.19	2.17, 2.09	$C^{\gamma}H$ 2.49, 2.44
K22	8.25	4.23	1.81	$C^{\gamma}H$ 1.46, 1.42, $C^{\delta}H$ 1.68 $C^{\varepsilon}H$ 2.99
S23	7.94	4.55	4.01, 3.85	
G24	8.14	4.13, 3.83		
K25	7.88	4.76	1.73	$C^{\varepsilon}H$ 3.03
P26		5.08	2.32, 1.92	$C^{\gamma}H$ 2.10, 2.01, $C^{\delta}H$ 4.00, 3.70
Y27	8.42	4.92	3.05, 2.95	$C^{\delta}H$ 6.92, $C^{\varepsilon}H$ 6.54
T28	8.46	4.65	3.90	$C^{\gamma}H$ 1.04
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
R30	8.71	4.73	1.40, 1.51	$C^{\gamma}H$ 1.63, 1.86, $C^{\delta}H$ 3.32, 3.22, $N^{\varepsilon}H_{2}$ 7.28
E31	8.83	5.14	1.74	$C^{\gamma}H$ 2.09, 1.97
Q32	8.22	4.50		

Table I (continued)

	$N^{\alpha}H$	$C^{\alpha}H$	$C^{\beta}H$	Others
E33	8.31	4.84		
A34	8.10	4.56	1.16	
Y35	8.72	5.76	2.78, 2.95	$C^{\delta}H$ 6.91, $C^{\epsilon}H$ 6.63
I36	9.95	5.33	1.34	$C^{\gamma}H$ 0.45
D37	8.30	4.61	2.73, 2.93	
L38	8.36	4.58	1.91	$C^{\gamma}H$ 1.64, $C^{\delta}H$ 0.58, 0.78
G39	8.36	4.41, 3.80		
G40	8.70	3.80, 4.18		
V41	8.39	3.58	1.74	$C^{\gamma}H$ 0.48 0.87
Y42	7.62	4.98	3.20, 2.69	$C^{\delta}H$ 7.16, $C^{\epsilon}H$ 6.81
P43		4.73	2.77, 1.88	$C^{\delta}H$ 3.81, 3.68
A44	9.15	4.73	1.50	
L45	8.38	4.46	1.64	$C^{\gamma}H$ 1.34, $C^{\delta}H$ 0.68, 0.44
F46	9.56	5.32	3.43, 2.97	$C^{\zeta}H$ 7.34, $C^{\delta}H$ 7.02, $C^{\epsilon}H$ 6.64
N47	8.65	5.33	2.09, 2.34	$N^{\gamma}H$ 7.42
F48	8.12	5.01	3.10, 3.20	$C^{\zeta}H$ 6.82, $C^{\epsilon}H$ 6.69
N49	8.42	5.01	2.78	
L50	8.48	4.41	1.42	$C^{\zeta}H$ 1.42, $C^{\delta}H_3$ 0.80, 0.49
E51	8.81	4.47	2.13	$C^{\gamma}H$ 2.26
D52	8.97	4.34	2.63	
G53	8.81	3.65, 4.09		
Q54	7.90	4.25	2.18	$C^{\gamma}H$ 2.39
Q55	8.73	4.37	1.76	$C^{\gamma}H$ 2.15, 2.39
P56		4.16	1.87	$C^{\delta}H$ 3.65, 3.28
Y57	10.12	4.40	2.54, 2.87	$C^{\delta}H$ 7.33, $C^{\epsilon}H$ 6.86
P58		4.54	2.42, 2.09	$C^{\delta}H$ 4.04, 3.92
A59	8.45	4.03	1.34	
G60	8.84	4.31, 3.88		
K61	8.03	5.12	1.74	
Y62	9.53	5.26	2.32, 2.79	$C^{\delta}H$ 6.72, $C^{\epsilon}H$ 6.57
R63	9.32	5.01	1.78	$C^{\gamma}H$ 1.58, 1.40
L64	8.46	4.32	1.53	$C^{\gamma}H$ 1.39 $C^{\delta}H$ 0.56 0.65
H65	8.93	4.52	2.88, 2.65	C^2H 7.84, C^4H 7.07
P66		4.42	2.48^b	$C^{\gamma}H$ 1.90, $C^{\delta}H$ 3.59
A67	11.35	4.64	1.54	
S68	8.81	4.64	3.32, 3.51	
F69	7.83	5.07	2.81, 3.06	$C^{\zeta}H$ 7.20, $C^{\delta}H$ 6.99, $C^{\epsilon}H$ 6.83
K70	9.11	4.74		

Table I (continued)

	$N^{\alpha}H$	$C^{\alpha}H$	$C^{\beta}H$	Others
I71	8.30	4.74	2.04	$C^{\gamma}H$ 1.11
N72	9.43	4.75	3.67, 2.75	
N73	8.48	4.40	2.60, 2.52	
F74	7.77	4.82	2.89, 3.50	$C^{\zeta}H$ 7.39, $C^{\delta}H$ 7.32, $C^{\epsilon}H$ 7.22
G75	8.31	4.13, 3.73		
Q76	7.65	5.26	2.33, 2.04	$C^{\gamma}H$ 2.58 2.44
V77	9.10	4.50	1.63	$C^{\gamma}H$ 0.68, 0.81
A78	9.10	4.76	1.40	
V79	8.51	4.05	2.05	$C^{\gamma}H$ 0.67 0.80
G80	9.07	3.52, 4.30		
R81	8.19	4.41	1.80	$C^{\gamma}H$ 1.65, 1.53, $C^{\delta}H$ 3.18 3.14
V82	8.23	3.41	1.80	$C^{\gamma}H$ 0.20, 0.95
L83	8.21	4.42	1.39[b]	
L84	7.15	4.91	0.37	$C^{\gamma}H$ 0.37, $C^{\delta}H$ -0.72, 0.11
Q85	8.99	4.91	1.98	$C^{\gamma}H$ 2.10, 2.31
S86	9.12	4.11	3.83	
V87	7.79	4.15	1.90	$C^{\gamma}H$ 0.72, 0.83
K88	7.85	4.16	1.64	$C^{\gamma}H$ 1.36, $C^{\delta}H$ 1.78, $C^{\epsilon}H$ 2.96

a   *A second β-resonance might be present but so far the COSY data don't allow discrimination of this resonance from a γ-resonance.*

b   *NMR data obtained so far could not provide unambiguous evidence for the non-degeneracy of this resonance.*

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 dαα-contacts can be observed (Figure 6). The mere observation of these contacts is indicative of the presence of antiparallel ß-sheet structure. Scrutiny of these and other NOE data resulted in elucidation of two ß-loops which partly combine into a triple stranded ß-sheet (Structure I), one ß-loop (Structure II) and one triple stranded ß-sheet (Structure III). Figure 7 depicts the non-sequential dαN- and dNN-contacts which, together with dαα-connectivities define a regular antiparallel ß-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
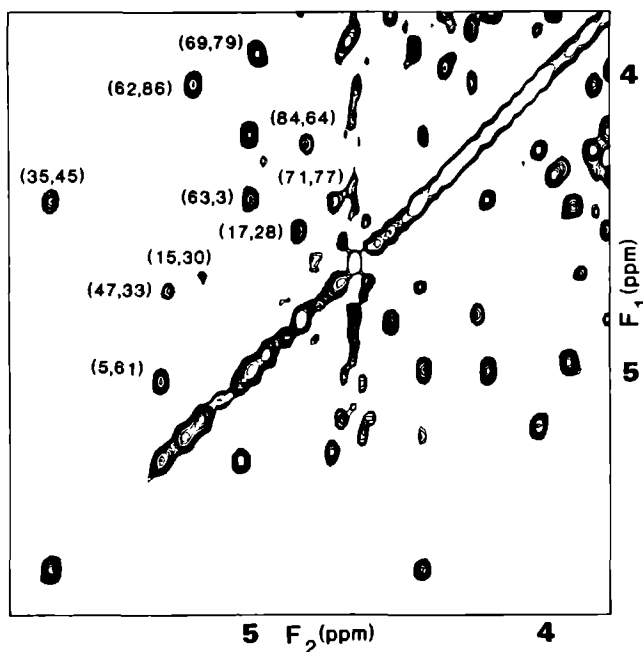
Figure 6. Detail of the Cα region of a 125 ms NOESY experiment performed on a 1.5 mM IKe GVP sample (pH 4.6, 303 K). Long range dαα-contacts which characterize the antiparallel β-sheet structures have been indicated by their (ω_2,ω_1)-coordinates.

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 non-sequential amide- to α-contacts. As is indicated in Figure 7, the ß-ladder in the ß-loop encompassing residues 31 to 49 comprises residues 37 and 43. The dαα-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 non-sequential dαN-contact between P43 and L38 that can be expected when these residues are involved in an extended ß-ladder is not clearly discernible. Clearly distinguishable, however, are side chain to side chain contacts between D37 and P43. The observation of a non-sequential dαN-contact between D37 and L44 provides additional evidence for the involve-ment of D37 and P43 in the ß-sheet structure. Structure II consists of a ß-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 residue 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 ß-sheet structures are reflected in the pattern of slowly exchanging amide protons in the amino acid sequence (Figure 4). 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 ß-sheet structure drawn in Figure 7. In the ß-loops running from residue 15 to 30 and from residue 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 ß-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 dNN-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 dαα-contacts observed in the NOESY spectra of IKe and M13 GVP have been depicted in an adjacency matrix. The stretches of dαα-contacts running perpendicular to the diagonal are highly characteristic of ß-sheet structures. Inspection of the adjacency matrix shows that the polypeptide chains of M13 and IKe GVP are almost completely identically folded in antiparallel ß-sheet structures. Note that the insertion of Q21 in the sequence of IKe GVP does not distort the ß-sheet struc-
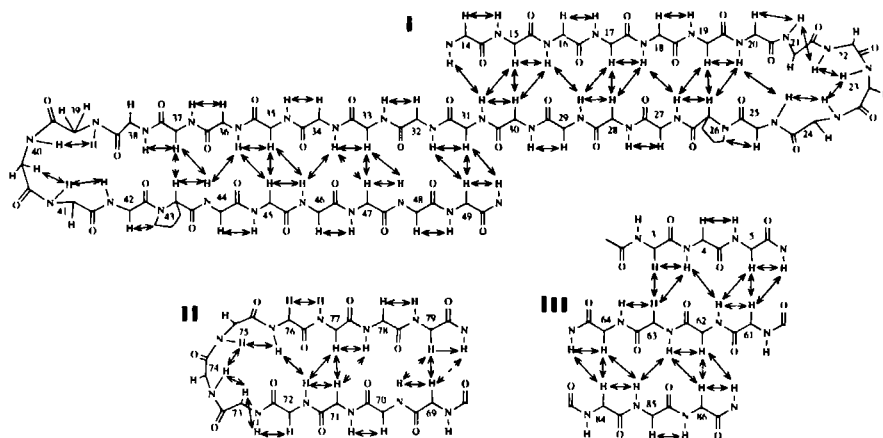
Figure 7. Schematic representation of the antiparallel β-sheet structures I-III (see text) in IKe GVP. Sequential (horizontal) and non-sequential (skewed) dαN-contacts and long range dNN- and dαα-contacts are indicated by arrows. The interrupted arrows indicate contacts whose observation is troubled by overlap.

ture in this region. When the other regions of the ß-sheet structures of these proteins are studied in more detail however, two minor differences can be observed. First, the ß-ladder in the ß-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 dαα-contact in IKe between V13 and Q32 while a strong dαα-contact exists between F13 and E31 in M13 GVP. Secondly, the ß-ladder in the ß-loop running from residues 31 to 50 comprises residues D37 and P43 as is indicated in Figure 7. In M13 GVP however, the ß-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 twofold axis of symmetry similar to that proposed for the crystal structure of M13 GVP (Brayer & McPherson, 1983). If, for the moment, we disregard the turn sections drawn in Figure 7 these considerations do not have consequences for the ß-sheet folding proposed above. However, the equivalence of the monomers in the NMR-spectrum
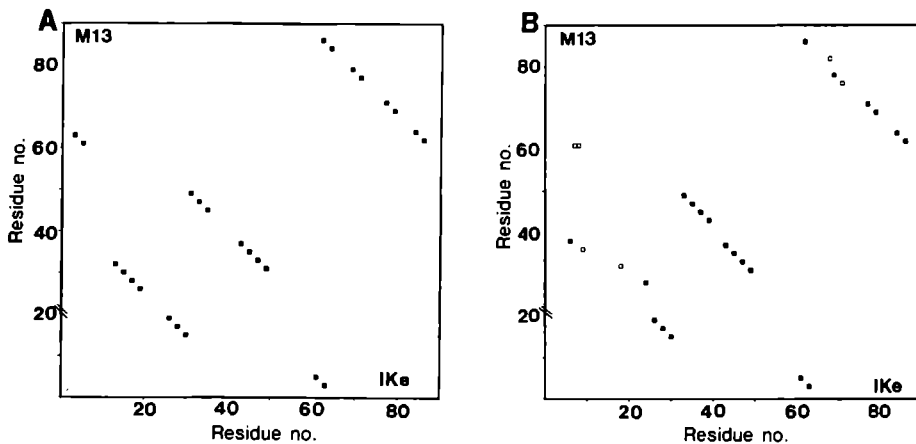
Figure 8. (A) Adjacency matrix showing, below the diagonal, the observed dαα-contacts (closed squares) in a 100 ms NOESY spectrum of IKe GVP (1.5 mM, 298 K, pH 4.6). The closed squares above the diagonal indicate the dαα-contacts that were observed in a NOESY experiment performed on M13 GVP (data taken 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 dαα-contacts in the crystal structure of M13 GVP (no intermonomer dαα-contacts (<4.0 Å) can be observed). Data were taken from the Protein Data Bank. Filled and unfilled dots correspond to distances smaller than 2.5 Å and 4.0 Å respectively. Below the diagonal dαα-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 dαα-contacts observed by NMR are also intramonomer.

has repercussions for distinguishing ß-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 intra-monomer. 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 ß-sheet. In Figure 8B the intramonomer dαα-distances (shorter than 4 Å) deduced from the crystallographic structure of M13 GVP and the dαα-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 dαα-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 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 semi-conserved 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 ß-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 ß-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 semi-conserved (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 ß-sheet structure. The T3-R63 pair in IKe GVP and the K3-T62 pair in M13 GVP are both connected by dαα-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 dα(δδ')-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
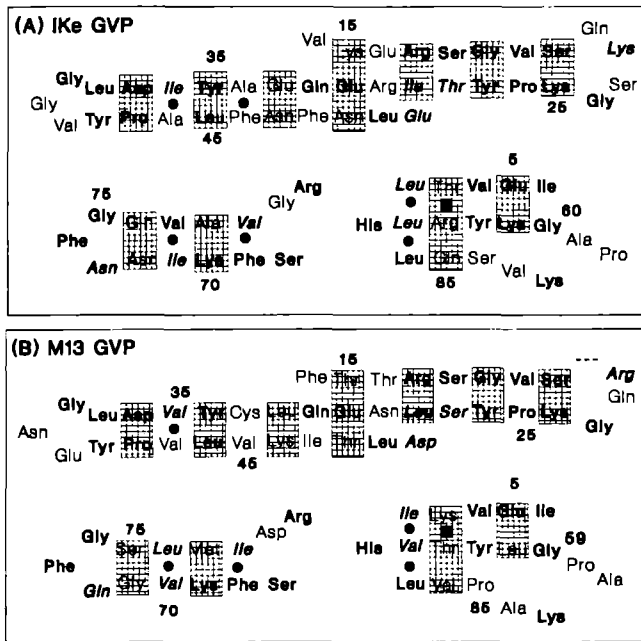
Figure 9. Comparison of the β-sheet structures derived for (A) IKe GVP and (B) M13 GVP. Residues that are conserved in IKe and M13 GVP are indicated in bold faces. Residues for which the replacement was conservative with respect to their polar, apolar or charged character are indicated in bold italic faces. The boxes enclose residues that are pointing towards 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 towards 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.

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.

In the previous Chapter we showed that the ß-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.

CHAPTER 6

Exploration of the ssDNA Binding Domains of the Gene V Proteins Encoded by the Filamentous Bacteriophages IKe and M13 by Means of a Spin-labeled Oligonucleotide and Lanthanide-DOTP Complexes[*]

ABSTRACT

Scrutiny of NOE data available for the protein encoded by gene V of the filamentous phage IKe (IKe GVP) resulted in elucidation of a β-sheet structure which is partly five-stranded. The DNA binding domain of IKe GVP was explored using a spin-labeled deoxytrinucleotide. The paramagnetic relaxation effects observed in the ^{1}H NMR spectrum of IKe GVP, upon binding of this DNA-fragment, could be visualized using 2D-difference spectroscopy. In this way, the residues present in the DNA binding domain of IKe GVP can be located in the structure of the protein. They exhibit a high degree of identity with residues in the gene V protein encoded by the distantly related phage M13 (M13 GVP) for which similar spectral perturbations are induced by such a spin-labeled oligonucleotide. Binding studies with negatively charged lanthanide-DOTP complexes showed that these bind to IKe and M13 GVP at two spatially remote sites whose affinities have different pH dependencies. Above pH 7, there exists one high affinity binding site for $Gd(DOTP)^{5-}$ per M13 GVP monomer, which coincides with the ssDNA binding domain as mapped with the aid of spin-labeled oligonucleotide fragments. The results show that ssDNA binds to conserved ("phosphate-binding") electropositive clusters at the surface of M13 and IKe GVP. These positive patches are interspersed with conserved or conservatively replaced hydrophobic residues. At pH 5, a second $Gd(DOTP)^{5-}$ binding site becomes apparent. The corresponding pattern of spectral perturbations is indicative for the accommodation of patches of conserved, or conservatively replaced, hydrophobic residues in the cores of the M13 and IKe dimers. Sequence comparison of the ssDNA binding protein encoded by the *Pseudomonas* bacteriophage Pf3 with that of M13 and IKe GVP support the hypothesis that this protein may adopt a similar structural organisation and that its interaction with ssDNA may be described by the same model.

INTRODUCTION

Gene V of the subclass I filamentous bacteriophages IKe (Peeters et al., 1983) and Ff (M13, f1 or fd) (Model & Russel, 1988) encodes for proteins (GVP) that widely have been used as model systems for non sequence-specific ssDNA-protein interactions. Both phages are conjugative plasmid specific and have *Escherichia coli* as host. Among other characteristics they differ from each other with respect to conjugative plasmid specificity (Inc N in IKe and Inc F in M13). In the replication cycle of both phages, GVP is indispensable role for sequestering of the viral strands from the dsDNA replication cycle. In addition, for the GVP encoded by M13, a gene regulatory function at the level of translation has been demonstrated (Model et al., 1982; Yen & Webster, 1982; Zaman et al., 1990, 1991).

IKe and M13 GVP consist of 88 and 87 residues respectively and in solution they predominantly occur as dimers of ~20 kDa molecular mass. Elaborate studies of their physico-chemical properties have been described and the essential features of binding to ssDNA have been characterized in thermodynamic terms. Important characteristics of the binding of IKe and M13 GVP to ssDNA are the high electrostatic contribution to the binding energy and the high binding cooperativity (de Jong et al., 1987a; Bulsink et al., 1985). This last property may be hold responsible for the formation of large aggregates upon stoichiometric addition of oligonucleotides to GVP. The increase in particle size and the concomitant decrease in ^{1}H NMR relaxation times seriously hampers NMR experiments aimed at an understanding of the ssDNA-GVP interaction at the sub-molecular level. For this reason spin-labeled DNA-fragments have come into use for studying the ssDNA-GVP interaction (de Jong et al., 1988). Even small additions of these ligands induce strong perturbations of the resonances belonging to amino acid residues residing in the DNA binding domains of IKe and M13 GVP and its surroundings. By using this tool, the involvement of a β-loop structure in ssDNA binding has been revealed in a 500 MHz study on IKe GVP (de Jong et al., 1989b). This β-loop structure, commonly denoted as the DNA binding loop, is highly conserved in M13 GVP (Chapter 4, van Duynhoven et al., 1990). For some time, further exploration of the ssDNA binding domain of IKe GVP has been obstructed by the incompleteness of the analysis of the ^{1}H NMR spectrum. The completion of the sequential assignment of the ^{1}H NMR spectrum of IKe GVP (described in the previous Chapter) opened

the prospect of mapping the complete ssDNA binding domain of IKe GVP which is described in the present paper. It will be demonstrated that this domain is organized similarly to that of M13 GVP.

Modelling studies, based on the crystal structure available for M13 GVP, suggested the involvement of specific electropositive clusters on the protein surface in the interaction with the phosphate backbone of ssDNA (Brayer & McPherson, 1984; King & Coleman, 1987). In Chapter 4 and Folkers et al. (1991b), however, several differences between the structure in solution (as studied by NMR) and the crystal structure have become apparent. The first data with respect to an involvement of positively charged residues in ssDNA binding have been derived from ^{1}H NMR studies by Alma et al. (1981b). Studies involving ^{13}C-methylation of lysine residues of M13 GVP indicated the involvement of three of these in ssDNA binding (Dick et al., 1988; 1989). The modified lysyl residues were also used in studies with paramagnetic Ln(DOTP)$^{5-}$ complexes. DOTP, a macrocyclic amine with phosphonomethyl groups attached to the nitrogen atoms, forms rigid and axially symmetric complexes with trivalent lanthanides (Sherry et al., 1988). The binding of these negatively charged complexes to GVP, which is competitive with the binding of the oligonucleotide dA$_7$ (Dick et al., 1989), resulted in significant paramagnetic broadening of those lysine residues that are involved in ssDNA binding. In the current work, the lanthanide chelates were further employed as probes for solvent accessible clusters of electropositive residues. These ligands proved useful for mapping those patches on the protein surfaces of both M13 and IKe GVP that are prone to undergo electrostatic interactions with the phosphate backbone of ssDNA.

MATERIALS AND METHODS
MATERIALS

Protein isolation and preparation. The procedure for isolation and purification of IKe GVP was essentially as described in the previous Chapter. Wild-type M13 GVP and mutant M13 GVP Y41H were isolated and prepared according to Chapter 3 and 4. Purified GVP was lyophilized and stored at -20 °C. The final concentration of GVP in the NMR samples varied between 0.5-1.5 mM, with no buffer components added and only a few millimolar salt present. The pH of the samples was adjusted using aliquots of diluted DCl. For concentration

determinations the following absorbtion coefficients were used: 7100 $M^{-1}cm^{-1}$ for IKe and M13 GVP and 5680 $M^{-1}cm^{-1}$ for M13 GVP Y41H (Chapter 3).

Spin-labeled oligonucleotide. A deoxytrinucleotide ($*(dA)_3*$) was synthesized with two spin label TEMPO groups covalently attached (through phosphodiester bonds) to the 5'- and 3'-ends (Claesen et al., 1986). The concentration of this oligonucleotide was determined by measuring the optical extinction at pH 7, using an absorbtion coefficient of 37,500 $M^{-1}cm^{-1}$ (de Jong et al., 1987b).

Ln(DOTP)$^{5-}$ complexes. The DOTP (1,4,7,10-tetraazacyclodecanetrayl-1,4,7,10-tetrakis (methylene)tetrakisphosphonic acid) reagent was synthesized by a procedure (Swinkels et al., 1991) which is an improvement of methods scattered in the literature. (Dr. A.D. Sherry is acknowledged for providing us with a batch of DOTP that was used for control experiments.) The Ln(DOTP)$^{5-}$ complexes (Ln = La, Gd or Tb) were prepared by adding aliquots of a LnCl$_3$ solution to a diluted DOTP solution (1-5 mM) at pH 8. During this titration, the mixture was vigorously stirred and kept at 80 °C until the lanthanide and DOTP were present in 1:0.95 ratio. The pH of the Ln(DOTP)$^{5-}$ stock solution was adjusted using aliquots of diluted DCl or NaOD. To prevent precipitation of Ln(DOTP)$^{5-}$ aggregates (Geraldes et al., 1989) these stock solutions were kept at concentrations of 2-5 mM.

METHODS

NMR measurements. NMR measurements were performed at 600 and 400 MHz on Bruker AM600 and AM400 spectrometers, respectively. Clean-TOCSY (MLEV17) (Bax & Davis, 1985; Griesinger et al., 1988) and NOESY (Jeener et al., 1979) experiments were conducted. All D$_2$O-TOCSY spectra were recorded with mixing times of 30 ms. 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).

H$_2$O-NOESY spectra were obtained with mixing times of 125 ms. The solvent resonance was suppressed using a semi-selective jump-return pulse sequence (Plateau & Gueron, 1982). Prior to the recording of the 2D-experiments, the receiver phase and a pre-

acquisition delay were adjusted so as to minimize baseline artifacts (Marion & Bax, 1988). In all experiments, the carrier was placed at the H_2O-resonance and TPPI was used for signal accumulation in the t_1 dimension (Marion & Wüthrich, 1983). Data were Fourier transformed in the phase-sensitive mode after weighting with shifted square sine bells or shifted gaussian functions. The data were processed either by standard Bruker software or by the NMRi package (installed on a Bruker ASPECT 3000 and a SUN 4 computer respectively). If necessary, additional baseline corrections were performed by home-written algorithms (programmed in C and installed on a SUN 4 computer).

Paramagnetic perturbations of ^1H NMR spectra. The methods adopted to investigate the binding of spin-labeled oligonucleotides and $Ln(DOTP)^{5-}$ complexes involve their paramagnetic effects on the ^1H NMR spectrum of GVP. The perturbations of interest in the present study are the lanthanide-induced shifts and the increased relaxation rates of the spins. The decrease in T_1 and T_2 relaxation times induced by the paramagnetic probes $Gd(DOTP)^{5-}$ and $*(dA)_3*$ can be described by (Dwek, 1973):

(1)
$$\frac{1}{T_{1,2\ pe}} = \frac{f_{1,2}(\tau_c,\omega)S(S+1)}{r^6}$$

In this equation r denotes the distance between the perturbed proton spin and the paramagnetic centre either in $Gd(DOTP)^{5-}$ or in $*(dA)_3*$ (the Gd3+ ion and the nitroxide in the TEMPO moiety respectively); $f_{1,2}(\tau_c,\omega)$ describes the different dependencies of T_1 and T_2 on the rotation correlation time, τ_c, and Larmor frequency, ω. S is the spin angular momentum of the paramagnetic centres; for equal values of the linebroadening and $f(\tau_c,\omega)$ the distances between the paramagnetic centres (Gd^{3+}: S=7/2; TEMPO: S = 1/2) and the perturbed proton spin will differ by a factor of 1.66.

The rotation correlation time of the system under study ($\tau_c \approx 15$ ns; Bulsink et al., 1986) is in the range where the paramagnetic effect on T_2 (and $T_{1\rho}$) is considerably stronger than on T_1 (de Jong et al., 1988). Therefore during the mixing time in TOCSY experiments, where relaxation is dominated by $T_{1\rho}$ (Bax & Davis, 1985), the intensity of the resonances is more effectively reduced than during a NOESY mixing time (where relaxation is dominated by T_1).

The definition of the association rate k^*_1 (which takes into account the concentration of free DOTP complex D) allows us to describe the exchange between the bound (GVPD) and unbound (GVP) state of the protein by the two-site model:

(2)
$$GVP \underset{k_{-1}}{\overset{k^*_1 = k_1 * D}{\rightleftharpoons}} GVPD$$

If only a small fraction of the protein is in the bound state then the decrease of T_2 is given by (Berliner, 1976; Dwek, 1973):

(3)
$$\Delta\left(\frac{1}{T_2}\right) = \frac{k^*_1}{k^*_1 + k_{-1}} \frac{1}{\frac{1}{k_{-1}} + T_{2,p-e}}$$

It follows from this equation that the observed increase in the relaxation rates will become limited by the rate of exchange when the reciprocal dissociation rate constant $(k_{-1})^{-1}$ is in the order of magnitude (or larger) than $T_{2,p-e}$. Consequently, below a certain distance between the paramagnetic centre and the perturbed proton spin, all perturbations will become of the same magnitude because then $T_{2,p-e}$ will dominate. Therefore relative and absolute distances can only be determined above a certain threshold value which is dictated by the exchange rate of the system.

The paramagnetic shifts induced by the $Ln(DOTP)^{5-}$ complexes (lanthanide induced shifts: LIS) are proportional to $(3\cos^2\theta - 1)/r^3$ where r again is the distance between lanthanide ion and perturbed proton and θ is the angle between the principal symmetry axis of the paramagnetic susceptibility tensor of the lanthanide complex and the ion-proton vector. The diamagnetic $La(DOTP)^{5-}$ complex serves as a control for the experiments with the paramagnetic $Ln(DOTP)^{5-}$ complexes.

2D-difference spectra. Absolute 2D-difference spectra ($Diff_{abs}$) were obtained by subtraction of TOCSY or NOESY spectra recorded before and after addition of spin-labeled oligonucleotide or $Gd(DOTP)^{5-}$:

(4) $Diff_{abs} = I(absence) - C*I(presence)$

I(absence) and I(presence) denote peak intensities in the 2D-spectra recorded before and after addition of paramagnetic ligand. The factor C corrects for systematic deviations in signal intensity in these spectra due to drift of the spectrometer, sample dilution etc. Typically, values of 1.2-1.4 are to be used for this constant. In addition, relative difference spectra ($Diff_{rel}$) were calculated using the relation:

(5) $$Diff_{rel} = \frac{I(absence) - C*I(presence)}{I(absence)}$$

In this calculation a minimum level was defined to discriminate between genuine and noise signals and the factor C is slightly overcorrected to eliminate artifacts due to random variation of signal intensity. A more elaborate discussion of these methods has been given by de Jong et al., 1988.

Using the relative 2D-difference representation the relative perturbations were calculated for all connectivities (with an accuracy of ~10%). Then, for every residue an average value for the perturbation of its connectivity pattern was calculated. These values were grouped into three classes: strong (>75%), medium (45-75%) and weak (<45%). The calculations were performed, using home-written algorithms programmed in Pascal and C, on Bruker ASPECT 3000 and SUN 4 computers respectively.

Quantitative analysis of ^{1}H NMR titration curves. Titrations of GVP with the shift reagent Tb(DOTP)$^{5-}$, in which the LIS, $\Delta\delta$, of a ^{1}H NMR resonance was monitored, were aimed at the determination of the stoichiometry, n, of the GVP-Tb(DOTP)$^{5-}$ complex and of its dissociation rate constant k_{-1}. Due to the high affinity of Tb(DOTP)$^{5-}$ for GVP and the large LIS values, the course of these titration curves are often partially ruled by intermediate exchange conditions. When we consider the equilibrium between complexed and non-complexed protein as a two-site exchange model (vide supra), we can use the analytical expressions that describe the spectral parameters for such a system (Kühne et al., 1979) to simulate these titration curves (Feeney et al., 1979). In this procedure, the association rate

constant k_1 is kept fixed at values estimated for diffusion limited complexation (10^7-10^9 $M^{-1}s^{-1}$). The maximal LIS, $\Delta\delta_{max}$, was determined graphically. The dissociation rate constant k_{-1} and the stoichiometry, n, were determined by a 2-dimensional amoeba (simplex) (Press et al., 1989) that minimized the RMS value of shifts, $\Delta\delta$, between simulated and theoretical titration curves (using an algorithm programmed in Pascal). In the midpoint of the titration, the experimental data were weighted to a lesser extent because of linebroadening and departures from Lorentz lineshapes (Feeney et al., 1979).

RESULTS

Elucidation of the antiparallel alignment of the N-terminal segment and the complex loop. The work described in the previous Chapter established the occurrence of several secondary structural elements in IKe GVP (van Duynhoven et al., 1992), i.e. a triple-stranded β-sheet and three β-loops. In analogy to the notation for the M13 GVP crystal structure (Brayer & McPherson, 1983), the β-loops were termed the dyad loop, the DNA binding loop and the complex loop (see Figure 1). Resumption of the analysis of the NOE data available for IKe GVP (previous Chapter) revealed several backbone to backbone, backbone to side chain and side chain to side chain NOE contacts between the N-terminal segment and the complex loop (arrows in Figure 1). In addition, the amide protons of residue 5, 37 and 35 were found to be slowly exchanging with the solvent. These observations indicate that, in addition to the secondary structure described earlier, the N-terminal segment and the complex loop are arranged in an antiparallel alignment which is stabilized by hydrogen bonding involving the mentioned amide protons (see Figure 1). In M13 GVP the N-terminal segment is similarly aligned with respect to the complex loop; evidence for this structure is presented in Folkers et al., 1992. The experiments described in the following sections are aimed at probing the electropositive surfaces on M13 and IKe GVP and, more specifically, mapping of the ssDNA binding domain in IKe GVP starting from the complete secondary structure of the proteins that has now become available.

Mapping of the DNA binding domain of IKe GVP using the spin-labeled oligonucleotide *(dA)$_3$*. Figure 2 shows a part of the fingerprint (dαN) region of the NOESY spectrum of

```
                                    k  88
          83  I L │ Q │ S  v
          65  h L │ R │ Y │ K  g  80    10
          1   m L │ T │ V │ E │ I  h n e  q v        DNA binding loop
                    N     α  (N)  α              s            20
                    H     H   H   H                v
                    ↕  /  ↕   ↕  /  ↕           K │ E │ R │ S │ G │ V │ S │ Q
                   (H)   α  (H)   α             R │ I │ T │ Y │ P │ K │ G    K
                    N         N                                             S
              G L │ D │ I │ Y │ A │ E │ Q │ E R              G
          40 G                                  I
              V Y │ P │ A │ L │ F │ N │ F │ N  L  e d g q q P y p a   59
                Complex loop                       50
```

Figure 1. Schematic representation of the secondary structure of IKe GVP. The β-loops have been designated according to the nomenclature introduced by (Brayer & McPherson, 1983). Newly recognized NOE contacts between amide (N) and α-protons are indicated by arrows. Lines indicate connectivities whose observation is troubled by overlap. Also the dαα-contacts are partly masked by overlap but multiple sidechain to sidechain connectivities indicate close proximity of residues V4-I36 and I6-A34. The circles around the amide protons indicate slow exchange of these nuclei with the solvent. Numbers indicate residue positions in the amino acid sequence. Coding of residues in capitals indicates that these residues are involved in the secondary structure elements recognized sofar (β-sheets and turns); coding in lower case indicates that, at the current state of the structure elucidation, such an involvement cannot be established. The typical pattern of amino acid side chains pointing alternatingly up- and downward in antiparallel β-sheets has been indicated by the boxes which enclose side chains pointing towards the reader.

IKe GVP dissolved in H_2O. Also shown is a difference spectrum, resulting from the subtraction of the aforementioned spectrum and a spectrum recorded after addition of an aliquot of the spin-labeled oligonucleotide *(dA)₃* to the IKe GVP sample. This figure serves to illustrate the potential of the 2D-difference experiment for simplifying crowded regions of 2D-spectra. The perturbed dαN- and dβN-connectivities that show up in this 2D-difference NOESY experiment allow for a straightforward confirmation of the sequential assignments of residues in close proximity to the bound *(dA)₃*. Several linebroadening effects of spin systems that could not be identified in an earlier 500 MHz study (de Jong et al., 1989b) can now, by virtue of the nearly complete assignment of the ^{1}H NMR spectrum, be interpreted in a sequence specific manner. The increased sensitivity of the present 600 MHz 2D-experiments allows for the recognition of even more perturbed connectivity patterns.

Though the difference NOESY experiments are useful during the process of assignment and structure elucidation, the more drastic signal filtering achieved during the
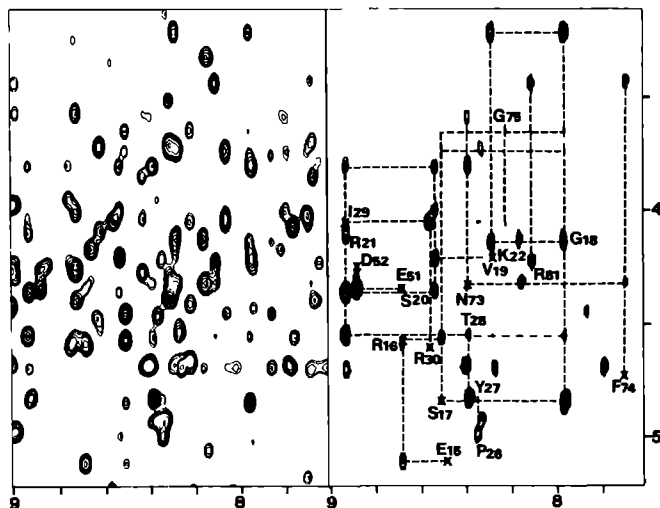
*Figure 2. (A) Fingerprint region of a 600 MHz NOESY spectrum of 1 5 mM IKe GVP (mixing time 125 ms) recorded at 298 K in H_2O (B) Absolute difference plot calculated by subtraction of the spectrum in (A) and a NOESY spectrum recorded after addition of 0 04 molar equivalent of *(dA)3* Sequential walks connecting the fingerprint connectivities of residues in close proximity of *(dA)3* are indicated by dashed lines.*

mixing time of TOCSY experiments (which is dominated by $T_{1\rho}$, vide supra) makes this experiment more useful for quantification (see METHODS) of the paramagnetic perturbations In Figure 3A, relaxation effects as induced by *(dA)3* in a 30 ms clean-TOCSY spectrum of IKe GVP have been quantified within the framework of the secondary structure of IKe GVP The typical character of antiparallel β-sheet structures where the amino acid side chains alternatingly point up- and downwards (indicated by boxes in Figure 3) is reflected in the relaxation induced perturbation pattern. In general, the spin systems of the boxed residues (which point towards the reader) are perturbed more strongly than those of their neighbours. The strong perturbations of the boxed residues in the sequence running form R16 to I29, the so-called DNA binding loop have been described in detail previously (de Jong et al., 1989b). The present data indicate the close proximity of several other residues to the spin-labeled oligonucleotide, i e E31, N47, N49, E51, K70, F74, Q76, A78, G80 and R81 Furthermore, it is noted that a major part of these newly recognized residues are located in or near the β-loop which is commonly referred to as the dyad loop
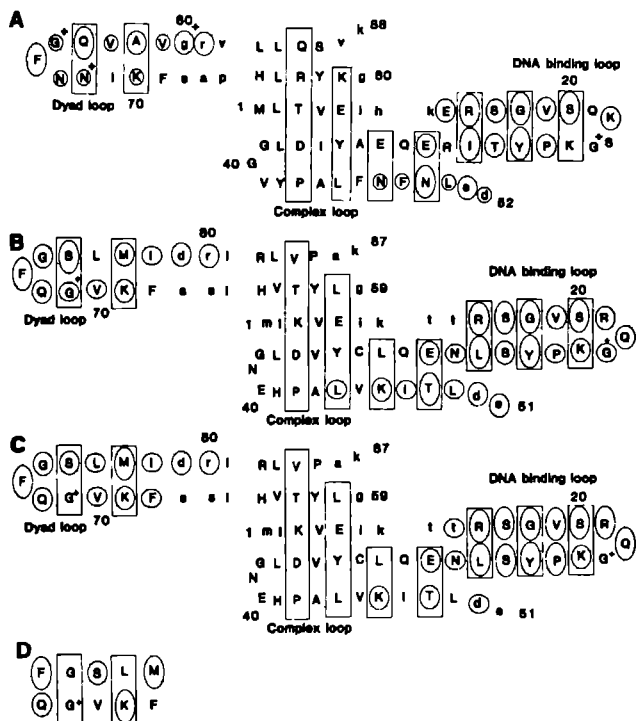
Figure 3. Presentation of the quantified ^1H NMR perturbations induced upon addition of (A) 0.04 equivalent of *(dA)$_3$* to IKe GVP (pH 4.8), (B) 0.04 equivalent *(dA)$_3$* to M13 GVP Y41H (pH 7, data taken from (Folkers et al., 1992), (C) 0.02 equivalent of Gd(DOTP)$^{5-}$ to M13 GVP Y41H (pH 7). In (D) the secondary structure of the segment running from 68 to 78 in M13 GVP as elucidated with X-ray diffraction studies (Brayer & McPherson) is presented. The strongest perturbations by Gd(DOTP)$^{5-}$ as they are present in C have been indicated. Relative and absolute difference spectra were calculated from clean-TOCSY (MLEV17, mixing time 30 ms) data, recorded before and after addition of the paramagnetic agent (*(dA)$_3$* or Gd(DOTP)$^{5-}$) to 1.2-1.5 mM GVP samples. For every residue, an average spin system perturbation was calculated (see METHODS). These values have been grouped in three classes which correspond to the different ellipses enclosing the residues (weak: o ; medium: O ; strong: O ; see METHODS). The residues marked with "+" manifest themselves in TOCSY spectra by poorly resolved connectivity patterns (due to antiphase lineshapes (glycines) and/or overlap) which hampers accurate quantification of the paramagnetic effects. In a few cases, however, scrutiny of 2D-difference NOESY experiments still enabled determination of their relative perturbations. Residues have been coded and emboxed as in Figure 1.
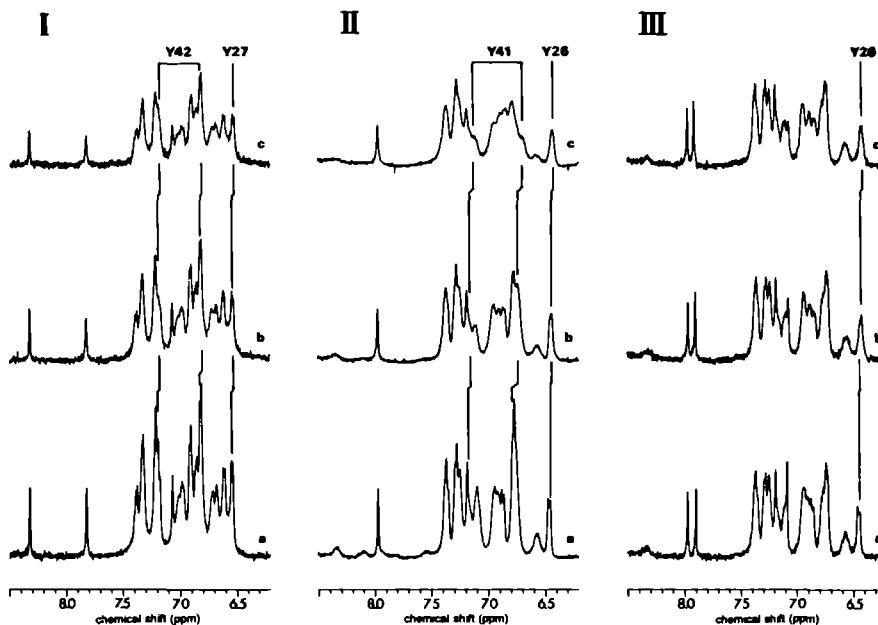
Figure 4. The aromatic regions of the 400 MHz ¹H NMR spectra of (I) 0.9 mM IKe GVP (pH 4.8), (II) 0.8 mM M13 GVP (pH 6.7) and (III) 1.0 mM M13 GVP Y41H (pH 7.1) before (a) and after addition of (b) 1/6 and (c) 1/3 equivalent of La(DOTP)⁵⁻. The shifts of the resonances of Y42 in IKe GVP and Y41 in M13 GVP (monitors for the aggregation state of these proteins) have been indicated. Also small upfield shifts can be observed for ε-Y26 in M13 and M13 GVP Y41H and ε-Y27 in IKe GVP.

Characterisation of the binding of Ln(DOTP)⁵⁻ complexes to GVP.

Diamagnetic effects: As is shown in Figure 4, an overall linebroadening and some minor shifts can be observed in the aromatic part of the ¹H NMR spectra of IKe and M13 GVP upon addition of the diamagnetic La(DOTP)⁵⁻ complex. These effects are analogous to those observed upon addition of other counter ions (e.g. NaCl) and can be attributed to an increased aggregation of the gene V protein which is most likely due to the shielding of its charged residues (Chapter 3). The upfield shift of the ring proton resonances of Y41 in M13 and Y42 in IKe GVP arise from protein-protein interactions in the GVP aggregates and are diagnostic for the aggregation state of M13 and IKe GVP (de Jong et al 1987b; King & Coleman, 1988; Chapter 3). The mutant protein M13 GVP Y41H, in which Y41 of wild-type M13 GVP is subsituted for a histidyl residue, was found to have aggregation properties far more favourable for ¹H NMR studies, probably because of a disturbance in the aforementioned

protein-protein interactions (Chapter 3 and 4). Indeed, titration of the La(DOTP)$^{5-}$ complex to this mutant GVP had a less dramatic effect on the linewidths. Still small shifts in the ^{1}H NMR spectrum of M13 GVP Y41H can be observed, analogous to those induced by raising the ionic strength of the solution. Therefore we attribute these to the binding of La(DOTP)$^{5-}$ to GVP as a counter ion. We conclude that complex formation of La(DOTP)$^{5-}$ with wild-type M13 and IKe GVP and mutant M13 GVP Y41H does not lead to (drastic) changes in the tertiary structures of these proteins which renders these ligands useful as probes for (electropositive) surfaces. The gradual increase in the magnitudes of the diamagnetic shifts indicates that fast exchange conditions apply to the binding of La(DOTP)$^{5-}$ to GVP. The contribution of exchange to the observed linebroadening can therefore be excluded. In the next section, it will be shown that for the paramagnetic shifts induced by Tb(DOTP)$^{5-}$, which can be orders of magnitude larger, under some conditions fast exchange conditions no longer apply.

Lanthanide-induced Shifts: Even upon addition of very small amounts of Tb(DOTP)$^{5-}$, paramagnetic shifts can be observed in the ^{1}H NMR spectra of IKe and wild-type M13 GVP and mutant M13 GVP Y41H. These shifts can be reversed by addition of salt to the solution; this is another indication that electrostatic interactions make a major contribution to the binding of Ln(DOTP)$^{5-}$ to GVP. Tracking of most of the shifted resonances during a titration with Tb(DOTP)$^{5-}$ beyond the saturation point, a prerequisite for a quantitative analysis of titration curves, is hindered for several reasons. First, the crowdedness of the ^{1}H NMR spectrum impairs observation of the separate resonance lines; except for some isolated resonances (e.g. ε-Y26 and δ-L83 in M13 GVP). Thus, the course of most lanthanide induced shifts during a titration is masked by overlap. Secondly, due to increased aggregation (vide supra), exchange and paramagnetic relaxation effects, the linebroadening is considerable. As pointed out in the previous section, linebroadening resulting from aggregation is significantly less in titrations of Ln(DOTP)$^{5-}$ with the mutant M13 GVP Y41H. Comparison of the paramagnetic shifts observed in the ^{1}H NMR spectra of wild-type M13 GVP and mutant M13 GVP Y41H upon addition of small amounts (0.020 equivalent) of Tb(DOTP)$^{5-}$, showed no significant differences. Therefore, M13 GVP Y41H was used for further characterization of the complexation of GVP with Ln(DOTP)$^{5-}$.

Figure 5. (A) The ^1H NMR LIS values (at 400 MHz) of the δ-L83 resonance of M13 GVP Y41H (0.5 mM) as observed during titrations with $Tb(DOTP)^{5-}$ at pH 8.8 (10 mM Tris-HCl) (□), pH 7.0 (10 mM Tris-HCl) (*) and pH 4.9 (+). The drawn line indicates a simulation of the titration at pH 8.8 with parameters that gave an optimal fit to the experimental data i.e. binding stoichiometry 1.07, dissociation rate $k_{-1}=1.2x10^2$ $M^{-1}s^{-1}$ and association rate constant $k_1=10^8$ $M^{-1}s^{-1}$.
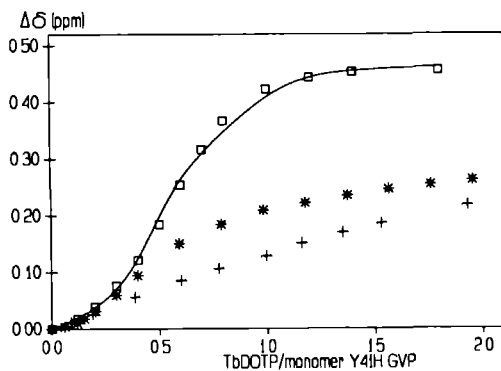
(B) The LIS values of ε-Y26 (0.5 mM) induced by 0.02 equivalent of $Tb(DOTP)^{5-}$ as function of pH (4.5-8.5). The pH was adjusted by addition of aliquots of diluted DCl to the solution.
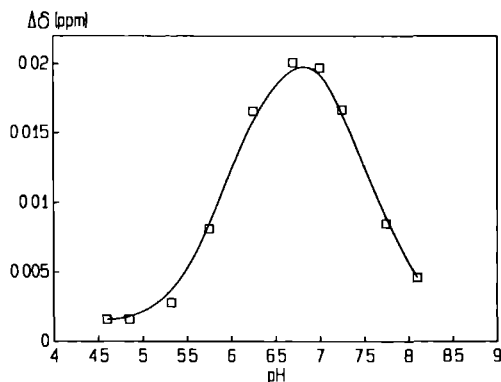
(C) The pH dependent k_{-1} rates (□, logarithmic left axis) and the fraction bound GVP (+, right axis) that match the LIS values presented in B, $\Delta\delta_{max}=1.1$ ppm, a stoichiometry of n=1, and an association rate constant of 10^8 $M^{-1}s^{-1}$. The ■ indicates the k_{-1} dissociation rate that was determined from the titration curve at pH 8.8 (10 mM Tris-HCl) in (A)
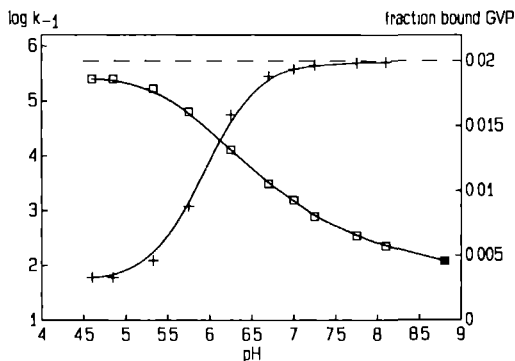
Figure 5A shows the course of the δ-resonance of L83, in M13 GVP Y41H, during titrations with $Tb(DOTP)^{5-}$ at three different pH values. A qualitative interpretation of the (sigmoidal) appearance of the titration curves indicates that above pH 7 the affinity of $Tb(DOTP)^{5-}$ for GVP is in the regime where at the start of these titrations intermediate exchange conditions apply which during the titration move over to fast exchange conditions (Feeney et al., 1979). In contrast, the curve recorded at pH 4.9 indicates that the affinity of the $GVP-Tb(DOTP)^{5-}$ has dropped to a value where fast exchange conditions determine the entire range of the titration. It is noted that the curves recorded at different pH values do not reach the *same* maximal LIS value ($\Delta\delta_{max}$). The values for $\Delta\delta_{max}$ are not only dependent on pH (Figure 5A) but also vary with the total protein concentration and salt strength (data not shown). It is our experience that conditions that stimulate aggregation of GVP tend to decrease the maximal shift, $\Delta\delta_{max}$, for δ-L83. At pH 8.8 the lanthanide induced shifts reach a maximal value for $\Delta\delta_{max}$ which indicates that self-association plays a minor role during the titration and that a two-site exchange model can be assumed. Using the iterative optimisation procedure, outlined in the METHODS section, simulated titration curves were fitted to the experimental data obtained at pH 8.8 (Figure 5A). Best agreement corresponded to a dissociation rate of the complex of $k_{-1}= 1.2\times10^2$ s^{-1} and a binding stoichiometry of n=1.07 ($Tb(DOTP)^{5-}$/GVP monomer). In this simulation an association rate constant of 10^8 $M^{-1}s^{-1}$ was used which is an estimate of the diffusion limit for the $GVP-Tb(DOTP)^{5-}$ association. Variation of this parameter in the range 10^7-10^9 $M^{-1}s^{-1}$ did neither significantly alter the appearance of the simulated curve nor significantly change the values for the dissociation rate and the binding stoichiometry. Due to the sensitivity of the LIS of the δ-L83 resonance to aggregation, the curves recorded at lower pH are not suitable for a quantitative interpretation as the assumption of a two-site exchange model no longer holds for this resonance. The lanthanide shift of the ε-Y26 resonance, however, is less sensitive to self-association of GVP and is therefore a better monitor of the pH dependence of the stability of the GVP-$Tb(DOTP)^{5-}$ complex. Due to exchange broadening, this resonance cannot be traced during the full course of a titration. Figure 5B shows the pH dependence of the LIS value of the ε-Y26 resonance induced by a small amount (0.02 equivalent) of $Tb(DOTP)^{5-}$. Below pH 7 the decrease in the LIS values can be attributed to dissociation of the GVP-$Tb(DOTP)^5$ complex in the fast exchange limit. Above pH 7 the decrease in the LIS values must be attributed to

an increase in affinity of Tb(DOTP)$^{5-}$: the dissociation rate becomes sufficiently slow to bring the system in the intermediate exchange regime. From the LIS at pH 6.8, where the system is in fast exchange and nearly all of the Tb(DOTP)$^{5-}$ is bound, we can estimate the LIS for ε-Y26 in the complexed GVP: $\Delta\delta_{max}$≈1.0 ppm. If we assume an association rate constant of $k_1=10^8$ M^{-1}s^{-1} and a binding stoichiometry of n=1 (Tb(DOTP)$^{5-}$/GVP monomer) we can estimate (as a function of pH) the fractions bound Tb(DOTP)$^{5-}$ and the dissociation rates k_{-1} by matching simulated LIS values with experimental ones; the result is shown in Figure 5C. The course of the fraction bound Tb(DOTP)$^{5-}$ and the dissociation rate of its complex with GVP match the qualitative description given previously. It is noted that the dissociation rate that was determined from the titration curve of the δ-L83 resonance obtained at pH 8.8 fits nicely into this picture. In the next section, it will be shown that the parameters estimated for the pH dependency of the stability of the GVP-Tb(DOTP)$^{5-}$ complexes are relevant towards the correct interpretation of linebroadening effects caused upon binding of Gd(DOTP)$^{5-}$.
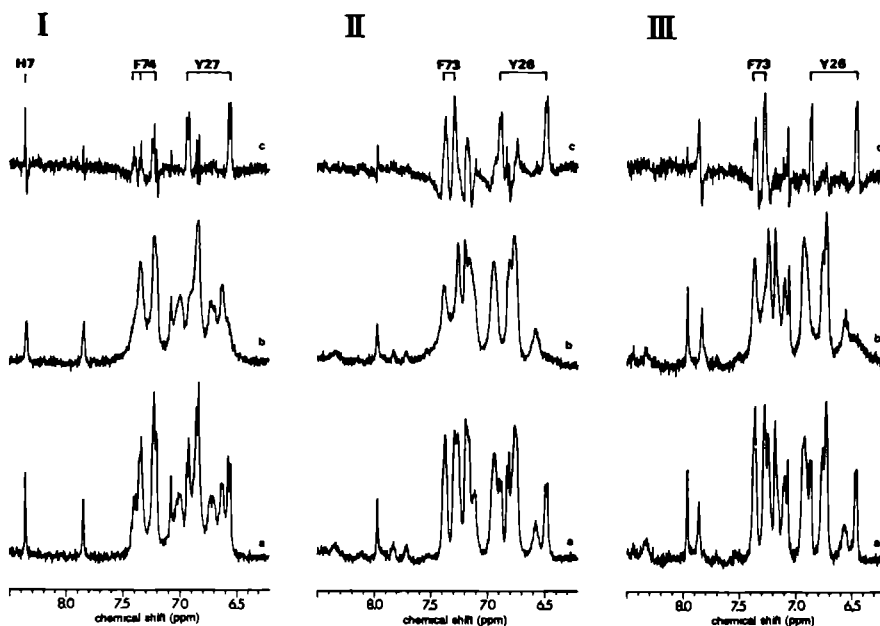

Figure 6. The aromatic regions of the 400 MHz ^{1}H NMR spectra of (I) 0.5 mM IKe GVP (pH 4.7), (II) 0.5 mM M13 GVP (pH 7.2) and (III) 0.6 mM M13 GVP Y41H (pH 7.1) (C) before (a) and after (b) addition of 0.02 equivalent of Gd(DOTP)$^{5-}$. In the difference spectra (c) the perturbed spin systems and their sequential assignments have been indicated.
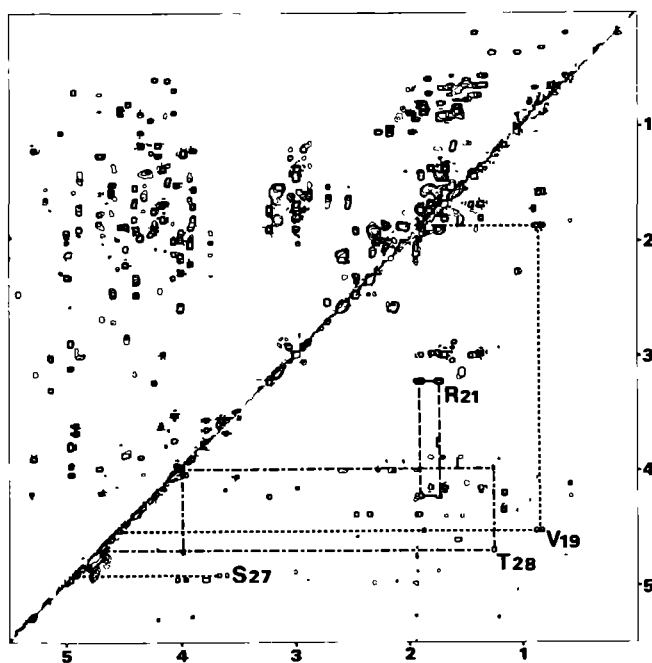
Figure 7. *Comparison of part of the clean-TOCSY spectrum and Gd(DOTP)$^{5-}$ induced difference TOCSY spectrum recorded for M13 GVP Y41H. Above the diagonal the 600 MHz clean-TOCSY(MLEV17) spectrum is shown recorded for a sample of 1.2 mM M13 GVP Y41H dissolved in D$_2$O at pH 7. The experiment was performed at 298 K with a mixing time of 30 ms. The difference spectrum is represented below the diagonal; it was computed from the spectrum presented above the diagonal and a spectrum recorded after addition of 0.02 equivalent Gd(DOTP)$^{5-}$. Several spinsystems and their sequential assignments are indicated.*

Mapping of the binding sites of Gd(DOTP)$^{5-}$ on M13 and IKe GVP. The pH range in which useful difference experiments can be performed is governed by the lifetime of the GVP-Gd(DOTP)$^{5-}$complex $(k_{-1})^{-1}$ (see METHODS). All difference experiments were therefore performed at pH 7 or lower, because estimations of k_{-1} (vide supra) indicate that above pH 7 all proton spins in the GVP monomer will be perturbed to the same extent. The line-broadening caused by the binding of Gd(DOTP)$^{5-}$ to GVP is visualized in the 1D-difference spectrum in Figure 6. Control experiments with La(DOTP)$^{5-}$ indicated that broadening and shifts due to diamagnetic effects were negligible under the experimental conditions studied. The limited resolution of the 1D-difference experiment only allows for the assignment of a few affected resonances. It is clear, however, that the mutation of Y41 into

a histidyl residue in M13 GVP does not have a significant effect on the paramagnetic linebroadening effects by Gd(DOTP)$^{5-}$. Apart from the negligible affinity of Gd(DOTP)$^{5-}$ for this residue it also indicates that the tertiary structures and of wild-type M13 and mutant M13 GVP Y41H are alike. A more powerful tool for visualisation and quantification of linebroadening effects is provided by 2D-difference spectroscopy. Figure 7 nicely illustrates the specific bleaching effects induced by Gd(DOTP)$^{5-}$ in a clean-TOCSY spectrum recorded for M13 GVP Y41H. A quantitative representation of these perturbations observed for the spins in different amino acids is provided by Figure 3C. A similar experiment performed on M13 GVP showed no significant differences in the bleaching patterns induced by Gd(DOTP)$^{5-}$. Comparison of Figure 3C (Gd(DOTP)$^{5-}$ bleachings) with Figure 3B (*(dA)$_3$* bleachings) shows that Gd(DOTP)$^{5-}$ binds to wild-type M13 GVP and mutant M13 GVP Y41H at sites that are roughly coincident with the ssDNA binding domain as characterized by the spin-labeled oligonucleotide. The dissociation rate constant for the GVP-Gd(DOTP)$^{5-}$ complex amounts, at pH 7, to approximately 10^3 s^{-1}. Consequently, for the resonances of protons within a distance of ~ 15 Å of the Gd^{3+} ion the correlation between the paramagnetic broadening and the distance to the paramagnetic centre will become negligible due to the long lifetime of the complex. Indeed, we observe that the strong perturbations by Gd(DOTP)$^{5-}$ of the resonances in the DNA binding loop are levelled off to the same magnitude. Taken together with the observed binding stoichiometry (at pH $\geq$ 7) of n~1 for Tb(DOTP)$^{5-}$/GVP (vide supra), we are led to conclude that most likely a high affinity binding site for Ln(DOTP)$^{5-}$ complexes resides in the DNA binding loop.

Reduction of the lifetime of the GVP-Gd(DOTP)$^{5-}$ complex by further lowering of the pH results in a better correlation between linebroadening effects and distances between paramagnetic centre and perturbed protons spins in GVP. As shown in Figure 8A, at pH 5.1 the strong-weak alternations in the perturbation pattern generated by Gd(DOTP)$^{5-}$ in the ^{1}H NMR spectrum of M13 GVP Y41H are more prominent than at pH 7 (Figure 3C). Also at this pH, the perturbations generated by Gd(DOTP)$^{5-}$ in the ^{1}H NMR spectra of wild-type M13 GVP and mutant M13 GVP Y41H did (within experimental error) not differ significantly. It is noted that the resonances from the tip of the so-called DNA binding loop in M13 GVP (residues 20-24), which are moderately perturbed by *(dA)$_3$*, almost completely disappear upon addition of Gd(DOTP)$^{5-}$. Furthermore, we now also observe bleaching of residues

residing in the β-sheet structure comprising the N-terminal segment, in contrast to the result at pH 7. For reasons explained below we take this as an indication of the presence of an extra low affinity binding site which is generated at low pH. Figure 8B summarizes the bleachings observed in the ^{1}H NMR spectrum of IKe GVP upon addition of Gd(DOTP)$^{5-}$ at pH 4.8. The similarity with the paramagnetic effects induced by Gd(DOTP)$^{5-}$ in the ^{1}H NMR spectra of M13 (and Y41H) GVP at low pH are striking. Furthermore, we note that Gd(DOTP)$^{5-}$ perturbs the same domain, except for the low affinity site, as the spin-labeled DNA-fragment *(dA)$_3$* (Figure 3B). Not only are perturbations observed for the region corresponding with those signals that are bleached at pH 7 in M13 (and Y41H) GVP, but we also observe perturbations of spin systems belonging to a domain corresponding to the low affinity binding site in M13 (and Y41H) GVP. NMR measurements on IKe GVP are only feasible in a very limited pH range (4.7-5.5) (de Jong et al., 1987b); though the analogy of the overall effects strongly suggests so, no experimental data can be provided for the pH dependence of the Gd(DOTP)$^{5-}$ binding sites of IKe GVP.

DISCUSSION

General remarks. Recently, departing from a stoichiometry of one Ln(DOTP)$^{5-}$ complex bound per GVP dimer, LIS values of ^{13}C-methylated lysines induced by Tb(DOTP)$^{5-}$ at pH 8.8 were used to fit this ligand in a fixed position on the surface of M13 GVP on the basis of the crystal structure (Dick et al., 1989). Most of the smaller ^{13}C-LIS values were in accordance with a model in which one Tb(DOTP)$^{5-}$ molecule is bound near the GVP dimer's C2 axis of symmetry. The largest ^{13}C-shift however, that of K24, severely violated this model. Therefore, the authors concluded that M13 GVP must undergo a large structural change upon binding of Tb(DOTP)$^{5-}$. Several results argue against this interpretation. Recent ^{1}H NMR studies established that in solution parts of the secondary structure elements of M13 GVP differ from the crystal structure. This is especially true for the regions where K24, K46 and K69 are located (Chapter 4; Folkers et al., 1991b). Therefore, the crystallographic coordinates do not provide a sound basis for detailed structural calculations. The effects observed, in this paper, in the ^{1}H NMR spectrum upon binding of the diamagnetic La(DOTP)$^{5-}$ complex do not support any drastic conformational changes, neither in wild-type

M13, nor in mutant Y41H or in IKe GVP. Furthermore, the effects induced by the paramagnetic complexes suggest the presence of two high affinity $Ln(DOTP)^{5-}$ binding sites per protein dimer (at pH $\geq$ 7), which most likely reside in the DNA binding loops.

The presence of at least two high affinity binding sites for $Ln(DOTP)^{5-}$ on the surface of the GVP dimer and the occurrence of intermediate exchange conditions complicates the use of LIS ^{1}H NMR data for detailed quantitative structural studies. For these reasons, in the present study, the application of the $Ln(DOTP)^{5-}$ complexes was restricted to probe for electropositive patches on the protein surfaces of IKe and M13 GVP.
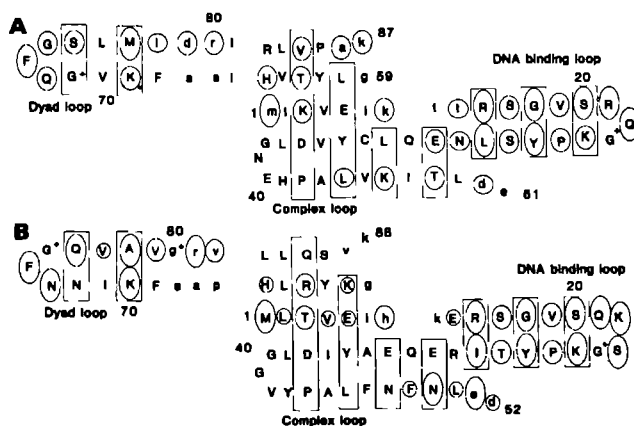

Figure 8. Perturbations, at low pH, in the ^{1}H NMR spectra of (A) M13 GVP Y41H (pH 5.1) and (B) IKe GVP (pH 4.7) upon binding of 0.04 equivalent of $Gd(DOTP)^{5-}$. The relative perturbations were calculated using difference clean-TOCSY spectra (30 ms) recorded at 298 K for 1.2 mM GVP samples. The bleaching effects are represented by ellipses as described in the legend of Figure 3.

The pH dependence of the affinity of the $Ln(DOTP)^{5-}$ complexes for GVP is most likely due to protonation of the phosphonate groups. Quantitative data available for $Dy(DOTP)^{5-}$ indicate that above pH 7 the deprotonated form prevails but that lowering of the pH will result in mono- and diprotonated species (Sherry et al., 1988). Reduction of the negative charge of the lanthanide complex is likely to reduce its affinity for the GVP. It is therefore all the more striking that the resonances of an additional group of residues in the N-terminal segment are bleached out of the spectrum at low pH. We attribute this to one or more positively charged residues of GVP which, at neutral pH, are involved in ion pairing

interactions; destabilisation of the salt bridge upon protonation of the negatively charged residue at low pH may increase the accessibility of the positively charged residue which may then involve in an ionic interaction with the negatively charged lanthanide complex. The protonation behaviour of one of the ^{13}C-methylated lysines (K3 or K7) indeed indicated an ion-pairing interaction for either one of these residues of M13 GVP (Dick et al., 1988). Data available for the terminal amino group suggest that it is buried into an environment that protects it from modification and impedes deprotonation (Dick et al., 1988). Though this does not unambiguously establish an ion-pairing interaction for this group, the pH dependence of the affinity of the Gd-complex for the N-terminal segment can be brought into accordance with an involvement of the amino terminus and/or K3 in salt bridges.

Table I

*Summary of residues in close proximity to bound *(dA)$_3$* that are conserved in M13 and IKe GVP. The spinsystems of the conserved residues listed in italics are perturbed weakly/moderately by *(dA)$_3$* but strongly by Gd(DOTP)$^{5-}$ (see text). The residues marked by superscripts have also been reported in NMR studies involving non spin-labeled oligonucleotides: (a) van Duynhoven et al., 1990; Alma et al., 1981a; (b) Dick et al., 1988; (c) de Jong et al., 1987; (d) King & Coleman, 1987.*

Conserved

M13 GVP:	R16	G18	S20	R21[a]	K24[b]	Y26[a]	E30	K69[b]	F73[a]	R80
IKe GVP:	R16	G18	S20	K22	K25	Y27[c]	E31	K70	F74[c]	R81

Conservative Replacement

M13 GVP:	L28[d]	T48	E50	Q72
IKe GVP:	I29	N49	D51	N73

Involvement of β-sheet structures in ssDNA binding. Perusal of the perturbation pattern induced by *(dA)$_3$* on the ^{1}H NMR spectrum of IKe GVP (Figure 3A) indicates that the perturbed spin systems are clustered in the DNA binding loop, the neighbouring part of the complex loop and in the dyad loop. Comparison with the effects reported for M13 GVP (Folkers et al., 1992; Figure 3B) shows a striking similarity. Conserved and conservatively replaced key residues that are in close proximity to the spin-labeled DNA fragment are

summarized in Table I. Here also the residues are indicated for which their involvement in ssDNA binding has been reported in NMR studies involving non spin-labeled oligonucleotides. The involvement of conserved key residues in the DNA binding loops of IKe (residues 16-29) and M13 GVP (residues 16-28) in the binding to ssDNA and implications for the structure-function relationship have been discussed elaborately (de Jong et al., 1989b; Chapter 4). With the experimental data now available, these considerations can be extended to the dyad loops of M13 (residues 68-78) and IKe GVP (residues 69-79) as well. Also in this element we observe that residues that are most proximal to the bound (spin-labeled) DNA fragment are conserved to a large extent. In this respect, the residues F73 and Q72 (in M13 GVP) and F74 and N73 (in IKe GVP) at the tip of the complex loop and the residues K69 (M13 GVP) and K70 (IKe GVP) in the stem of the loop are worth mentioning. In the part of the complex loop that runs contiguous to the DNA binding loop also perturbed residues are conserved or conservatively replaced i.e. E30, T48, D50 in M13 GVP and E31, N49, E51 in IKe GVP. A peculiar exception in the conservation pattern is formed by the replacement of the charged residue K46 in M13 GVP by the polar residue N47 in IKe GVP which is only weakly perturbed by $*(dA)_3*$. The conserved, perturbed, positively charged residues R80 (M13 GVP) and R81 (IKe GVP) are most likely located in turn-like structures; more precise statements have to await detailed tertiary structure calculcations that are currently in progress.

In the previous sections, it has been noted that at pH 7 the paramagnetic relaxation effects of the spin-labeled oligonucleotide $*(dA)_3*$ can be mimicked by the relatively small, highly negatively charged, ligand $Gd(DOTP)^{5-}$. This endorses the observations made in a number of GVP-ssDNA binding studies that electrostatic interactions make a major contribution to the binding of ssDNA to M13 and IKe GVP (Bulsink et al. 1985; de Jong et al 1987a). A survey of the $Gd(DOTP)^{5-}$ perturbations in the ^1H NMR spectra of IKe and M13 GVP (Figure 3 and Figure 8) leads to the recognition of conserved electropositive residue clusters formed by R16, K22, K25, K70, R81 and R16, R21, K24, K69, R80 respectively. These domains are interspersed with several conserved hydrophobic residues (Y26, L28, F73 and Y27, I29, F73 respectively) for which indications have been found that they participate in hydrophobic (stacking) interactions with nucleotide bases (King & Coleman, 1987; de Jong et al 1987b). The perturbation patterns induced by $Gd((DOTP)^{5-}$ and $*(dA)_3*$ suggest that a

tentative role in hydrophobic interactions with ssDNA may be assigned to M77 in M13 GVP and A78 in IKe GVP. Typically, the mentioned positive and hydrophobic residues can be observed to occur pairwise in close proximity of each other e.g. R16-L28, R24-Y26 and K69-M77 in M13 GVP and R16-I29, K25-Y27 and K70-A78 in IKe GVP.

The high affinity of Gd(DOTP)$^{5-}$ for the DNA binding loop (vide supra) suggests that a major part of the electrostatic interactions with ssDNA may take place in this region. This is supported by the peculiar pK value of the (modified) surface residue K24 in M13 GVP (Dick et al., 1988) which suggests that its surroundings are highly positively charged. The clustering of positive charges in the DNA binding loop may also explain the aberrant binding mode of oligonucleotides that possess terminal dianion phosphates (O'Connor & Coleman, 1982).

In the turn of the DNA binding loops the spin systems of the positively charged residues R21 and K24 in M13 GVP and K22 and K25 in IKe GVP show a weaker perturbation by *(dA)$_3$* than by Gd(DOTP)$^{5-}$. Experiments with non spin-labeled oligonucleotides have shown however, that the turn residues R21 and K24 in M13 GVP are involved in ssDNA binding (Chapter 4; Alma et al., 1981b; Dick et al., 1988). We conclude that the geometry of the *(dA)$_3$*-GVP complex may restrain the possible conformational space of the paramagnetic TEMPO group to positions in which some parts of the ssDNA binding domain (e.g. the tip of the DNA binding loops) are more weakly perturbed than others. For the smaller Ln(DOTP)$^{5-}$ ligands such considerations seem less important which allows us to use them as probes for patches of the protein surfaces of M13 and IKe GVP that may accommodate the negatively charged phosphate-backbone of the longer ssDNA.

Interestingly, both the DNA binding loop and the dyad loop possess almost only amide protons that are in fast exchange with the solvent (Folkers et al., 1991b; previous Chapter). This indicates a high flexibility for these electropositive structural elements, a property which may be required for their binding to ssDNA in an non sequence-specific manner.

Conserved hydrophobic cores in IKe and M13 GVP. The nearly complete ^{1}H NMR spectrum assignment and secondary structure elucidation of M13 and IKe GVP has enabled us to employ the Ln(DOTP)$^{5-}$ complexes as surface probes. In a different approach, earlier work departed from the crystallographic coordinates and a limited number of LIS values to

arrive at far-reaching structural implications. Previously, we argued that the crystallographic data available for M13 GVP provide a weak basis for structural calculations (vide supra). This is illustrated in Figure 3D where the bleaching pattern induced by $Gd(DOTP)^{5-}$ in the M13 GVP ^1H NMR spectrum is indicated in a schematical representation of the dyad loop folding as it exists in the crystal structure (Brayer & McPherson, 1983). Previously, it has been suggested that the dyad loops accommodate the ethylenic groups of $Gd(DOTP)^{5-}$ in a solvent exposed hydrophobic pocket composed of L76 and I78 and their symmetry related counterparts (Dick et al., 1989). The pattern of resonances of M13 GVP bleached by $Gd(DOTP)^{5-}$ matches poorly, however, with the solvent accessibilities as proposed by the crystal structure. The NMR data indicate a β-loop folding in which the hydrophobic residues F68, V70, L76 and I78 are situated on one side of the β-ladder. It is noted that in IKe GVP the hydrophobic nature of these residues is conserved in an analogously folded β-loop. The pattern of perturbations induced by $Gd(DOTP)^{5-}$ in wild-type M13, mutant Y41H and IKe GVP (Figure 3) indicate that these hydrophobic residues are folded into the interior of the protein. The other side of the β-sheet in IKe and M13 GVP accommodates positive (K70 and K69 respectively) and polar residues (Q72 and S75 in M13 GVP and N72, N73 and Q76 in IKe GVP) which may more easily facilitate binding of negatively charged ligands (e.g. ssDNA, $Ln(DOTP)^{5-}$) to the protein surface than the hydrophobic pocket proposed by the crystal structure. The same pattern of hydrophobic residues on one side of the β-sheet and polar residues on the other side can be recognized in other regions of the β-sheet structure of IKe and M13 GVP. In M13 and IKe GVP, the β-sheet structure comprising the N-terminal segment which is perturbed by $Gd(DOTP)^{5-}$ at low pH contains, on one side of the β-sheet, several polar and positively charged residues. The pattern of perturbations by $Gd(DOTP)^{5-}$ indicates that this side of this structural element is faced towards the solvent in M13 and IKe GVP. The other side of this β-sheet structure is predominantly occupied with hydrophobic residues, e.g. Y61 in M13 and Y62 in IKe GVP which are most likely accommodated in the interior of the protein. Experimental evidence is provided by the ring flip of Y61 which is intermediate on the NMR time scale while the motions of other aromatic residues are fast (Folkers et al., 1991a). This is indicative for a restrained mobility of this side chain. Furthermore, the hydroxy protons of the aromatic rings of Y61 in M13 GVP and Y62 in IKe

GVP are slowly exchanging with the solvent, which is also indicative for accommodation of these residues in the interior of the protein (unpublished results).
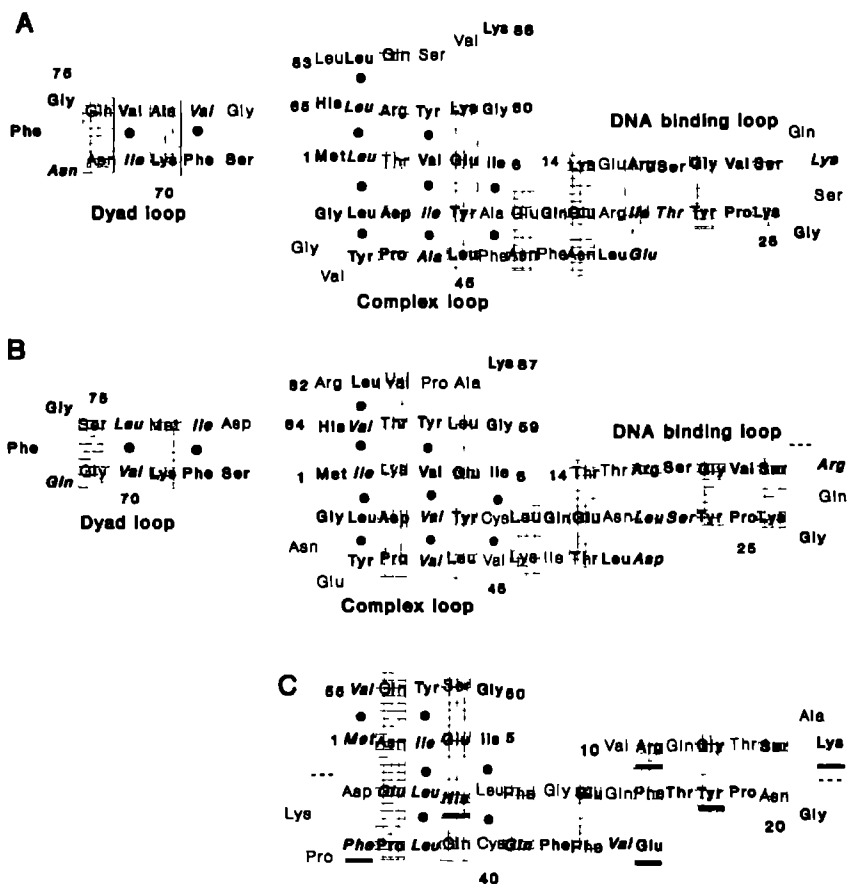
Figure 9. Pictorial representation of the amino acid sequences of (A) IKe and (B) M13 GVP showing the β-sheet folding of the amino acid chain. Conserved residues and semi-conserved residues are indicated in bold and bold italic respectively. Dashes represent insertions. The residues enclosed by boxes are pointing towards the reader. Filled circles indicate pairs of conserved or semi-conserved apolar residues that reside at opposite sides of antiparallel aligned amino acid chains. In (C) a putative β-sheet structure is depicted for Pf3 DBP for the amino acid segments running from residues 1 to 5, 10 to 45 and 50 to 54. The underlined residues are also conserved in other ssDNA binding proteins (see text).

Taken all the experimental evidence together, we can recognize in the β-sheet structures delineated for IKe and M13 GVP two hydrophobic surfaces (i.e. in the so-called dyad loop and in the β-sheet comprising the N-terminal segment) which are, according to the experimental data, folded into the interior of the protein. Globally, these data are in accordance with the spatial folding of the β-sheet elements as proposed by crystallographic data. According to the crystal structure of the M13 GVP dimer, the N-terminal segment protrudes into the solvent as an extended chain that is not involved in any β-sheet structure (Brayer & McPherson, 1983). The crystallographic data suggest a hydrophobic interaction between the N-terminal segment and the dyad loop. In contrast to this, the NMR data indicate an involvement of the N-terminal segment in a regular β-sheet structure. Furthermore, in both the dyad loop (vide supra) and the N-terminal segment the hydrophobic residues are now far more regularly distributed on one side of the β-sheet (indicated by dots in Figure 9). This provides a more solid basis for the hydrophobic intermonomer interaction between dyad loop and N-terminal segment than proposed on the basis of the crystal structure.

Sequence similarity to Pf3 DBP and other ssDNA binding proteins. It is of interest whether the structural features of IKe and M13 GVP and their involvement in ssDNA binding are similar for other proteins. Structural information for ssDNA binding proteins other than M13 and IKe GVP is not available, therefore only homologies in amino acid sequences are at our disposal to uncover such similarities. A straightforward amino acid sequence comparison of IKe GVP with the ssDNA binding protein Pf3 DBP encoded by the phage Pf3 of subclass II yields a sequence homology that is significantly less than with M13 GVP (Luiten et al., 1985). The key residues present in the DNA binding loops of M13 and IKe GVP are strikingly well conserved in the amino acid sequence of Pf3 DBP, however, and gave further support to the proposal of this β-loop element as a motif in ssDNA binding (de Jong et al., 1989b). Folding of the amino acid chains as indicated in Figure 9C leads to an alignment in which the conservation of many other residues in M13 and IKe GVP may be extrapolated to Pf3 DBP. This representation corresponds to the triple-stranded β-sheet structure in IKe and M13 GVP; for other secondary structure elements (e.g. the dyad loop) searches for an appropriate Pf3 DBP sequence alignment yielded far less impressive results. Residue Q45 in Pf3 DBP can be considered as a conservation of the β-breakers E51 and D50

in IKe and M13 GVP respectively. Of interest is the pairwise conservation of hydrophobic residues which are accommodated in the interiors of IKe and M13 GVP (vide supra). Note how these are conserved to a larger extent in Pf3 DBP than the solvent exposed residues. This in contrast to the putative DNA binding loop of Pf3 DBP where residues involved in ssDNA binding in M13 and IKe GVP are neater conserved than the residues located at the other side of the β-sheet. These two patterns of conserved residues are characteristic of elements involved in maintaining the structural integrity of a protein and elements involved in function (i.e. ssDNA binding) respectively.

Using the amino acid sequence of M13 GVP as a template, domains in the primary sequences of a variety of other ssDNA binding proteins have been searched for that yielded optimal alignment of aromatic and charged residues (Prasad & Chiu, 1987). The underlined residues in Figure 9C are not only conserved in Pf3 DBP, M13 and IKe GVP but also in the gene product 32 of phage T4 (GP32) (Williams et al., 1981), RecA (Sancar et al., 1980), SSB (Sancar et al., 1981) and SSF (Chase et al., 1983) (the latter three encoded by *Escherichia coli*). This suggests that a triple-stranded β-sheet structure may be a common denominator in ssDNA binding proteins.

CHAPTER 7

A Transferred NOE Study on the Conformation of Oligonucleotides in Complexes with the Protein Encoded by Gene V of the Bacteriophage M13[*]

ABSTRACT

The conformation of the oligonucleotides TGCAACGA, TCGTTGCA, D–$(dA)_6$ and its enantiomer L–$(dA)_6$ in complexes formed with the ssDNA binding protein encoded by gene V of the bacteriophage M13 was studied by means of the transferred NOE effect and CD measurements. The TrNOEs were quantified by iterative back-calculation, in analogy of the NO2DI approach described for conventional NOESY spectra (van de Ven, F.J.M., Blommers, M.J.J., Schouten, R.E., & Hilbers, C.W. (1991) J. Magn. Reson. 94, 140-151). The NMR data indicate that the bound oligonucleotides preferentially adopt extended structures, i.e. partially destacked bases in anti-conformations. The experiments with TCGTTGCA and TGCAACGA suggest that interactions between neighbouring bases in the oligonucleotide govern the degree of destacking in the bound conformations. Experiments with the L- and D-enantiomers of $(dA)_6$ showed that the binding of M13 GVP to oligonucleotides is only to a limited extent selective for the chirality of the sugar-phosphate backbone. This suggests an flexible and mobile type of interaction between oligonucleotides and GVP which may be biologically relevant towards sliding and translocation mechanisms in the nucleoprotein complex.

INTRODUCTION

Interactions between proteins and nucleic acids play an important role in the regulation of genetic processes in the living cell. Crucial roles in DNA replication are played by helix destabilizing proteins, i.e., proteins with a higher affinity for single-stranded than for double-stranded DNA. An important model for ssDNA binding proteins is encoded by gene V of the filamentous bacteriophage M13, that uses *Escherichia coli* as host (Model & Russel, 1988). In the life-cycle of this conjugative plasmid specific phage this ssDNA binding protein (M13 GVP) plays an indispensable role. After a certain threshold concentration has been reached it forms large nucleoprotein complexes with the viral DNA strand which is thereby sequestered from the replication cycle.

interactions (e.g. between enzyme and substrate) this feature has been well established. In nature the sugar backbone of native (D-)DNA consists exclusively of 2-deoxy-D-ribose units. Recently, enantio- or L-DNA (DNA with 2-deoxy-L-erythro-pentose as sugar backbone) has been employed to study the chiral specificity of the formation of nucleic acid structures (Shimamoto & Utiyama, 1983; Urata et al., 1991). To our knowledge this paper is the first occasion in which the DNA-protein interaction in ssDNA binding proteins is probed for its enantio-selectivity.

MATERIALS AND METHODS
MATERIALS

The isolation and purification of M13 GVP were as described in Chapter 3 and 4. Lyophilized M13 GVP was stored -20 $^{\circ}$C. Its concentration was determined using a molar extinction coefficient of 7100 $M^{-1}cm^{-1}$ (per monomer) at the absorption maximum (276 nm) (Day, 1973). The L-adenosine units were prepared according to Asseline et al. (1991). The oligonucleotides TGCAACGA, TCGTTGCA and $(dA)_6$ were synthesized using the phosphotriester method (van Boom et al., 1983). All oligonucleotides lack the 5' and 3' terminal phosphates and were used in their Na^+-forms. Oligonucleotide concentrations are given in mononucleotide units and were determined by measuring the UV absorbance at 257 nm using estimated molar extinction coefficients (Cassani & Bollum, 1969).

METHODS

Experimental. CD measurements were performed, at 293 K, on a Mark V circular dichrometer interfaced to an AT Personal Computer which was equipped with a home-written acquisition program (courtesy of the Laboratory of Biochemistry, Wageningen Agricultural University, the Netherlands). The CD spectra were smoothed by means of a simplified least-squares procedure (13-point moving polynomial smooth) (Savitzky & Golay, 1964) and are presented as ε_L-ε_R ($M^{-1}cm^{-1}$, in mononucleotide units). The contribution of M13 GVP to the CD spectrum was corrected for by subtraction of the blank (protein without oligonucleotide).

Physico-chemical studies resolved much of the thermodynamics by which this process can be characterized. Generally speaking, two binding modes exist whose prevalence is governed by the length of the nucleotide chain and by the ionic strength of the solution (Alma et al., 1982; Bulsink et al., 1986; Kansky et al., 1986; Bulsink et al., 1988b). The so-called polynucleotide binding mode is characterized by high cooperativity factors and a stoichiometry of one protein per four nucleotides. In the oligonucleotide mode, which dominates at low ionic strength and/or chain lengths of less than 25 residues, three nucleotides bind to one protein molecule. The cooperativity factors for the binding to the short chain lengths are two orders of magnitude smaller than for the longer ones.

Recently, much progress has been made in resolving the structure of M13 GVP by means of ^{1}H NMR. The DNA-binding domain of M13 GVP, as it has been mapped by spin-labeled DNA fragments and inorganic, negatively charged, lanthanide complexes, consists of electropositive β-sheet structures patched with hydrophobic residues (Folkers et al., 1992; Chapter 6). The conservation of these features in the gene V protein of the distantly related phage IKe suggests that non sequence-specific binding of ssDNA to this class of proteins is mediated by a common structural denominator.

At the level of the ssDNA conformation in the nucleoprotein complex our knowledge pertains mainly to its overall structure. Neutron scattering (Gray et al., 1982a) and electron-microscopic studies (Gray et al., 1982b; Gray, 1989) showed that the ssDNA strands are packed inside rod-like left-handed helical protein assemblies. Thus far, ^{1}H NMR studies (Alma et al., 1983a; King & Coleman, 1987) have provided only limited insight in the ssDNA conformation at the sub-molecular level. For a meaningful approach towards model-building studies of the nucleoprotein assembly, both the overall structure and microscopic data on the ssDNA conformation are required. The present study aims at a quantitative interpretation of the structural details of oligonucleotides in their complexes with M13 GVP. For this purpose complexes of the single-stranded octanucleotides TCGTTGCA, TGCAACGA, the hexanucleotide D–(dA)6 and its enantiomer L–(dA)$_6$ with M13 GVP were subjected to transferred NOE studies. In contrast to the hexanucleotides, the octamers showed far less overlap in the ^{1}H NMR spectrum and were therefore more useful for the quantitative evaluation of the observed transferred NOEs. The enantiomeric forms of (dA)$_6$ were employed to study the chiral specificity of the ssDNA-GVP interaction. For many other protein-ligand

The NMR experiments were performed at 400 or 500 MHz on a Bruker AM400 or on a AM500 spectrometer respectively. All oligonucleotide samples were dissolved in D_2O (pH 7) and were subjected to phase-sensitive DQF-COSY experiments (Rance et al., 1983). Before and after addition of aliquots of M13 GVP, inversion-recovery and NOESY experiments with mixing times ranging from 75 to 500 ms were performed. Preacquisition delays of 1.2-1.7 ms were used during which the HDO resonance was suppressed by coherent low-power irradiation (Zuiderweg et al., 1986). Prior to the recording of the 2D-experiments the receiver phase and a delay after the read-out pulse were adjusted to minimize baseline artifacts (Marion & Bax, 1988a). In all 2D-experiments, TPPI was used for signal accumulation in the t_1 dimension (Marion & Wüthrich, 1983).

NMR data were processed either on a Bruker ASPECT 3000 data station using Bruker software, or on a SUN 4 computer using the NMRi package (Release 1.1.). Standard Bruker software was used for the analysis of the inversion-recovery experiments. Baseline corrections by polynomial functions were performed using home-written algorithms run on the SUN 4 computer. (C-programs were kindly provided by Dr. A. Lommen). Volumes of peaks in NOESY spectra were determined by integration using the standard NMRi option.

Quantitative interpretation of the two-dimensional transferred NOE effect. The time evolution of the longitudinal magnetizations during the mixing time, τ_m, in the presence of chemical exchange between a free (F) and bound (B) form of a spin system can be described by a set of coupled differential (Bloch) equations:

$$(1) \qquad \frac{d}{d\tau_m}\begin{pmatrix} M_B^z \\ M_F^z \end{pmatrix}(\tau_m) = \left[\begin{pmatrix} L_B(\tau_{c,B}) & 0 \\ 0 & L_F(\tau_{c,F}) \end{pmatrix} + \begin{pmatrix} -K_B & K_F \\ K_B & -K_F \end{pmatrix} \right] \begin{pmatrix} M_B^z - M_B^0 \\ M_F^z - M_F^0 \end{pmatrix}(\tau_m)$$

Here M_B and M_F are n-dimensional vectors that represent the longitudinal magnetizations of the spins system in the bound and free form respectively. The corresponding longitudinal relaxation matrices are represented by $L_B(\tau_{c,B})$ and $L_F(\tau_{c,F})$, where $\tau_{c,B}$ and $\tau_{c,F}$ are the rotational correlation times of the bound and free forms respectively. In a typical transferred NOE experiment, the rotational correlation times of the bound ($\tau_{c,B}$) and free ($\tau_{c,F}$) forms

differ by an order of magnitude. K_B and K_F are rate matrices obtained by multiplying the unity matrices with the corresponding rate constants k_B and k_F that describe the pseudo-first-order reaction between free and bound form of the oligonucleotide. In Chapter 2, it has been shown that in the presence of fast exchange on the longitudinal relaxation scale, i.e.

$$(2) \qquad k_F + k_B >> [L_B(\tau_{c,B}) - L_F(\tau_{c,F})]_{ij}$$

equation (1) simplifies to:

$$(3) \qquad \frac{d}{d\tau_m}(M_B^z - M_F^z)(\tau_m) = L_{av}[(M_B^z - M_F^z - M_B^0 - M_F^0)(\tau_m)] \quad L_{av} = p_B L_B(\tau_{c,B}) - p_F L_F(\tau_{c,F})$$

In this expression the motion of the longitudinal sum magnetizations of bound and free forms is described by the weighted average, L_{av}, of the relaxation matrices L_B and L_F.

It is noted that during the t_1 and t_2 times and the mixing time, τ_m, of the 2D-NOE experiment, different time-scales pertain, i.e. the chemical shift and longitudinal relaxation scales respectively. For the transferred NOE experiments described in this paper, $\tau_{c,F}$ is in the regime where the off-diagonal elements in the relaxation matrix of the free form, L_F, approach zero. The diagonal elements can therefore be set equal to the reciprocal values of the longitudinal relaxation times of the free oligonucleotide $T_{1,F}$ which can easily be determined by inversion-recovery experiments.

By rearranging and integration of Equation (3) we arrive at:

$$(4) \qquad N(\tau_m) = S \, e^{-L^{av}(p_B \tau_m)}$$

which describes the evolution of the measured transferred NOE effects $N(\tau_m)$. Here a scaling factor, S, is introduced to relate the experimental with the simulated peak intensities. The elements of the matrix L^{av} are given by:

$$(5) \qquad L_{ii}^{av} = \frac{p_F}{p_B} T_{1,F}^{\;1} + \frac{1}{10}\sum \frac{\gamma^4 \hbar^2}{d_{ij}^6}[J_0 + 3J_1 + 6J_2]$$

(6)
$$L_{ij,i\neq j}^{av} = \frac{1}{10} \frac{\gamma^4 \hbar^2}{d_{ij}^6} [6J_2 - J_0]$$

In these expressions d_{ij} is the distance between spins i and j in the bound conformation. The spectral density functions, J_n (n = 0,1,2), take the form:

(7)
$$J_n = \frac{\tau_{c,B}}{[1-(n\omega\tau_{c,B})^2]}$$

Here ω is the Larmor precession frequency. As can be seen in Equation 4, the fraction p_B has the main effect of scaling down the mixing time, τ_m. As a secondary effect, an extra leakage term is introduced in the relaxation matrix which accounts for the longitudinal relaxation in the unbound form F ($T_{1,F}^{-1}$).

These considerations were, in a rather straightforward manner, incorporated in the so-called NO2DI algorithm for the determination of distances, r_{ij}, from a set of experimental NOESY intensities (van de Ven et al., 1991). This algorithm, here tailored to TrNOESY applications (and henceforth referred to as TO2DI), adjusts a set of ^{1}H-^{1}H distances in a sequential manner by matching simulated and experimental TrNOEs. As is discussed in the original description of NO2DI (van de Ven et al., 1991) this method is likely to work if the distances are refined in the order of descending magnitudes of the corresponding NOEs and if the first-order build-up terms (corresponding to $L_{ij}^{av}(\tau_m)$, vide supra) still dominate in $N_{ij}(\tau_m)$.

RESULTS

Characterisation of the $(dA)_6$ enantiomers, and their binding to GVP, by CD measurements. The CD spectrum of D–$(dA)_6$ (thick line in Figure 1A) is characteristic of a single-stranded right-handed helical structure with anti-conformations of the stacked bases (Olsthoorn et al., 1980). Signals with the same position and magnitude but of opposite sign were observed in the CD spectrum of L–$(dA)_6$ (thick line in Figure 1B). Thus, the L-oligo-
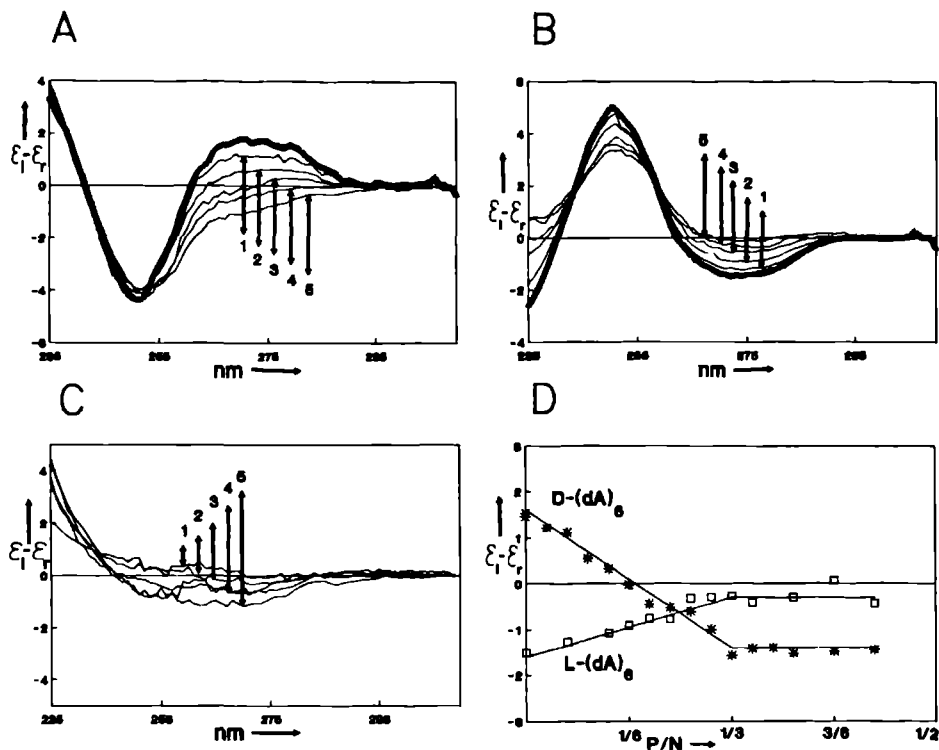
Figure 1. The thick lines in (A) and (B) are the CD spectra of $D-(dA)_6$ and $L-(dA)_6$ respectively (0.26 mM mononucleotide). The thin lines represent the CD spectra of these oligonucleotides after addition of aliquots of M13 GVP (P/N ratios: (1) 0.066, (2) 0.132, (3) 0.198, (4) 0.264, (5) 0.330). In these CD spectra the contribution of M13 GVP (the blank spectrum) has been subtracted. At the higher P/N ratios parts of the short wavelength (<230 nm) CD spectra were perturbed by protein scatter. Addition of the CD spectra in (A) and (B) results in the titration series listed in (C). Figure D shows the elipticity at 270 nm during titrations of L- and $D-(dA)_6$ with M13 GVP. Note the sharp deflection points of the curves at P/N = 1/3.

nucleotide adopts a conformation which is the mirror image of the $D-(dA)_6$ conformation, i.e. a single-stranded left-handed helical structure. Upon addition of M13 GVP to $L-(dA)_6$, we observe changes in the positive as well as the negative CD bands which are indicative of partial base-destacking and extension of the helical structure (Olsthoorn et al., 1980). In the long and short wavelength bands of the CD spectrum of $D-(dA)_6$ similar changes can be observed which also point towards partial destacking. The 250 nm band, however, undergoes relatively small changes upon addition of M13 GVP. The loss of mirror symmetry between

the CD spectra upon addition of M13 GVP to the $(dA)_6$ enantiomers is visualized in Figure 1C, which shows the summations of the spectra shown in Figure 1A and 1B. Upon complexation to M13 GVP we observe the appearance of a weak, broad band in the CD sum-spectrum, extending from 245 to 280 nm, and a short wavelength band near 230 nm. The short wavelength band overlaps with the intrinsic protein CD band at 228 nm which arises from one or more tyrosyl residues (Day, 1973). The long wavelength band is not congruent with the CD spectrum of unbound $D–(dA)_6$ which indicates that it is not just the *degree* of destacking of the left- and right-handed helices that causes the appearance of these bands (Olsthoorn et al., 1980). At this point, we may explain the appearance of both bands by the loss of mirror-image symmetry of the enantio-oligonucleotide structures and/or by protein (tyrosyl) chromophores that become optically active at these wavelengths.

The titration curves in Figure 1D (elipticity at 270 nm vs. P/N ratio) show that both $(dA)_6$ enantiomers bind to M13 GVP with stoichiometries that are in good correspondence with the expected n=3 binding mode (Alma et al., 1982; Kansy et al., 1986). The linearity of the titration curves and the sharp deflections at the equivalence points indicate that also the association constants must be of the same order of magnitude.

Characterisation of binding of oligonucleotides to M13 GVP by ^{1}H NMR titrations.
$\underline{D-(dA)_6 \text{ and } L-(dA)_6}$: As the spin systems in the unbound $(dA)_6$ enantiomers are in magnetically equivalent environments, NMR measurements will produce identical spectra. Figure 2A shows the course of the H8 resonance positions during the titrations of $D–(dA)_6$ and $L–(dA)_6$ with M13 GVP. To a large extent, the magnetic equivalence of the free forms of the enantiomers is maintained upon complexation with M13 GVP. Upon addition of GVP to $D–(dA)_6$ and $L–(dA)_6$, their H8 resonances shift to nearly coincident (but not identical) positions. The similarity of the titration curves can be attributed to unwinding of both the left- and right-handed single-stranded helical oligonucleotide structures upon binding to M13 GVP (vide supra).

Also in the protein, both enantiomers produce similar effects. As an illustration, Figure 2B shows the course of the C^εH-Y26 resonances during titrations of M13 GVP with $D–(dA)6$ and $L–(dA)6$. In correspondence with the previous results both curves indicate similar binding modes (n=3) and binding affinities of the same order of magnitude. The
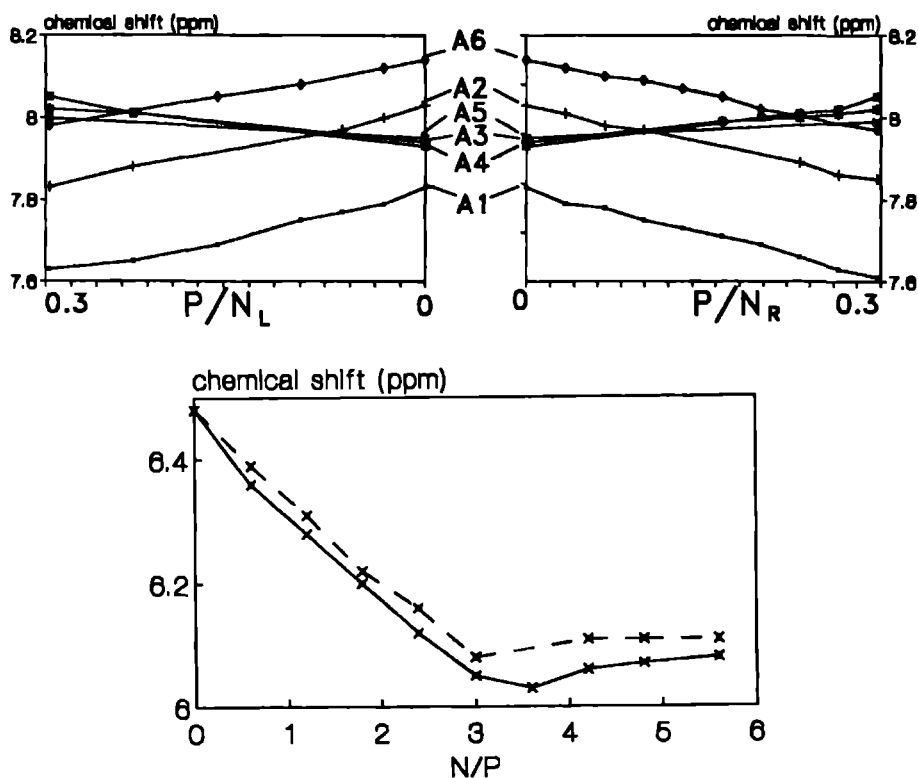
Figure 2. (A) *Chemical shifts (400 MHz) of the H8 resonances of 3.0 mM L–(dA)$_6$ (right half) and D–(dA)$_6$ (left half) observed upon addition of M13 GVP. The curves are presented as a function of the ratio of protein monomers to mononucleotide units (P/N). Note the symmetric appearance of the titration curves of these enantiomers.*
(B) *Chemical shift (500 MHz) of the C$^\epsilon$H-Y26 resonance of M13 GVP (0.2 mM) upon addition of D–(dA)$_6$ (———) and L–(dA)$_6$ (- - -). The titration curve is presented as a function of mononucleotide to protein monomer units (N/P).*

maximum shifts induced by D–(dA)6 and L–(dA)6 differ only slightly.

The gradual shift of the oligonucleotide and protein resonances during the titrations indicates that the bound and free forms are in fast exchange on the chemical shift scale. We can therefore assume that exchange on the spin-lattice relaxation scale is also in the fast exchange regime. Further evidence is provided by measurement of T_1 relaxation times of the observed oligonucleotide resonances as a function of equivalents of added M13 GVP (Figure 3). The linear dependence of T_1^{-1} on the P/N-ratio that can be observed in this plot indicates

that the spin-lattice and cross-relaxation are ruled by the average of the longitudinal relaxation matrices of free and bound oligonucleotides (Chapter 2 of this thesis).
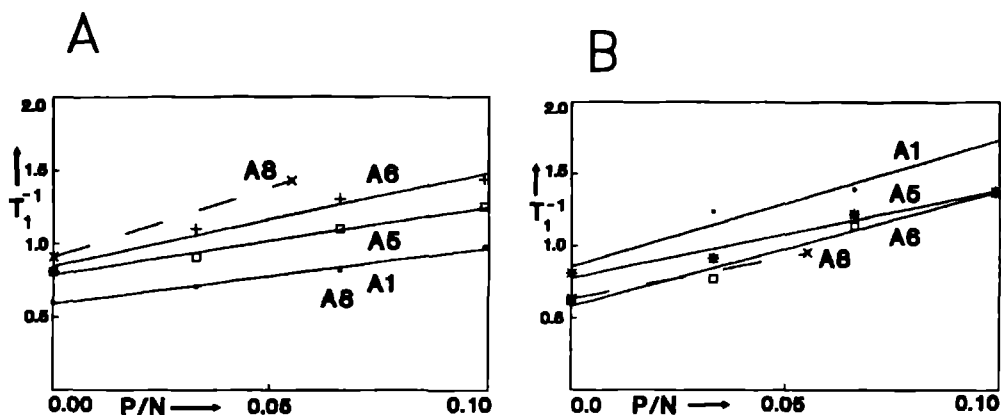

Figure 3. T_1^{-1} values (reciprocal longitudinal relaxation times) of the H8 (A) and H1' (B) spins during a titration of 3.0 mM D-(dA)$_6$ with aliquots of M13 GVP. Due to overlap of broadened resonances only a limited range of the titration can be monitored by inversion-recovery experiments. The dashed line indicates the course of T_1^{-1} of the H1' and H8 proton spins of the A8 nucleotide TGCAACGA upon addition of M13 GVP.

<u>TGCAACGA and TCGTTGCA:</u> Upon addition of M13 GVP to these octamers we observe that their resonances undergo relatively small shifts. Concomitantly, a considerable linebroadening can be observed which indicates that their binding affinities are sufficiently high to bring free and bound oligonucleotide into the intermediate exchange regime (on the chemical shift scale; estimates for the dissociation rate constants for the octamer-GVP complexes lie in the range 200-300 s^{-1}). The decrease of the T_1 relaxation rate of the octamer proton spins upon complexation with M13 GVP, is in the same order of magnitude as the decrease in longitudinal relaxation time of the corresponding D-(dA)$_6$ proton spins. This indicates that the bound and free forms are in fast exchange on the spin-lattice relaxation scale. This justifies the use of average longitudinal relaxation matrices for the quantitative interpretation of the observed transferred NOE effects (see METHODS section).

Sequential assignment of the ^{1}H NMR spectra of (dA)$_6$, TCGTTGCA and TGCAACGA.
Using the through-bond connectivities, as they are manifested in DQF-COSY spectra (not

Table I

^{1}H NMR resonance assignments of (A) (dA)[6], (B) TCGTTGCA and (B) TGCAACGA at 298 K. Chemical shifts (± 0.03 ppm) are quoted relative to DSS.

(A) (dA)6

residue	H8	H1'	H2'	H2"	H3'
A1	7.85	5.93	2.10	2.38	
A2	8.06	5.65	2.40	2.54	
A3	7.99	5.65	2.30	2.43	
A4	7.95	5.65	2.30	2.43	
A5	7.97	5.79	2.31	2.50	
A6	8.16	6.13	2.50	2.32	4.60

(B) TGCAACGA

residue	8/6	*	1'	2'	2"	3'
T1	7.37	1.18	6.00	1.77	2.24	4.62
G2	7.96		2.67	2.67	2.59	4.92
C3	7.34	5.63	5.96	1.67	2.13	4.68
A4	8.16	7.90	5.90	2.56	2.60	4.92
A5	8.21	7.79	6.10	2.57	2.57	4.94
C6	7.41	5.77	5.98	1.73	2.21	4.96
G7	7.75		5.66	2.34	2.34	4.93
A8	8.30	8.03	6.33	2.77	2.53	4.85

(C) TCGTTGCA

residue	8/6	*	1'	2'	2"	3'
T1	7.54	1.81	6.14	2.24	2.42	4.70
C2	7.58	5.91	6.12	2.42	2.23	4.77
G3	7.91		6.01	2.76	2.63	4.95
T4	7.48	1.69	6.11	1.78	2.18	4.77
T5	7.40	1.79	6.01	1.95	2.23	4.92
G6	7.94		6.00	2.69	2.57	4.92
C7	7.49	5.78	6.02	1.82	2.27	4.92
A8	8.33	8.10	6.37	2.77	2.55	4.72

*   *Position of either the H5, H2 or methyl resonance (of C, A or T respectively).*

shown), the major part of the resonances in the ^{1}H NMR spectra of the free oligonucleotides could be assigned to a specific type of protons. Due to overlap, most H4', H5' and H5" resonances could, however, not be identified unambiguously. In the customary assignment
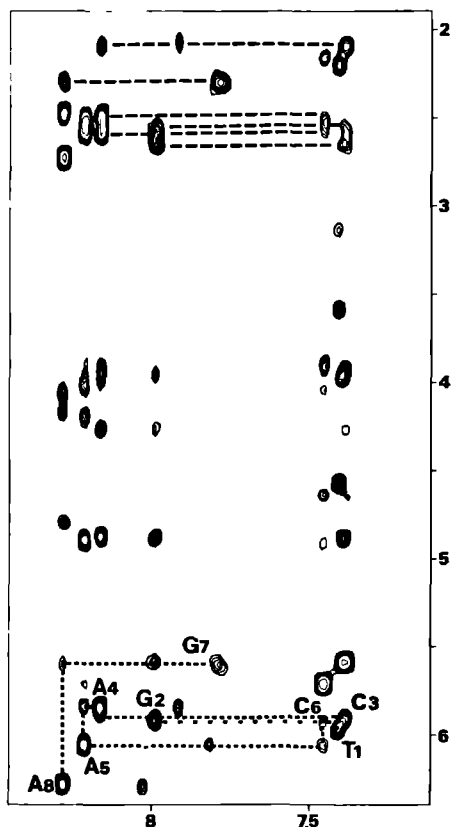
Figure 4. Sequential walks in the H8-H1' and H8-H2'/H2" regions of the 500 ms NOESY spectrum TGCAACGA (P/N =0.055). The sample, dissolved in D_2O, contained 11.2 mM mononucleotide.

protocol (Wüthrich, 1986) inter-nucleotide NOEs are used for identification of connectivity patterns in a sequence specific manner. At room temperature, however, for all of the oligonucleotides free in solution, neither intra- nor inter-nucleotide NOEs can be observed, even for mixing times as long as 600 ms. The absence of visible NOE effects can be attributed to unfavourable rotation correlation times and rapid internal motions (e.g. rapid stacking-destacking) in these molecules. Addition of small aliquots of M13 GVP to the nucleotide solutions, however, results in the appearance of strong intra- and inter-nucleotide contacts. An example is presented in Figure 4 which shows the NOESY spectrum of TGCAACGA, after an aliquot of M13 GVP had been added to the sample. The observed connectivities arise from the transferred NOE effect and reflect the conformation of the oligonucleotide in its complex with M13 GVP. The inter-nucleotide contacts, summarized in Figure 5, allow for the sequential assignments of the ^1H NMR spectra of D- and L-(dA)6
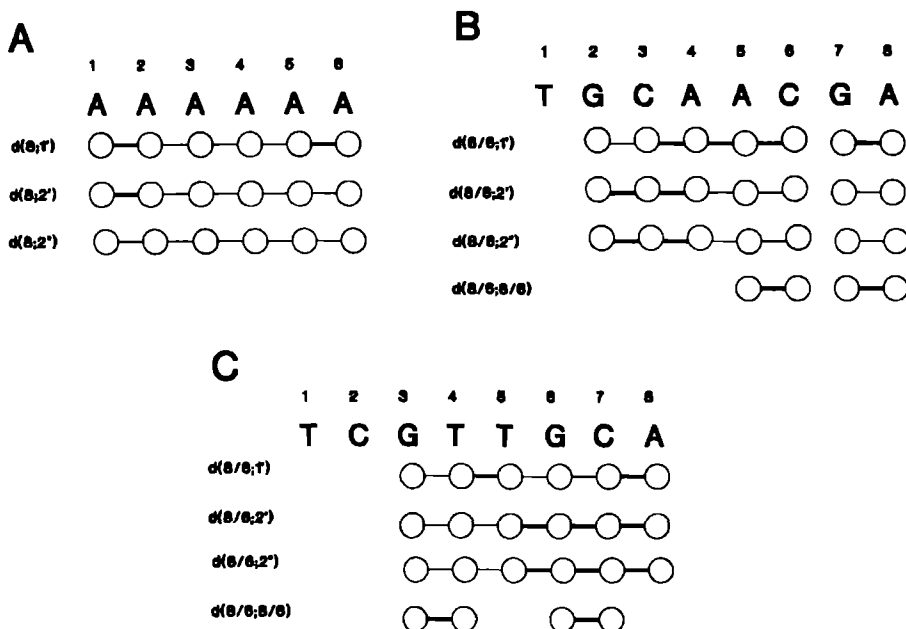
Figure 5. A graphical summary of the observed sequential transferred NOE contacts in 400 MHz NOESY spectra (298 K, τ_m= 75-500 ms) of solutions of (A) 10.2 mM $(dA)_6$, (B) 11.2 mM TCGTTGCA and (C) 11.2 mM TGCAACGA to which M13 GVP has been added to final P/N ratios of 0.075, 0.055 and 0.055 respectively. Dashed lines indicate transferred NOEs which could not be quantified accurately because of overlap.

(both enantiomers have identical sets of resonances, vide supra), TGCAACGA and TCGTTGCA (Table I). With equal amounts of M13 GVP added, both D–$(dA)_6$ and L–$(dA)_6$ yield indistinguishable transferred NOESY spectra. A straightforward sequential analysis of these spectra is precluded by the large degree of resonance overlap. Scrutiny of a series of transferred NOESY experiments, recorded at different P/N ratios, enabled us to recognize sufficient inter-nucleotide contacts to make a complete sequential assignment.

In the octanucleotides the degree of overlap was far less. In spite of that, even at prolonged mixing times, no sequential contacts between T1-C2 and C2-G3 in TCGTTGCA and T1-G2 and C6-G7 in TGCAACGA could be observed; these residues could be assigned however by elimination.

The prevalence of the H8-H2"-contacts in the sequential walks is suggestive of a B-DNA type helical conformation of the oligonucleotide fragments in their complexes with M13

GVP. Compared to most intra-nucleotide contacts these sequential NOEs only become apparent at rather long mixing times. This indicates that the bound oligonucleotide must have a rather extended helical structure with partially destacked bases.

In the next section the observed NOEs are interpreted in a quantitative manner by iterative back-calculation of the transferred NOE effects.

Table II

Average T_1 relaxation times of proton spins in the free form of $(dA)_6$. The relaxation times of the proton spins in the octanucleotides TCGTTGCA and TGCAACGA differ only to a minor extent from the listed values. Therefore only one set of T_1 relaxation times is used in the calculations.

resonance	T_{1F} [s]
H8,H5,H6	1.4
H2	5.0
H1'	1.4
H2',H2"	0.5
H3'	0.7
H4'	0.7
H5',H5"	0.4

Quantitative evaluation of the Transferred NOEs. The experimental results show that the bound and free forms of the oligonucleotides are in fast exchange on the longitudinal relaxation scale. In addition, $\tau_{c,F}$ is in the regime where the off-diagonal terms in L_F can be neglected. Both requirements for the quantification of the transferred NOE effects, as described in the METHODS section, are therefore obeyed. As the $T_{1,F}$ relaxation times for the different types of spins do not differ significantly among the mononucleotide units, only one set of $T_{1,F}^{-1}$ values was used in the NOE back-calculations (Table II). The algorithm also requires the fraction of bound oligonucleotide, p_B, the rotation correlation time, $\tau_{c,B}$, and the scaling factor, S, as input (see Methods). Elaboration of Equations 4-6 shows that the quantities $\tau_{c,B}$ and p_B are strongly correlated with S, i.e. an overestimation of S can be corrected by an underestimation of τ_c and p_B (and vice versa). The fraction of bound material, p_B, can be estimated from total M13 GVP and oligonucleotide concentrations (one can assume that virtually all M13 GVP is bound when the oligonucleotide is present in excess).

Table III

Inter-proton distances [Å] within the nucleotide unit of a B-DNA type structure. Only distances that may contribute significantly to spindiffusion are listed, i.e. distances smaller than 2.7 Å. The inter-proton distances in italics do not differ significantly over the conformational space spanned by sugar pucker P, pucker amplitude Φ_{max} and γ (<(5'O-5'C-4'C-3'C)).

	H5	H6/8	H1'	H2'	H2"	H3'	H4'	H5'	H5"
H5	---	2.4							
H6/8		---		2.2					
H1'			---	*2.4*					
H2'				---	*1.8*	*2.4*	*2.4*[1]		
H2"					---	*2.7*			
H3'						---	*2.7*	*2.5*[2]	
H4'							---	*2.5*	*2.4*
H5'								---	*1.8*
H5"									---

1 *N-pucker sugar ring conformation, in any other ring conformation: $d_i(2"-4') > 2.7\,Å$*

2 *$\gamma \approx 60°$ (+ synclinal), for any other conformation: $d_i(3'-5') > 2.7\,Å$*

Using standard B-DNA distances (Table III) and starting values for S and τ_c, the cross and diagonal TrNOESY peak intensities were calculated. Subsequently, new and better values for the scaling factor, S, and rotation correlation time, τ_c, were calculated by the algorithm TO2ST, by matching the simulated and experimental TrNOEs (N_{ij}^{sim} and N_{ij}^{exp} respectively):

$$(8) \qquad S(new) = S(old)\sum_i \frac{N_{i,i}^{sim}}{N_{i,i}^{exp}}$$

$$(9) \qquad \tau_c(new) = \tau_c(old)\sum_i \sum_j \frac{N_{i,j}^{sim}}{N_{i,j}^{exp}}$$

It is our experience that a few recalculations (<5) suffice to obtain a combination of p_B, τ_c and S for which the set of distances roughly corresponds with the experimental data. Using these values for S, τ_c, and p_B a run of the TO2DI algorithm resulted in refined distances. In turn, these distances were used to obtain better estimates for S and τ_c using the procedure

outlined above. With these values the TO2DI algorithm was applied again to obtain further refined estimates for the distances etc. This procedure, which is summarized in Figure 6, is not a completely model-free one. First, standard B-DNA inter-proton distances are used to obtain rough initial estimates for S and τ_c. Secondly, because the overlap of nucleotide reson-ances with each other, or with protein resonances, precludes an accurate determination of TrNOE volumes, in many cases standard interproton distances were introduced in the relaxation matrix. In most cases this concerns TrNOEs between the H2', H2" H3', H4', H5' and H5" sugar ring protons. Most of these distances are almost invariant over the conformational space available for the deoxyribose ring, when the pseudorotation wheel is transversed and torsion angle γ is varied (Wijmenga et al., 1992). Therefore, in the TrNOE simulations, the set of invariant interproton distances listed in Table III will account for the major part of the spindiffusion within the sugar ring. Only the $d_i(2";H4')$ and $d_i(3';5')$ distances vary, to some extent, as a function of sugar ring conformation (Wüthrich, 1986). In the simulations, substitution of the smallest and largest values for $d_i(2";H4')$ and $d_i(3';5')$ into the distance matrix did not yield significantly different results.

Figure 6. Representation of the algorithm used for the calculation of S, τ_c and the distance matrix D. First, values for S and τ_c are searched for that give the best fit of a set of distances to experimental crosspeak and diagonal volumes (TO2ST, see text). In the second stage, an adapted NO2DI algorithm (TO2DI, see text) performs a further refinement of the distances.
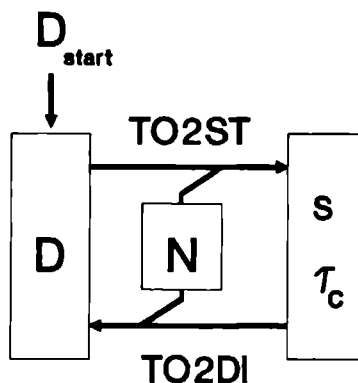
In Table IV the results are presented of the application of the protocol, outlined above, to the transferred NOESY spectra of TGCAACGA and TCGTTGCA. Most intra-nucleotide $d_i(8,6;2')$ and $d_i(8,6;2")$ distances point to B-DNA type anti-conformations ($\chi \approx 240°$) for the bases in the bound oligonucleotide. The inter-nucleotide $d_s(8,6;2")$ distances, however, are considerably longer than those reported for a typical B-DNA type double helix (≈ 2.2 Å,

Table IV

Intra- and inter-nucleotide d(8,6;1') and d(8,6;2") distances [Å] as determined by the TO2DI algorithm from the peak-volumes of transferred NOEs. The inter-nucleotide $d_s(8,6;2")$ distances are placed in brackets. Distances d_{ij} for which the corresponding TrNOE $N_{ij}(\tau_m)$ has a predominant (> 75%, see text) contribution of spindiffusion are not included in the Tables. As criteria for appropriate choices for S, and τ_c, the ratio (C/P) has been given between calculated (TO2DI) and physical (invariant) values for the distances $d_t(1';2")$ and $d_t(5;6)$. Intra- and inter-nucleotide TrNOEs that could not be quantified accurately due to overlap were accounted for by insertion of standard (B-DNA, Table III) and "infinitely long" distances (5 Å) respectively.

(A) TGCAACGA (N = 11.4 mM, P/N = 0.055, τc = 15 ns, τ_m = 75 ms)

i	$d_i(8,6;1')$	$d_i(8,6;H2")$	$d_s(8,6;H2")$	C/P
T1		---		0.99[1]
G2	3.2	---		

C3		2.2		1.00[2]
			3.1	
A4	2.8	2.0		

A5	3.0	---		
			2.9	
C6	3.0	2.2		
G7	3.5	---		0.97[1], 1.01[2]
			3.1	
A8	3.2	2.5		

(B) TCGTTGCA (N=11.4 mM, P/N = 0.055, τc = 14 ns, τ_m = 75 ms)

i	$d_i(8,6;H1')$	$d_i(8,6;H2")$	$d_s(8,6;H2")$	C/P
T1	2.9	2.4		1.03[1]
C2	---	2.4		0.90[1], 1.18[2]
G3	---	---		
T4	2.5	2.2		0.90[1]
T5	2.9	2.3		
G6	---	2.3		1.03[1]
			3.3	
C7	3.2	2.3		1.11[2]
			3.5	
A8	3.1	2.4		

[1]   *Ratio between calculated and known (nearly invariant) value (2,36 Å) for $d_t(1';H2")$*

[2]   *Idem for $d_t(5;6)$ (2.4 Å).*

(Wüthrich, 1986) but are still indicative of a helical conformation for certain parts of the oligonucleotide backbone. The intensities of the $d_s(8,6;2')$ and $d_s(8,6;1')$ inter-nucleotide contacts were for a major part determined by spindiffusion (via the H8→H2'→H2"→H1' pathway). For this reason the corresponding distances could not be determined accurately enough and were therefore not included in Table IV.

Due to extensive spectral overlap, the TrNOE spectra of D–(dA)$_6$ and L–(dA)$_6$ did not lend themselves to quantitative calculations. Both spectra, however, yielded (within experimental error) identical sets of (overlapping) TrNOE volumes. The TrNOE intensities at short mixing time indicates that the adenyl bases adopt anti-conformations except for the A1 base which is in a syn-conformation. This is in good agreement with previous NMR studies on oligodeoxyadenylic acids bound to M13 GVP (Alma et al., 1983a; King & Coleman, 1987).

Figure 7 shows the calculated build-up curve of the TrNOEs based on the values for p_B, τ_c and the interproton-distances listed in Table IV. At prolonged mixing times the TrNOEs calculated for the terminal nucleotide A8 still rather well match with the experimental ones. This is in support of the validity of the underlying assumption of our model, i.e. fast exchange on the longitudinal relaxation scale. For the spin systems of the C3 and C6 residues the correspondence between model-predicted and experimental values becomes rather poor. Typically, the abberation tends to be larger for the sugar-sugar than for the base-base TrNOEs. Most likely spinpools with different mobility and/or external leakage pathways exist that are not accounted for by the model (which assumes an isotropic mixing time and external spin-lattice relaxation via the free form). It is well established that differences in mobility or external leakage pathways become apparent at prolonged mixing times. The most likely candidates for these effects are the restricted mobility of the sugar ring spin system or leakage of its magnetization to the protein via cross-relaxation. In either case this points to an interaction of the oligonucleotides with the M13 GVP surface via the sugar-phosphate backbone. It is noted that intermolecular crosspeaks between oligonucleotide and protein can only be observed at mixing times that are much longer than the ones that were employed to generate the intramolecular transferred NOEs.
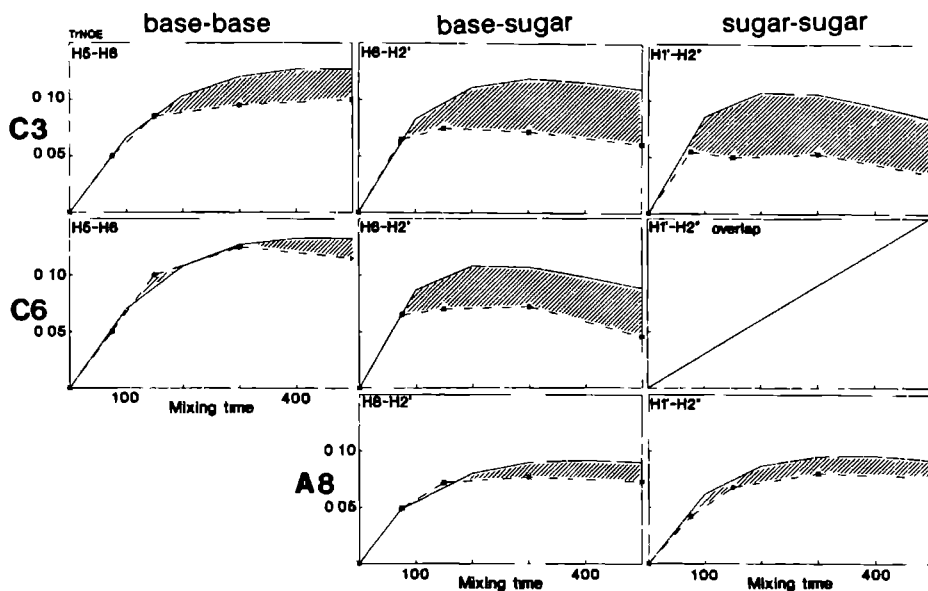
Figure 7. Calculated (————) and experimental (- - -) build-up curves for intra-nucleotide base-base (H5-H6), sugar-base (H8/6-H2') and sugar-sugar (H1'-H2") TrNOEs, observed for TGCAACGA bound to M13 GVP. In the calculations the parameters (τ_c, p_B and interproton distances) listed in Table IV were used. The shaded areas correspond to discrepancies between experimental and model-predicted TrNOE intensities.

DISCUSSION

The oligonucleotide-GVP interaction. Both the NMR and CD data provide evidence for base-destacking and concomitant disruption of the single-stranded helical structures of the $(dA)_6$ enantiomers upon complexation to M13 GVP. The differences observed between the NMR spectra of bound L–$(dA)_6$ and D–$(dA)_6$ are rather small which suggests that these oligonucleotides both adopt similar extended 'helical' structures. This is well in line with the CD data of the bound $(dA)_6$ enantiomers. The differential effects visualized in the long wavelength CD sum-spectra seem larger, however, than those observed by NMR. A possibility is that tyrosyl chromophores become optically active upon binding of DNA fragments. It should also be noted that in the CD and NMR experiments different physical quantities are observed, (chromophore optical activity and chemical shift tensor respectively)

in which some structural features may be weighted stronger than others. We conclude that the interactions of the $(dA)_6$ enantiomers with M13 GVP share a common feature (i.e. base-destacking and disruption of the single-stranded helix) which does not discriminate between the different chirality of the sugar rings. This may be related to the non-specific character of the ssDNA-GVP interaction. This is in contrast to the enantio-selectivity of many other protein-ligand interactions, e.g. the insensitivity of L-$(dA)_6$ towards snake venom phophodiesterase (Tazawa et al., 1970).

The C^εH-Y26 resonance, which can be used as a monitor for the complexation of M13 GVP with DNA fragments (Alma et al., 1982; King & Coleman, 1987), exhibits similar shifts upon complexation of the $(dA)_6$ enantiomers with GVP. It has been suggested that the strong upfield shifts of C^εH-Y26, brought about by binding of DNA fragments, can be attributed to ring current effects from the bound nucleotide bases. Although the $(dA)6$ enantiomers differ in chirality of their sugar-phosphate backbones they are both able to induce C^εH-Y26 shifts of comparable size. This can be brought in agreement with a large degree of flexibility in the oligonucleotide-GVP complex in which the average conformation of the enantiomers induce similar ring current effects on the C^εH-Y26 spin.

The TrNOE data obtained at longer mixing times for the TGCAACGA-GVP complex can also be explained in terms of internal mobility (both at the intra- and inter-nucleotide level). The dynamic nature of the oligonucleotide-GVP interaction is also reflected in the long mixing or irradiation times and the experimental conditions that other authors found necessary to observe intermolecular protein-nucleotide NOEs of significant intensity (Alma et al., 1982; King & Coleman, 1987). This indicates that in the complex, the protons of GVP and oligonucleotide can be considered as spinpools with uncorrelated internal motions. Even in the GVP-$(dA)_{40-60}$ assembly the NMR data are indicative of a considerable degree of motional freedom within the complex (King & Coleman, 1988).

The intramolecular distances determined for the TCGTTGCA and TGCAACGA fragments indicate that in the bound form these oligonucleotides predominantly adopt extended structures with partially destacked bases. For the pyrimidine-purine and pyrimidine-pyrimidine sequences (CG, TG, TC), however, the inter-nucleotide ds(6,2") distances are often stretched to such an extent that the corresponding TrNOEs could not or hardly be observed. It is well known that these dinucleotides form poor base-stacks in B-DNA type structures

(Saenger, 1984). This suggests that intramolecular neighbour-neighbour interactions determine the structure of the bound oligonucleotide.

From these data an oligonucleotide-GVP binding model emerges in which the electrostatic interaction between electropositive patches on the protein surface (Chapter 6) and the negatively charged sugar-phosphate backbone drives the conformational changes (base destacking and helix extension) in the oligonucleotide chain. Besides by electrostatic interactions with GVP, the conformation of the bound oligonucleotide is also determined by intramolecular neighbour-neighbour base-stacking interactions. In this interaction model non-electrostatic intermolecular interactions (e.g. between aromatic rings of GVP and the nucleotide units) are of a highly dynamic nature and play a minor role in oligonucleotide-GVP complex stabilization. This interaction model permits binding of an enantio-oligonucleotide with a binding affinity that is within the same order of magnitude as for the native DNA fragment.

This type of interaction may be resemblant to the dsDNA binding that occurs under specific experimental conditions (Sang & Gray, 1987). In the same terms we can explain the mimicking of the linebroadening effects induced upon binding of spin-labeled oligonucleotides by the small, negatively charged, inorganic ligand $Gd(DOTP)^{5-}$ (Chapter 3).

The 'apparent sequence-specificity' of the oligonucleotide-GVP model is in line with the data on the characterizing binding properties of polynucleotides. It is well known that the base-composition of polynucleotide chains determines their binding affinity to GVP (e.g. poly(dA) binds more strongly than poly(dT) and poly(dC)) (Alma et al., 1983b; Bulsink et al., 1985; Kansy et al., 1986; de Jong et al., 1987a). This has been explained by the difference in (de–)stacking energy of purine and pyrimidine bases. In same terms an 'apparent sequence-specificity' has been explained for the binding of M13 GVP to hetero-polynucleotides (Sang & Gray, 1989). The present results demonstrate, at the molecular level, that in the bound oligonucleotide some base-sequences indeed adopt more destacked conformations than others.

Physiological relevance. The rather loose oligonucleotide-GVP interaction might bear relevance towards sliding or translocation mechanisms operative in the partially occupied polynucleotide lattices. Such mechanisms have shown to be indispensable for a rapid formation of nucleoprotein clusters (Shimamoto & Utiyama, 1983; Pörschke & Rauh, 1983;

Bulsink et al., 1988a). The non-specific electrostatic interaction between the sugar-phosphate backbone and the electropositive protein surfaces and the absence of a firm interaction of the nucleotide bases with aromatic rings on the protein surface (as found in the n=3 oligonucleotide mode) allows a rapid sliding of GVP along the ssDNA lattice. Upon formation of GVP clusters on the DNA lattice, more specific (e.g. non-electrostatic) interactions take place which may be responsible for the difference in n=3 (oligonucleotide) and n=4 (polynucleotide) mode. Such two-stage interaction mechanisms are commonly employed by sequence-specific dsDNA binding proteins where they facilitate the search for the base-sequence targets on the dsDNA double helix (Saenger, 1984).

The recent completion of the assignment of the ^{1}H NMR spectrum of M13 GVP has opened the prospect of its tertiary structure elucidation, an enterprise that is currently in progress. The present quantitative structural data will be of use in future model building studies of the ssDNA-GVP complex. It is noted that the present results pertain to the oligo-nucleotide-GVP complex. In the nucleoprotein assembly some aspects of the ssDNA conformation may be different (vide supra).

References

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

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

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

Alma, N.C.M., Harmsen, B.J.M., van Boom, J.H., van der Marel, G. & Hilbers, C.W. (1983a) *Biochemistry 22*, 2104-2115.

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

Altona, V. & Sundaralingham, M. (1972) *J. Ac. Chem. Soc. 94*, 8205-8212.

van Amerongen, H., Kwa, S.L.S. & van Grondelle, R. (1990) *J. Mol. Biol. 216*, 717-727.

Asseline, U., Hau, J-.F. Czernecki, S., Le Diguarher, T., Perlat, M-.C., Valery, J-.M. & Thuong, N.T. (1991) *Nucl. Ac. Res. 19*, 4067-4074.

Balaram, P., Bothner-By, A.A. & Dadok, J., (1972) *J. Am. Chem. Soc. 94*, 4015-4017.

Bax, A. (1984) *Nuclear Magnetic Resonance in Liquids*, Delft University Press, Delft.

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.

Berliner, L.J. (1976), *Spin labeling theory and applications*, Academic Press, London.

Bevington, P.R. (1969), *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill Company, New York.

van Boom, J.H., van der Marel, G.A., Westerink, H.P., van Boeckel, C.A.A., Mellema, J.-R., Altona, C., Hilbers, C.W., Haasnoot, C.A.G., de Bruin, S.H. & Berendsen, R.G. (1983) *Cold Spring Harbor Symp. Quant. Biol. 47*, 403-409.

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

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

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

Bulsink, H., Harmsen, B.J.M. & Hilbers, C.W. (1985), *J. Biomol. Struct. Dyn.* **3**, 227-247.

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

Bulsink, H., Harmsen, B.J.M. & Hilbers, C.W. (1988a) *Eur. J. Biochem.* **176**, 589-596.

Bulsink, H., Harmsen, B.J.M. & Hilbers, C.W. (1988b) *Eur. J. Biochem.* **176**, 597-608.

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

Campbell, A.P. & Sykes, B.D. (1991) *J. Magn. Reson.* **93**, 77-92.

Cavanagh, J., Chazin, W.J. & Rance, M. (1990) *J. Magn. Reson.* **87**, 110-131.

Chase, J.W., Merril, B.M. & Williams, K.R. (1983) *Proc. Nat. Acad. Sci USA* **80**, 5480-5484.

Churchill, M.E.A. & Travers, A.A. (1991) *TIBS* **16**, 92-97.

Clore, G.M. & Grorenborn, A. (1982) *J. Magn. Reson.* **48**, 402-417.

Clore, G.M. & Gronenborn, A.M., (1983) *J. Magn. Reson.* **53**, 423.

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

Cuypers, T., van der Ouderaa, F.J. & de Jong, W.W. (1974) *Biochem. Biophys. Res. Commun.* **59**, 557-563.

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

Day. L.A. (1988) *Ann. Rev. Bioph. Bioph. Chem.* **17**, 509-539.

Dick, L.R., Sherry, A.D., Newkirk, M.M. & Gray, D.M. (1988), *J. Biol. Chem.* **263**, 18864-18872.

Dick, L.R., Geraldes, F.G.C., Sherry, A.D., Gray, C.W. & Gray, D.M. (1989), *Biochemistry* **28**, 7896-7904.

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

van Duynhoven, J.P.M., Folkers, P.J.M., Prinse, C.W. Harmsen, B.J.M., Konings, R.N.H., Hilbers, C.W. (1992) *Biochemistry* **31**, 1254-1262.

Duynhoven, J.P.M., Folkers, P.J.M., Nooren, I.M.A., Swinkels, D.W., Harmsen, B.J.M., Konings, R.N.H., Tesser, G.I. & Hilbers, C.W. (1992) *submitted*.

Dwek, R.A. (1973), *NMR in Biochemistry*, Clarendon Press, Oxford.

Eley, C.G., Moore, G.R., Williams, G. & William, R.J.P. (1982) *Eur. J. Biochem.* **124**, 295-303.

Ernst, R.R., Bodenhausen, G. & Wokaun, A. (1987) *Principles of Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford.

Falaschi, A. Cobianchi, F. & Riva, S. (1980) *TIBS 5*, 154-157.

Feeney, J., Batchelor, J.G., Albrand, J.P. & Roberts, G.C.K. (1979) *J. Magn. Reson. 33*, 519-529.

Folkers, P.J.M., Stassen, A.P.M., van Duynhoven, J.P.M., Harmsen, B.J.M., Konings, R.N.H. & Hilbers, C.W. (1991a) *Eur. J. Biochem. 200*, 139-148.

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

Folkers, P.J.M., van Duynhoven, J.P.M., van Lieshout, E., Harmsen, B.J.M., Konings, R.N.H. & Hilbers, C.W. (1992) *in preparation*.

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

Fujimori, S., Shudo, K. & Hashimoto, Y. (1990) *J. Am Chem. Soc. 112*, 7436-7438.

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

Garssen, G.J., Kaptein, R., Schoenmakers, J.G.G. & Hilbers C.W. (1978) *Proc. Natl. Acad. Sci. USA 75*, 5281-5285.

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

Geraldes, C.F.G.C., Brown, R.D., Cacheris, W.P., Koenig, S.H., Sherry, A.D. & Spiller, M. (1989) *Magn. Reson. Med. 9*, 94-104.

Gilbert, W. & Dressler, D. (1968) *Cold Spring Harbor Symp. Quant. Biol. 43*, 473-484.

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

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

Gray, C.W., Kneale, G.G., Leonard, K.R., Siegrist, H. & Marvin, D.A. (1982b) *Virology 116*, 40-52.

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

Gray, C.W. (1985) *Fourth Conversation in Biomolecular Stereodynamics*, SUNY at Albany, Abstracts, 219-220.

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

Hofschneider, P.H. & Preuss, A. (1963) *J. Mol. Biol. 7*, 450-451.

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. 714*, 4546-4553.

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 (1987b) *Eur. J. Biochem. 167*, 563-572.

de Jong, E.A.M., Claesen, C.A.A., Daemen, C.J.M., Harmsen, B.J.M., Konings, R.N.H., Tesser, G.I. & 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. (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.

Kansy, J.W., Clack, B.A. & Gray, D.M. (1986) *J. Biomol. Struc. 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.

Kühne, R.D., Schaffhauser, T., Wokaun, A. & Ernst, R.R. (1979) *J. Magn. Reson. 35*, 39.

Landy, S.B. & Rao, B.D.N. (1987) *J. Magn. Reson. 81*, 371.

Lippens, G.M., Cerf, C. & Hallenga, K. (1992) *J. Magn. Reson. in press*.

Lohman, T.M., Bujalowski, W. & Overman, L.B. (1988) *TIBS 13*, 250-255.

Luiten, R.G.M., Putterman, D.G. Schoenmakers, J.G.G. Konings, R.N.H. & Day, L.A. (1985) *J. Virol.* **56**, 268-276.

Macura, S., Ernst, R.R. (1980) *Mol. Phys.* **41**, 95-117.

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

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

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

van Mierlo C.P.M. & Wijmenga, S.S. (1991) *Eur. J. Biochem.* **195**, 807-822.

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

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

Mooren, M.M.W., Hilbers, C.W., van der Marel, G.A., van Boom, J.H. & Wijmenga, S.S. (1991) *J. Magn. Reson.* **94**, 101-111.

Mooren, M.M.W., Egberts, E.M.D.T., van Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1992) *in preparation.*

Morris, G.A. & Freeman, R. (1978) *J. Magn. Reson.* **29**, 433-462.

Newport, J.W., Lonberg, N., Kowalkzykowski, S.C. & von Hippel, P.H. (1981) *J. Mol. Biol.* **145**, 105-121.

Noggle, J.H. & Schirmer, R.E. (1971) *The Nuclear Overhauser effect: Chemical applications,* Academic Press, New York.

O'Connor, T.P.O. & Coleman, J.E. (1982) *Biochemistry* **21**, 848-854.

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

Olsthoorn, C.S.M., Haasnoot, C.A.G. & Altona, C. (1980) *Eur. J. Biochem.* **106**, 85-95.

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

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

Plyte, S.E. & Kneale, G.G. (1991) *Protein Engineering* **4**, 553-560.

Pörschke, D. & Rauh, H. (1983) *Biochemistry* **22**, 5869-5878.

Prasad, B.V.V. & Chiu, W. (1987) *J. Mol. Biol. 193*, 579-584.

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

Press, W.H., Flanney, B.P., Teukolsky, S.A. & Vettingly, W.T. (1989) *Numerical Recipes*, pp 292-293, Cambridge University Press, Cambridge.

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

Rajavashisht, T.B., Taylor, A.K., Andalibi, A., Svenson, K.L. & Lusis, A.J. (1989) *Science 245*, 640.

Rinkel, L.J. & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621-649.

Savitzky, A. & Golay, M.J.E. (1964) *Analytical Chemistry 38*, 1627-1639.

Sancar, A., Stachelek, C., Koningsberg, W. & Rupp, W.D. (1980) *Proc. Nat. Acad. Sci. USA 78*, 2611-2615.

Sancar, A., Williams, K.R., Chase, J.W. & Rupp, W.D. (1981) *Proc. Nat. Acad. Sci. USA 78*, 4274-4278.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Sang, B.-C. & Gray, D.M. (1989) *J. Biomol. Struct. Dyn. 7*, 693-706.

Schwabe, J.W.R. & Rhodes, D. (1991) *TIBS 16*, 291-296.

Shimamoto, N. & Utiyama, H. (1983) *Biochemistry 22*, 5869-5878.

Sherry, A.D., Malloy, C.R., Jeffrey, F.M.H.,Cacheris, W.P. & Geraldes, C. F.G.C. (1988) *J. Magn. Reson. 76*, 528-533.

Stassen, A.P.M., Zaman, G.J.R., van Deursen, J.M.A., Schoenmakers, J.G.G. & Konings, R.N.H. (1992) *Eur. J. Biochem. 204*, 1003-1014.

Summers, M.F. (1991), *J. Cell. Bioch. 45*, 41-48.

Swinkels, D.W., van Duynhoven, J.P.M., Hilbers, C.W. & Tesser, G.I. (1991) *Recl. Trav. Chim. Pays-Bas 110*, 124-128.

Tazawa, I., Tazawa, S., Stempel, L.M. & Ts'o, P.O.P. (1970) *Biochemistry 9*, 3499-3513.

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

Urata, H., Shinohara, K., Ogura, E., Ueda, Y. & Kagi, M. (1991) *J. Am. Chem. Soc. 113*, 8174-8175.

van de Ven, F.J.M. & Hilbers, C.W. (1988) *Eur. J. Biochem. 178*, 1-38.

van de Ven, F.J.M., Blommers, M.J.J., Schouten, R.E. & Hilbers, C.W. (1991) *J. Magn. Reson. 94*, 140-151.

Vroom, E. (1988) *Ph. D. Thesis*, State University Leiden, The Netherlands.

Vuister, G.W. Boelens, R., Padilla, A., Kleywegt, G. & Kaptein, R. (1990) *Biochemistry 29*, 1829-1839.

Waugh, J.S. (1986) *J. Magn. Reson. 68*, 189-192.

Wijmenga, S.S., Mooren, M.M.W. & Hilbers, C.W. (1992) *Practical Approach Series*, Oxford University Press, in press.

Williams, K.R., LoPresti, M.B. & Setogushi, M. (1981) *J. Biol. Chem. 256*, 1754-1762.

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

Wüthrich, K. (1990) *J. Biol Chem. 265*, 22059-22062.

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

von Hippel, P.H. & Berg, O.G. (1989) *Protein Nucleic Acid Interactions*, (Saenger, W., Ed.), MacMillan, London.

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

Zaman, G., Smetsers, A., Kaan, A., Schoenmakers, J. & Konings, R. (1991) *Biochim. Biophys. Acta 1089*, 183-192.

Zaman G. (1991) *Ph.D. Thesis*, University of Nijmegen, Nijmegen

Zabin, H.B., Terwilliger, T.C. (1991) *J. Mol. Biol. 219*, 257-275.

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

Summary

In the living cell, protein-DNA interactions play important roles in the genetic process. As is discussed in Chapter 1, most of our current knowledge in this field pertains to the double-stranded DNA binding proteins. Though the DNA binding proteins of this type may differ significantly in function, only a limited number of structural elements, so-called DNA binding motifs, are employed in the direct interactions with the DNA duplex. In the area of the single-stranded DNA binding proteins our knowledge is rather limited. The work described in this thesis aimed at gaining insight in this type of DNA-protein interaction at the sub-molecular level. For this purpose the proteins encoded by gene V of the filamentous bacteriophages M13 and IKe (GVP) were employed as model systems for the single-stranded DNA protein interaction. Most of the structural information on these proteins described in this thesis stems from nuclear magnetic resonance (NMR) studies.

In Chapter 2 two NMR techniques, the transferred NOESY (TrNOESY) and the TOCSY experiment, are treated theoretically. Especially those aspects are discussed that are relevant towards a quantitative interpretation of peak-intensities in the two-dimensional spectrum. A cyclic dinucleotide was used to demonstrate the feasibility of the quantitative interpretation of TOCSY data. Some of the theoretical considerations discussed for the TrNOESY experiment have found practical applications in Chapter 7.

Chapter 3 describes some of the most relevant characteristics of mutant and wild type M13 GVP for NMR experiments. The majority of these proteins has poor solubility characteristics (i.e. a strong tendency to form aggregates). This seriously hampers the recording of good quality NMR spectra. However, the mutant protein in which tyrosine-41 was replaced by a histidyl residue, distinguished itself by its exceptionally low tendency to form aggregates. In addition, optimization of the solution conditions led to spectra of reasonable quality for the wild-type en most of the other mutant M13 GVPs.

These results enabled the structure elucidation of a β-loop structure in Y41 GVP (Chapter 3). Evidence could be provided that an identical β-loop was present in M13 GVP. In both cases this structural element was intimately involved in DNA binding. In earlier work, an analogous DNA binding β-loop structure was found in IKe GVP. In both β-loops (commonly referred to as DNA binding loops) especially those amino acid residues that are involved in the interaction with the bound DNA are conserved to a high degree. These

findings are in support of an earlier proposal of this type of β-loop as a structural motif in single-stranded DNA binding proteins.

The nearly complete ^{1}H NMR spectral assignment and secondary structure elucidation of IKe GVP is described in Chapter 5. At the level of the secondary structure the folds of M13 and IKe GVP are nearly identical. Evidence exists that also the tertiary folding of these proteins is similar. Comparison of the secondary structure obtained by NMR for M13 GVP, with the crystal structure proposed by Brayer & McPherson (1983) showed several local, but significant, differences: The striking similarity of the NMR structures of M13 and IKe GVP is in support of the validity of these structures in solution.

Chapter 6 describes the mapping of the DNA binding domain of IKe GVP, using the spin-labeled oligonucleotide *(dA)$_3$*. The specific linebroadening effects induced upon binding of this ligand were visualized by 2D-difference spectroscopy. Besides the aforementioned DNA binding loop, also other β-sheets could be identified that were in close proximity to the bound oligonucleotide. At neutral pH the negatively charged Gd(DOTP)$^{5-}$ induced similar linebroadening effects as *(dA)$_3$*. From the perturbation patterns induced by these paramagnetic ligands in the ^{1}H NMR spectra of M13 and IKe GVP the picture emerges of a DNA binding domain consisting of electropositive β-sheet elements, interspersed with hydrophobic residues. The perturbation pattern induced by Gd(DOTP)$^{5-}$ at low pH indicated that at least one intramolecular salt-bridge is present in the β-sheet region comprising the N-terminal segment.

In the last Chapter (Chapter 7) conformational and dynamical aspects of the oligonucleotide-GVP interaction were studied by means of the transferred NOE-effect. The conformation of oligonucleotides in the complex with GVP is mainly dictated by hydrophobic (stacking) interactions between neighbouring (nucleotide) bases. In the complex with GVP the bound oligonucleotide has a considerable degree of mobility. This type of interaction might be relevant for rapid sliding of the GVP molecule along extended polynucleotide lattices. This sliding mechanism has been inferred from other types of experiments and has been shown to be essential for a rapid formation of nucleoprotein assemblies.

Samenvatting

Interacties tussen DNA en eiwitten spelen een belangrijke rol in tal van moleculair biologische processen in levende organismen. Zoals beschreven in Hoofdstuk 1, is met name op het gebied van de dubbelstrengs-DNA bindende eiwitten al veel inzicht verkregen. Hoewel de DNA bindende eiwitten van dit type onderling sterk verschillen in functie en werkingsmechanisme blijken steeds weer dezelfde struktuurelementen, zogenaamde DNA bindings motieven, gebruikt te worden. Op het gebied van de enkelstrengs-DNA bindende eiwitten reikt onze kennis minder ver. Het in dit proefschrift beschreven onderzoek beoogde het inzicht in dit type DNA-eiwit interactie te verruimen. De enkelstrengs-DNA bindende eiwitten welke gecodeerd worden door gen V van de bacteriofagen M13 en IKe (GVP), werden hiervoor als modelsystemen gebruikt. Voor het onderzoek van de struktuur en struktuur-functie relatie van deze eiwitten werd met name gebruik gemaakt van kernspinresonantie (angelsaksische afkorting: NMR). Met name de multi-dimensionale NMR technieken zijn uitermate geschikt voor het in oplossing bestuderen van de struktuur biomacromoleculen zoals eiwitten en nucleinezuren.

In Hoofdstuk 2 worden twee van deze technieken, het TrNOESY en het TOCSY experiment, nader uitgediept. Met name die aspecten komen aan de orde die van belang zijn voor een kwantitatieve interpretatie van piekintensiteiten in deze twee-dimensionale NMR spectra. Voor het TOCSY experiment wordt een en ander toegelicht aan de hand van een cyclisch dinucleotide modelsysteem. De theoretische beschrijving van het TrNOESY experiment vindt zijn toepassing in Hoofdstuk 7.

In Hoofdstuk 3 worden wild-type en mutant M13 gen V eiwitten gekarakteriseerd met behulp van NMR. De meeste van deze eiwitten bezitten een sterke neiging tot aggregeren. Met name met betrekking tot NMR experimenten is dit een zeer ongunstige eigenschap. Het mutant eiwit, waarin tyrosine-41 gemuteerd is in een histidine residu (M13 GVP Y41H), onderscheidde zich echter in zeer gunstige zin van het wild-type eiwit. Onder sommige condities blijkt echter ook de aggregatie van wild-type M13 GVP tot aanvaardbare proporties te kunnen worden teruggebracht waardoor zinvolle NMR-metingen mogelijk worden.

Deze resultaten gaven de aanzet tot de struktuuropheldering van een zogenaamde β-lus in M13 GVP Y41H (Hoofdstuk 4). Aangetoond kon worden dat deze β-lus ook aanwezig is in het M13 GVP. Dit struktuurelement bleek nauw betrokken te zijn bij de DNA binding.

Reeds eerder was een dergelijk struktuurelement, met dezelfde DNA-bindende eigenschappen, aangetroffen in IKe GVP. In beide β-lussen ("DNA binding loops") zijn met name die aminozuren geconserveerd die sterk bij de DNA-eiwit interactie betrokken zijn. Deze bevindingen ondersteunen een eerder gepostuleerde hypothese dat dit type β-loop een terugkerend motief is in de enkelstrengs-DNA bindende eiwitten.

In Hoofdstuk 5 wordt de voltooiing beschreven van de toekenning van het ^{1}H NMR spectrum van IKe GVP en de opheldering van de secundaire structuur in oplossing. De secundaire vouwing van IKe GVP bleek vrijwel identiek te zijn aan die van M13 GVP. Er bestaan sterke aanwijzingen dat deze gelijkenis zich uitstrekt tot de tertiaire struktuur (de globale vouwing) van deze eiwitten. In voorgaande studies bracht vergelijking van de NMR-struktuur van M13 GVP met een eerder door Brayer en McPherson (1983) gepostuleerde kristalstruktuur enkele lokale, maar niet onbelangrijke, verschillen aan het licht. De treffende gelijkenis van de NMR-structuren van M13 en IKe GVP ondersteunt de juistheid van deze strukturen in oplossing.

De volledige toekenning van de ^{1}H NMR spectra van M13 en IKe GVP maakte het mogelijk om de interactie van deze eiwitten met DNA in detail te bestuderen. Hoofdstuk 6 beschrijft het in kaart brengen van de DNA-bindende domeinen van M13 en IKe GVP met het gespinlabelde DNA fragment *(dA)$_3$*. De specifieke lijnverbredingen van de NMR resonanties van de aminozuren in het DNA-bindingsdomein van IKe GVP werden zichtbaar gemaakt met 2D-verschilspectroscopie. Behalve de reeds eerder beschreven "DNA binding loops" konden in M13 en IKe GVP nog andere β-sheet structuren aangewezen worden die deel uitmaken van het DNA-bindende domein. Bij neutrale pH bleek het sterk negatief geladen paramagnetische Gd(DOTP)$^{5-}$ complex, globaal gezien dezelfde resonanties te verbreden als het gespinlabelde DNA fragment. Vergelijking van de uitdovingspatronen van *(dA)$_3$* en Gd(DOTP)$^{5-}$ in M13 en IKe GVP levert het beeld op van een DNA bindend domein bestaande uit electropositieve β-sheet structuren waarin geconserveerde hydrofobe aminozuren verspreid voorkomen. Uit het uitdovingspatroon bij lagere pH bleek dat zich in de regio nabij de N-terminale segmenten van M13 en IKe GVP zich minstens één intra-moleculaire zoutbrug moet bevinden.

Het laatste Hoofdstuk (Hoofdstuk 7) beschrijft de conformatie en dynamica van oligonucleotiden (korte DNA fragmenten) gebonden aan M13 GVP. Hierbij werd voor een

belangrijk deel gebruik gemaakt van het in Hoofdstuk 2 nader uitgewerkte TrNOESY experiment. De struktuur die een oligonucleotide aanneemt in het complex met M13 GVP lijkt met name bepaald te worden door hydrofobe (stacking) interacties tussen naburige nucleotide basen. Verder bezit het gebonden oligonucleotide een aanzienlijke bewegingsvrijheid. Dit dynamische oligonucleotide bindingsmodel zou van belang kunnen zin voor een snelle voortbeweging van een eiwitmolekuul langs een polynucleotide keten zoals dat reeds met andere technieken aangetoond is. Dergelijke "sliding' mechanismen zijn van essentieel belang voor een snelle en volledige bezetting van een DNA keten met GVP.

Curriculum Vitae

Johannes P.M. van Duynhoven werd geboren op 19 juli 1962 te Langenboom (N.-Br.). Na behalen van het Atheneum B diploma op het Merletcollege te Cuyk werd in 1980 aangevangen met de scheikunde studie aan de Katholieke Universiteit Nijmegen. Het kandidaatsexamen (S2) werd afgelegd in 1983. De doctoraalstudie ("oude stijl") omvatte de bijvakken Biochemie (Prof. dr. J.J.H.H.M. de Pont) en Analytische Chemie (Prof. drs. G. Kateman) en het hoofdvak Biofysische Chemie (Prof. dr. C.W. Hilbers). Daarnaast werd in 1984 het Diploma Stralingsdeskundigheid[III] behaald. Na het afleggen van het doctoraalexamen in juni 1986 was hij tot 1 juli 1991, in dienst van N.W.O., werkzaam als wetenschappelijk assistent ("promovendus oude stijl") op de afdeling Biofysische Chemie van de Katholieke Universiteit Nijmegen. Hier werd het in dit proefschrift beschreven onderzoek verricht.