Protein Chemistry-Nuclear Magnetic Resonance Approach to Mapping Functional Domains in Single-Stranded DNA Binding Proteins

Joseph E. Coleman, Kenneth R. Williams, Garry C. King, Richard V. Prigodich, Yousif Shamoo, and William H. Konigsberg

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510

DNA binding proteins that bind preferentially to single-stranded DNA with relatively little sequence specificity are widely distributed in both prokaryotic and eukaryotic cells. These proteins function in DNA replication and recombination and are synthesized in high enough copy number to suggest that they act in vivo by forming stoichiometric complexes with exposed single-stranded regions of DNA. Some proteins of this class also have significant affinity for single-stranded RNA and may thus influence translation as well. The single-stranded DNA binding proteins coded for by gene 32 of bacteriophage T4 and by gene 5 of the filamentous bacteriophage fd (M13) are the most extensively investigated examples of this class of proteins. The location of the DNA binding domains in the primary structure, the thermodynamics of DNA binding and the specific amino acid residues responsible for the nucleotide-protein interactions have been successfully determined for these proteins by a variety of physicochemical techniques, which include limited proteolysis to determine minimum DNA binding domains, chemical modification to detect specific types of amino acid side chain involved in DNA binding, and a number of one- and two-dimensional [1]H-nuclear magnetic resonance (NMR) techniques to detect protein-oligonucleotide interactions. A crystal structure is available for the unliganded gene 5 protein, facilitating the detailed study of these interactions. Recently, a program of site-directed mutagenesis has been untertaken to detect specific residues participating in DNA binding and to assign the ^{1}H-NMR spectra residue by residue. Examples of the information obtained by these methods and the synthesis of the findings into models for the binding of single-stranded DNA by these proteins are covered in this review.

Biological Roles of the Bacteriophage T4 Gene 32 and fd Gene 5 ssDNA Binding Proteins

The bacteriophage T4 gene 32 protein (g32P) has served as a prototype for a class of proteins that bind without sequence specificity to single-stranded nucleic acids. These proteins play a stoichiometric role in DNA replication, repair, and recombination. Consistent with g32P functioning in a stoichiometric rather than catalytic role is the finding that there are approximately 10,000 g32P molecules per infected cell or perhaps 125 molecules per replication fork [1]. Studies on condition-ally lethal mutations provide a striking demonstration of the importance of g32P in T4 DNA metabolism. Within 2 minutes after cells infected with T4 phage carrying the tsP7 mutation in gene 32 are shifted to a nonpermissive temperature, all T4 DNA replication ceases [2,3], and the normal 600–1,000-S intracellular T4 DNA is com-pletely converted to small DNA segments (34–80 S) that range from one-fourth to two genomes in length. Most, if not all, of the various functions ascribed to g32P in vivo have been attributed to its ability to bind tightly and cooperatively to single-stranded nucleic acids. Because g32P binds approximately 1,000 times more tightly to ssDNA than to dsDNA it is an effective equilibrium destabilizer of dsDNA [5]. This helix destabilization undoubtedly contributes to the 150–fold increase in the rate of DNA synthesis on a nicked dsDNA template that is seen upon the addition of g32P [6,7]. Helix destabilization per se cannot be the only function of 32P in DNA replication because the *Escherichia coli* SSB protein, which is also a strong helix-destabilizing protein, cannot substitute for g32P in these reactions [8,9]. This result suggests that g32p and SSB may either impose quite different conformations onto the ssDNA, or that g32P but not the *E coli* SSB protein may be able to interact with other proteins in the T4 DNA replication complex and that these protein:protein interactions may be an essential aspect of g32P function. Support for this latter idea derives both from genetic studies [10] and from the results of protein affinity chromatography using columns containing covalently bound g32P [11].

Besides being an ideal substrate for T4 DNA polymerase, the ssDNA in a g32P:ssDNA complex is resistant to attack by nucleases, yet it can readily base pair with homologous ssDNA. Gene 32P plays an essential role in DNA repair by protecting from nuclease attack those transiently single-stranded regions of DNA that occur normally in vivo and thus preventing lethal dsDNA cuts [12]. While most attention is directed toward its helix-destabilizing properties, g32P can also increase the rate of renaturation of strand-separated homologous ssDNA by up to 1,000-fold [1]. This "activity" of g32P could undoubtedly play an important role in recombina-tion. Whether the addition of g32P favors the denaturation or renaturation of dsDNA depends upon the relative stabilities of the base-paired dsDNA on the one hand versus the g32P:ssDNA complex on the other under the particular salt concentrations being studied.

While cooperative binding of g32P to ssDNA is essential for most of its functions in DNA metabolism, the ability of g32P to also bind ssRNA allows it to control its own rate of synthesis at the level of translation [13–15]. Based on extensive in vivo and in vitro data, a model for g32P autogenous regulation has been proposed [16–19], which requires that, as the protein is synthesized in vivo, it bind first to all ssDNA sequences that are present. At this point, the "free" g32P concentration rises until it reaches a particular threshold level (approximately 0.1 mg/ml) necessary for

g32P to bind specifically to its own mRNA, thus preventing further g32P synthesis. The 10^1 to 10^4 higher affinity (depending upon the particular homopolymer tested, see reference [20]) of g32P for ssDNA over ssRNA assures that all of the intracellular ssDNA will be saturated with g32P before binding occurs on mRNA. The specificity for g32P mRNA (as opposed to other T4 mRNAs) seems to result from the presence of a uniquely unstructured region that spans about 50 nucleotides near the presumed ribosome binding site of g32P mRNA [17,18]. It is important to realize that at the estimated in vivo concentration of free g32P (2–3 μm), all available ssDNA and ssRNA sequences should be saturated with g32P. In addition, g32P should be able to melt most adventitious stem-loop (hairpin) structures in ssDNA that are stabilized by less than ~5–7 base pairs, but, because of the lower affinity of g32P for ssRNA, similar structures in mRNA would be stable to g32P. Hence, in vivo g32P cannot denature any fully duplex DNA or RNA structures, nor would it be expected to be able to denature those hairpin-like structures essential for mRNA function [19].

The ability of g32P to stimulate rates of DNA replication both in vivo and in vitro sets it apart from the bacteriophage fd gene 5 protein, whose primary role is to prevent replicative DNA synthesis [21]. Thus g5P binds to the emerging single-stranded fd DNA as it is synthesized on the complementary strand of the nicked RFs and in so doing g5P prevents its use by host enzymes for synthesis of the complementary strand. Addition of gene 5 protein (in amounts stoichiometric to the DNA) to an in vitro reconstituted RF-synthesizing system halts DNA synthesis [21]. The ssDNA:g5P complex apparently migrates to the cytoplasmic surface of the cell membrane where the DNA passes through the membrane and the gene 5 protein is replaced by coat proteins. As in the case of g32P, g5P remains behind in the cytoplasm and is not found in significant quantities in the intact virions. Although g32P and g5P have different functions in vivo and in addition fail to share any significant amino acid sequence homology, they are similar in that both proteins bind preferentially to ssDNA and are required in stoichiometric amounts. It has been estimated that there are approximately 10^5 molecules of g5P per infected bacterium. In both protein:ssDNA complexes, the ssDNA is held in an expanded form so that the base-to-base distance is considerably larger than that found in fd ssDNA alone [22]. The physicochemical studies described below suggest that the similarity between these two ssDNA binding proteins extends to their mechanism of binding.

Limited proteolysis studies suggest that the T4 Gene 32 protein contains three functional domains. Based on the amino acid sequence of g32P (Fig. 1) and the nucleotide sequence of its gene [18,23,24], g32P contains 301 amino acids and has a molecular weight of 33,488. Although g32P should have a net charge of −10 at pH 7, which is in agreement with an experimentally determined isoelectric point of 5.0 [25], the charge distribution is quite asymmetric. While the NH_2-terminal half of g32P has a net charge of +10, the COOH-terminal half has a net charge of −20. The monomeric g32P behaves hydrodynamically as though it were a prolate ellipsoid with an axial ratio of 4:1 and an overall length approaching 12 nm [1]. Since in its complex with ssDNA each g32P molecule may only span a distance of about 3.3 nm (assuming a binding site size of 7) and an internucleotide spacing of about 0.47 nm [26], adjacent molecules of g32P probably overlap one another to some extent. This overlap presumably contributes to the cooperativity that is observed when g32P binds ssDNA [1,5,26]. Although g32P exists as a monomer in dilute solution (<0.5 μM [27]), it undergoes indefinite self association, so that at the estimated free
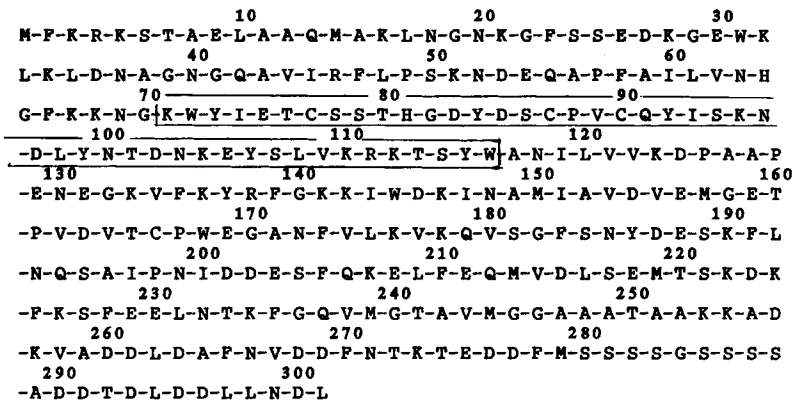
```
                    10                  20                  30
M-F-K-R-K-S-T-A-E-L-A-A-Q-M-A-K-L-N-G-N-K-G-F-S-S-E-D-K-G-E-W-K
                 40                  50                  60
L-K-L-D-N-A-G-N-G-Q-A-V-I-R-F-L-P-S-K-N-D-E-Q-A-P-F-A-I-L-V-N-H
       70 ————————————— 80 ————————————— 90 —————————
G-F-K-K-N-G|K-W-Y-I-E-T-C-S-S-T-H-G-D-Y-D-S-C-P-V-C-Q-Y-I-S-K-N
——————— 100 —————————————— 110 ———————————— 120
 -D-L-Y-N-T-D-N-K-E-Y-S-L-V-K-R-K-T-S-Y-W|A-N-I-L-V-V-K-D-P-A-A-P
      130              140              150              160
 -E-N-E-G-K-V-F-K-Y-R-F-G-K-K-I-W-D-K-I-N-A-M-I-A-V-D-V-E-M-G-E-T
              170              180              190
 -P-V-D-V-T-C-P-W-E-G-A-N-F-V-L-K-V-K-Q-V-S-G-F-S-N-Y-D-S-K-F-L
             200              210              220
 -N-Q-S-A-I-P-N-I-D-D-E-S-F-Q-K-E-L-F-E-Q-M-V-D-L-S-E-M-T-S-K-D-K
            230              240              250
 -F-K-S-F-E-E-L-N-T-K-F-G-Q-V-M-G-T-A-V-M-G-G-A-A-A-T-A-A-K-K-A-D
          260              270              280
 -K-V-A-D-D-L-D-A-F-N-V-D-D-F-N-T-K-T-E-D-D-F-M-S-S-S-S-G-S-S-S-S
          290              300
 -A-D-D-T-D-L-D-D-L-L-N-D-L
```

Fig. 1. Primary structure of gene 32 protein. The boxed sequence contains six of the eight tyrosyl residues in the protein and is believed to constitute a part of the DNA binding domain.

g32P concentration in vivo (~ 3.0 μM [19]) it would have an apparent molecular weight corresponding to that of a dimer or trimer [27]. It seems likely that the g32P:g32P interactions responsible for indefinite aggregation are related to those involved in cooperative binding of g32P to ssDNA.

The finding that limited proteolysis can be used to generate functionally active fragments of g32P [28–30] has facilitated more detailed structure/function studies. This approach has allowed three different functions of g32P (ie, those mediated by g32P:g32P, g32P:T4 replication protein, and g32P:ssDNA interactions) to be mapped onto specific regions of the g32P primary structure. These studies have taken advantage of the fact that the domains made up of residues 9–21 and 253–275 in g32P are particularly susceptible to cleavage by a wide variety of different proteinases [23,25,30–33]. Clearly these two regions are exposed on the surface of g32P. Cleavage at any point between residues 9 and 21 removes the basic NH$_2$-terminal "B" region and produces g32P-B. Cleavage at any point between residues 253 and 275 removes the acidic, COOH-terminal "A" region and produces g32P-A. Cleavage at both sites results in a 26,000-dalton fragment called g32P-(A+B). Since the in vitro properties of these individual cleavage products appear to be identical irrespective of the enzyme used in their preparation it would appear that the essential residues within the "A" and "B" domain must be contained within residues 276–301 and 1–9, respectively.

The COOH-terminal A region is essential for g32P:DNA replication protein interactions and for controlling the helix-destabilizing activity of g32P.
Removal of the acidic COOH-terminal "A" region to produce g32P-A has relatively little effect on the equilibrium binding properties of this protein [34,35]. Both g32P and g32P-A bind cooperatively to ssDNA [34,35] and as a result have binding constants of approximately 10^8–10^9 M^{-1} for poly(dT) compared to 10^6 M^{-1} for oligo d(pT)$_8$ [34]. In view of their similar affinities for ssDNA it is surprising that at least with native dsDNA, g32P-A is a considerably better helix-destabilizing protein than g32P. In 0.01 M NaCl, g32P-A lowers the T_M dsDNA from about 70°C to less than 1°C [36]. In contrast, g32P has no effect on the T_M of T4 dsDNA [5,25]. It appears that in the case of g32P there is a kinetic block that prevents the attainment of the expected equilibrium [5]. That is, ssDNA binding proteins such as g32P destabilize

dsDNA by preferentially binding to single-stranded regions that might occur normally as a result of "breathing" of the dsDNA. This mechanism differs from that used by helicases, which utilize energy derived from ATP hydrolysis to drive the helicase through the dsDNA, resulting in strand separation.

A "slow step" that could account for the g32P kinetic block might be the conformational change that has been shown to accompany cooperative binding. This conformational change can be most easily detected by limited proteolysis experiments, which demonstrate that cooperative binding of g32P to ssDNA significantly increases the rate of trypsin cleavage of the COOH-terminal "A" region and decreases the rate of removal of the NH₂-terminal "B" region [31]. Conceptually the "A" region has been pictured as an arm or flap of g32P that is thought to partially occlude the ssDNA binding site [35]. By removing this domain it may be possible to avoid an otherwise obligatory conformational change, and for this reason g32P-A (but not g32P) is able to nucleate cooperative binding on transient ssDNA regions. Both proteins can destabilize poly[d(A-T)] because this partially dsDNA contains relatively stable single-stranded regions that allow both g32P and g32P-A to effectively nucleate binding.

While the exact mechanism by which the COOH-terminal "A" region prevents g32P from denaturing dsDNA is not yet clear, the possible in vivo functional consequences of the presence of a g32P lacking the "A" region are readily apparent in an in vitro T4 DNA replication system. Studies with this system demonstrate that although g32P-A can substitute for g32P in leading-strand synthesis, there is complete absence of lagging-strand synthesis in the presence of g32P-A [9]. The absence of RNA primer extension (lagging-strand synthesis) appears to be due to destabilization of the 3'-hydroxy chain terminus by g32P-A. Although these in vitro studies demonstrate that the helix-destabilizing activity of g32P is greatly increased by removal of its acidic, COOH-terminal domain, it is unclear whether the primary in vivo function of this region is to provide a means of controlling or of merely limiting the ability of g32P to denature dsDNA. Since there is no evidence that any significant partial proteolysis of g32P occurs in vivo, it has been suggested that heterologous protein-protein interactions involving the COOH-terminal "A" region of g32P and other T4 replication proteins may result in an effect similar to that of the actual removal of this domain [9]. If this notion is correct, then the "A" region would enable other T4 replication proteins to specifically enhance the helix-invasion potential of g32P only when g32P is situated in front of an advancing replication fork [9]. Regardless of whether this control is actually exercised in vivo or not, it now seems clear that the "A" region is essential for g32P:protein interactions with the T4 DNA replication complex. At least two of the ten proteins that have so far been shown to bind g32P—that is, g61P, the T4 primase, and g43P, the T4 DNA polymerase—fail to interact with g32P-A [9]. These in vitro results suggest that at the very least, the COOH-terminal "A" region is essential for gp32:replication protein interactions and for limiting the helix-destabilizing ability of g32P.

This particular functional domain appears to be shared by at least three other functionally homologous proteins. The ssDNA binding proteins encoded by bacteriophage T7 [37,38], E coli SSB [39], and E coli F-plasmid [40] all contain a very acidic region at their COOH terminus that is more exposed to proteolysis when these proteins are bound to ssDNA. In each case, removal of this domain results in an increase in helix-destabilizing ability. Although it is not yet certain if eukaryotic high mobility group proteins (HMG's) are functionally homologous to g32P-like ssDNA

binding proteins, it is interesting that high molecular weight HMG proteins also seem to contain a domain analogous to the COOH-terminal "A" region of g32P. As in the case of g32P, proteolytic removal of a very acidic, COOH-terminal region results in increasing the helix-destabilizing ability of HMG-1 [41], an effect that again may result from this region, partially occluding a positively charged ssDNA binding site [42]. Instead of interacting with other replication proteins, the acidic COOH-terminus of HMG-1 is thought to interact with basic histones [43].

The NH$_2$-terminal "B" region is essential for cooperative binding to ssDNA. The first indication that the basic NH$_2$-terminal "B" region (here defined as the first nine amino acids in g32P: Met-Phe-Lys-Arg-Lys-Ser-Thr-Ala-Glu) might play a role in cooperative g32P:g32P interactions came from limited proteolysis studies on g32P in the presence and absence of ssDNA [31]. While cooperative g32P binding to poly(dT) decreased the rate of proteolytic cleavage of the NH$_2$-terminal "B" region, noncooperative binding to oligo-[d(pT)$_8$] did not retard this cleavage [31]. Differential scanning microcalorimetry [44] and fluorescence-quenching measurements [34] both confirm that the NH$_2$-terminal "B" region is essential for cooperative binding. On the basis of fluorescence-quenching measurements, g32P, g32P-A, and g32P-(A+B) all bind oligonucleotides such as d(pT)$_8$ with an affinity of about 10^6 M^{-1} [34]. While limited proteolysis has no significant effect on the affinity of g32P for d(pT)$_8$, loss of the "B" region decreases the affinity of g32P for poly(dT) by at least 200-fold. Thus, the dissociation constant for the g32P:poly(dT) complex is at least 1.4×10^{-9} M, whereas that for the g32P-(A+B):poly(dT) complex is only 3×10^{-7} M [34]. More recent fluorescence studies have confirmed that g32P-(A+B) binds to poly(dT) with approximately the same affinity as to d(pT)$_8$, with a cooperativity parameter of close to unity [35]. Since g32P-B does not undergo indefinite self-aggregation [45], it appears that the first nine amino acids are essential for g32P:g32P protein interactions whether they occur between "free" g32P molecules in solution or between adjacent g32P molecules bound to a ssDNA lattice.

The protease-resistant core of g32P contains the ssDNA binding region. Despite the loss of cooperative protein:protein interactions, g32P-(A+B) (also known as g32P core or g32P*) results in comparable polynucleotide lattice deformation to that observed with g32P [35]. Along with the fact that both g32P and g32P-(A+B) have similar affinities for oligo-d(pT)$_8$ [34], these data suggest that removal of the "A" and "B" regions from g32P has only a minimal effect on its intrinsic interaction with nucleic acids. It is thus reasonable to conclude that the ssDNA binding domain is contained within the region spanning residues 22–253. This idea is further supported by an ultraviolet light-induced crosslinking study that demonstrated that when g32P:ssDNA complexes are photolysed, all of the site(s) of crosslinking are located within the g32P-(A+B) core region [45].

To further define individual amino acids involved in DNA binding, we attempted to take advantage of the P7 temperature-sensitive mutant gene 32 protein. In vivo studies suggested that although the P7 mutant protein was unable to bind ssDNA at nonpermissive temperatures, it nonetheless retained the ability to participate in several g32P:protein interactions involving other *E coli* and bacteriophage T4 proteins [46]. These findings lead to the proposal that the P7 mutation is located close to the ssDNA binding site [47]. Using comparative high-performance liquid chromatography (HPLC) tryptic peptide mapping we demonstrated that the P7 mutant protein contains a cysteine in place of arginine 46 [48]. ssDNA binding studies verified that

the P7 protein is indeed unable to bind ssDNA at temperatures above 37°C in vitro. Differential scanning microcalorimetry, however, indicates that in contrast to the wild-type protein, which denatures at 55°C [44], the P7 protein undergoes a single thermal transition at 37°C [49]. This result suggested that the inability of the P7 protein to bind ssDNA above 37°C is simply due to the unfolding of the protein. That the single transition at 37°C corresponds to complete unfolding of the P7 g32P was verified by limited proteolysis studies. At 25°C, g32P-(A+B) can be easily obtained by limited trypsin digestion of g32P from either the wild type or P7 temperature-sensitive strains. In contrast, at 43°C, the core protein can only be obtained from the wild-type protein. At the restrictive temperature, trypsin immediately cleaves the P7 g32P into a large number of small peptides, none of which is larger than 5,000 daltons. While arginine 46 is obviously essential to maintain the thermal stability of g32P, our results indicate that it need not, and in fact probably does not, play a direct role in ssDNA binding. Although an unusual clustering of missense mutations that map between residues 36 and 125 has led to the suggestion that this region contains the DNA binding domain [50], the ^{1}H-NMR, metal binding, and in vitro mutagenesis experiments detailed below provide the first direct evidence for identifying individual amino-acid residues at the interface of the g32P:ssDNA complex.

Nitration of Gene 32 Protein Identifies Tyr Residues as Involved in DNA Binding

Nitration of five of the eight Tyr side chains in g32P with tetranitromethane (TNM) was found to completely prevent DNA binding as assayed with a circular dichroism binding assay employing single-stranded fd DNA [28]. In addition, binding of fd DNA to g32P prior to treatment with TNM completely prevented nitration of the five Tyr residues [28]. Thus, one or more Tyr side chains of the protein appeared to be involved in the nucleotide binding surface.

^{1}H-NMR Detection of Tyr and Phe Side Chains Involved in Oligonucleotide Binding to G32P*

High-resolution ^{1}H-NMR of a protein and its ligand complexes can be a powerful spectroscopic probe of those amino acid residues involved in ligand binding if individual proton resonances, or at least resonances of one class, undergoing perturbation consequent to ligand binding can be assigned. Difference spectra coupled with selective deuteration of amino acid types can aid in such a study, and this approach has been successfully applied to the identification of side chains of g32P that participate in nucleotide binding [51], as illustrated below.

Native g32P oligomerizes in solution even in the absence of nucleotides, resulting in extreme broadening of its ^{1}H-NMR resonances, with the exception of a few signals that can be assigned to amino acid residues in the N- and C-terminal regions of the protein [52]. Much evidence now indicates that both the N- and C-terminal regions are separate domains that must undergo considerable rotational motion relative to the main core of the protein. This accounts for the sharp ^{1}H-NMR signals from protons on amino acid side chains located in these domains. The extreme broadening of the ^{1}H-NMR spectrum from most of the protein places a severe limitation on NMR studies. The broadening problem was solved by the discovery that both the N- (residues 1–21) and C-terminal (residues 254–301) domains can be removed by limited proteolysis with trypsin. The remaining core protein, (g32P*)

still retains its high affinity for single-stranded DNA but loses the cooperative feature of the binding. Most importantly for NMR studies, the core protein does not oligomerize even when oligonucleotides as long as 12 residues are present [52]. This feature results in ^{1}H-NMR (500 MHz) spectra of relatively high resolution for both g32P* alone as well as for a variety of oligonucleotide complexes of g32P* [52]. While not widely employed, limited proteolysis may be useful in preventing oligomerization or in isolating domains from proteins that are not otherwise suitable for NMR studies because of protein-protein interactions.

In view of the size of g32P* (Mr 26kd), the proton spectra are still composed of many overlapping resonances. By forming one-dimensional differences spectra of the complexes minus the uncomplexed protein, it was found that a large number of aromatic ^{1}H resonances assignable to Tyr and possibly also to Phe and Trp side chains were shifted upfield by oligonucleotide binding. Progressively more of these resonances shifted as the bound nucleotide was increased from two to ten residues in length [52].

While the initial ^{1}H difference spectra were valuable in qualitatively demonstrating the close approach of the nucleotide bases to aromatic rings in g32P*, the large number of signals from a protein of this size made assignments to residue type difficult. With a protein produced in a bacterium or from an overexpression vector, the overlapping resonance problem can be greatly simplified by producing the protein from bacterial auxotrophs growing on selectively deuterated amino acids. This approach to simplifying the ^{1}H-NMR spectra of g32P* has proven very successful for assigning aromatic proton resonances in g32P* and its oligonucleotide complexes as illustrated below [51].

Gene 32P was overproduced from a tyrosine auxotroph of *E coli* growing on perdeuterated tyrosine (dTyr). The protein was then proteolyzed to g32P*. If the ^{1}H spectrum of the perdeuterated protein is then subtracted from that of the protonated protein, a ^{1}H spectrum of the Tyr residues alone is obtained, as shown in Figure 2A. The most upfield signal in this spectrum is that of a single 3 or 5 proton of a Tyr, and integration of the area under the complete spectrum corresponds closely to that of the 32 protons expected from the eight Tyr residues of the protein. The same technique applied to the protein containing perdeuterated Phe yields equally good results (Fig. 2B).

With the ability to generate difference ^{1}H-NMR spectra of the isolated Tyr or Phe protons, the shifts in these signals on nucleotide binding can then be assessed by superimposing a second set of difference spectra, ie, those formed by subtracting the ^{1}H spectrum of the oligonucleotide complexes of the perdeuterated protein from the same complexes of the protonated protein. A progression of three such spectra for the ApA, d(pA$_4$), and d(pA)$_6$ complexes of dTyr g32P* is pictured in Figure 3A–C. The difference spectra of the complexes are superimposed on the Tyr difference spectra of the unliganded protein to determine which resonances shift in the complex. The darkened resonances are those that shift, and the sequential spectra show that dinucleotide binding shifts resonances corresponding to two Tyr residues, tetranucleotide binding shifts resonances corresponding to four Tyr residues, and hexanucleotide binding shifts resonances attributable to five Tyr residues (Fig. 3A–C).

Similar spectra for the dPhe g32P* show the dinucleotide to shift no Phe resonances, the tetranucleotide to shift signals corresponding to two Phe residues, and a hexanucleotide to shift the same signals as the tetranucleotide, ie, those
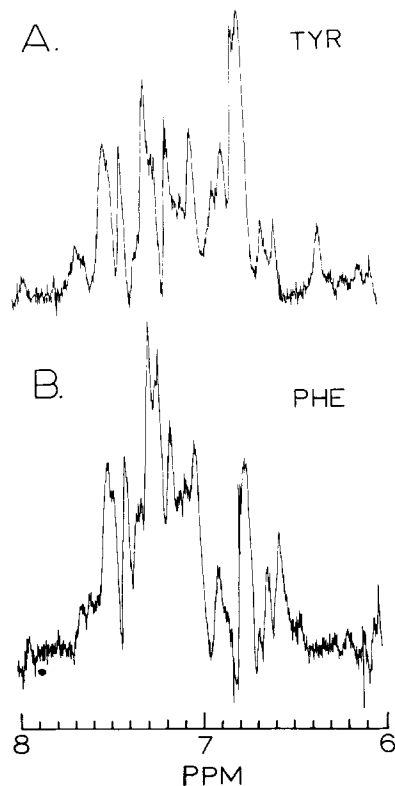
Fig. 2. Aromatic ^{1}H-NMR (500 MHz) difference spectra. A. Gene 32P* minus dTyr gene 32P*. This difference spectrum represents resonances that are due to tyrosine protons alone. B. Gene 32P* minus dPhe gene 32P*. This difference spectrum represents resonances that are due to phenylalanine protons alone.

attributable to two Phe residues. The difference spectra for the tetranucleotide complex are shown in Figure 3D. By comparing the sum of the shifted resonances in the Tyr and Phe difference spectra with the total observed in a protonated protein, they are found to account for all the shifts induced by nucleotide binding in the aromatic proton region of the ^{1}H-NMR spectrum, suggesting that the Trp side chains do not closely approach the base rings of a bound nucleotide. This was confirmed by the use of a g32P* containing perdeuterated Trp residues.

With one or two exceptions, the shifts of the Tyr and Phe proton resonances induced by nucleotide binding are upfield in direction and have maximum $\Delta\delta$ values of 0.2 to 0.3 ppm, substantially less than the >0.7 ppm upfield shifts that might be expected from tight intercalation with ring-to-ring stacking distances of ~ 3.4 Å [53]. A more probable model than a "glove-fit" intercalation is one in which the rings of the aromatic amino acids approach the bases in some regular fashion down the nucleotide chain and form part of a series of hydrophobic pockets accepting the base rings. The postulate that these aromatic amino acid side chains participate in the formation of hydrophobic binding pockets for the base rings is a more cautious interpretation of the data than the term "intercalation." As base rings are brought near the amino acid rings in a model, the ring current shifts are determined by both
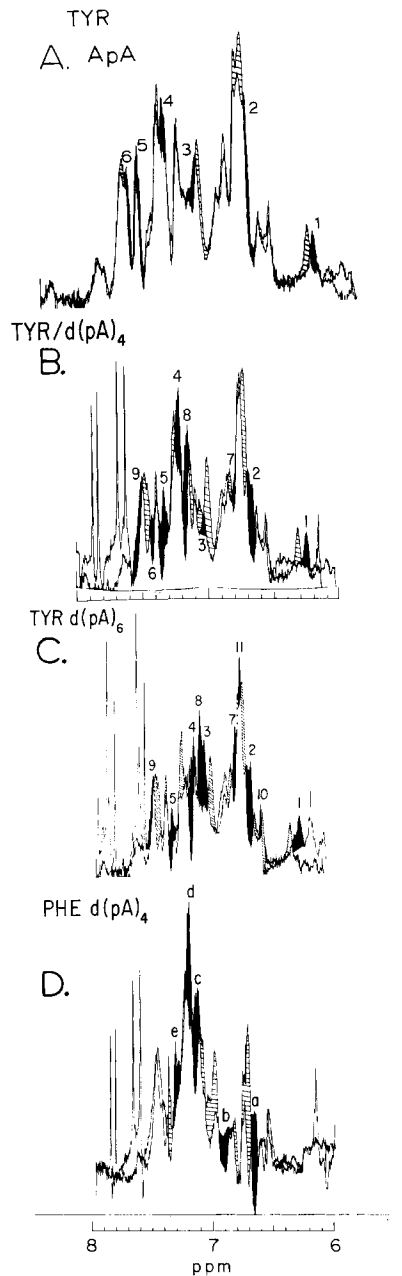
Fig. 3.   A. Aromatic ^{1}H-NMR (500 MHz) difference spectra of the g32P*:ApA complex. Spectrum of Tyr resonances from Figure 2A (envelope of hatched peaks) superimposed on the Tyr resonances of the dTyr g32P*:ApA complex generated in the same manner, ie, g32P*:ApA complex minus dTyr g32P*:ApA complex as represented by the envelope over the filled-in peaks. B. Spectrum of Tyr resonances (Fig. 2A) superimposed with g32P*:d(pA)$_4$ complex minus dTyr g32P*:d(pA)$_4$ complex. C. Spectrum of Tyr resonances (Fig. 2A) superimposed on the Tyr resonances of the g32P*:d(pA)$_6$ complex. D. Spectrum of Phe resonances (Fig. 2B) superimposed on the Phe resonances of the g32P*:d(pA)$_4$ complex.

distance and angle and can give a wide range of magnitudes as well as be both upfield or downfield. Both phenomena are observed in Figure 3. An additional factor to be considered is the likely possibility that in the protein-oligonucleotide complex the base rings are not rigidly fixed relative to the aromatic amino acid rings, and therefore observed "ring current" shifts are the dynamic average of a limited exchange process.

Using a cut-and-weigh method and assuming a standard line shape, we found that the shifted resonances in Figure 3C assignable to tyrosyl protons appear to account for five Tyr residues, while those in Figure 3D account for two Phe residues. Thus, the DNA binding surface of g32P* appears to contain five Tyr and two Phe residues that are within 4.5 Å of the base rings of the bound nucleotide. The NMR results as well as the nitration experiments focus attention on the amino acid sequence of the molecule extending from Lys 71 to Trp 116, which contains six of the eight Tyr residues in g32P*. This region of the peptide chain must form at least part of the DNA binding domain (Fig. 4).

We have recently shown that gene 32 is a zinc metalloprotein, the single Zn(II) ion per molecule being one of the structural elements that maintain the required conformation for binding to DNA with high affinity [54]. Titration with organic mecurials, which displace the Zn(II), as well as visible and near-UV absoprtion spectra of the Co(II) derivative show that the metal ion is bound to three S^- groups and probably one nitrogen ligand. The most likely coordination site involves Cys 77, His 81, Cys 87, and Cys 90, since the molecule contains only one other Cys at position 166. Removal of the Zn(II) results in a dramatic decrease in the DNA binding affinity of the Apo g32P, further evidence that the region of the primary structure from residue 71 to 116 is involved in forming a DNA binding domain [54].

Site-Directed Mutagenesis as a Means of Assigning ^{1}H-NMR Signals to Specific Amino Acid Residues Involved in Nucleotide Binding

In theory, site-directed mutagenesis can be used to assign the ^{1}H resonances from any specific residue, if that residue is mutated to one whose proton resonances have very different chemical shifts. The high-resolution difference spectra that can be constructed with the newer NMR spectrometers should resolve such resonances in the ^{1}H-NMR spectrum of native protein minus mutant protein. This approach would be practical only if directed at either relatively infrequently represented residues in a given sequence or at a particular region of the sequence indicated to be involved in some function by other probes. Both conditions apply to the Tyr residues of g32P*, since the previous data have focused on the sequence from residues 71 to 116 as forming part of the DNA binding domain, and six of the eight Tyr residues are present in this sequence. Since gene 32 was cloned, it was relatively straightforward to begin site-directed mutagenesis designed to change the six Tyr residues, one at a time, to Ser. The first residue changed was Tyr 115 to Ser 115.

-Lys[71]-Trp-Tyr-Ile-Glu-Thr-Cys-Ser-Ser-Thr-His-Gly-Asp-Tyr-Asp-

Ala-Cys-Pro-Val-Cys-Glu-Tyr-Ile-Ser-Lys-Asn-Asp-Leu-Tyr-Asn-Thr-

Asp-Asn-Lys-Glu-Tyr-Ser-Leu-Val-Lys-Arg-Lys-Thr-Ser-Tyr-Trp[116]

Fig. 4. Amino acid sequence of gene 32 protein from Trp 71 to Trp 116 believed to form part of the DNA binding domain.

The resultant Ser 115 g32P* still binds to DNA, but with reduced affinity [51]. The ^{1}H-NMR difference spectrum, wild-type g32P* minus Ser 115 g32P* shows two major difference peaks (2 and 4 in Fig. 5) clearly assignable to Tyr 115 and representing the 3,5 and 2,6 protons of one of the Tyr residues interacting with the nucleotide bases (Fig. 3). This procedure should allow identification of all the Tyr residues involved in DNA binding. Since there are a relatively large number of Phe residues scattered throughout the sequence, and none between 71 and 116, the same approach to identifying the Phe sidechains involved in nucleotide binding will be more complex. Several minor peaks in the mutant aromatic ^{1}H-NMR difference spectrum (less than one proton) suggest that the Tyr 115 to Ser 115 mutation has shifted other aromatic resonances slightly (Fig. 5). While this does not interfere with interpretation in the present example, such shifts, if large, could seriously complicate the procedure.

Mapping the Nucleotide Binding Domain of Gene 5 Protein

Gene 5 protein (a dimer of Mr 9.7 kd/monomer) is a considerably more tractable system than g32P; consequently, work on this protein has reached a stage at which detailed structural descriptions are possible. Earlier progress in the study of g5P has been reviewed [22] and will not be covered in detail here. Instead, an outline of earlier work will be given, followed by recent progress in the ^{1}H-NMR studies of g5P.

In vivo, g5P forms long superhelical assemblies with fdDNA [55]. The central questions in the study of g5P relate to the details of the DNA binding mechanism and to the topographic properties of the protein that governs assembly into the supermolecular structure. A refined crystal structure of the unliganded protein has been determined [56], facilitating progress towards answering these questions.
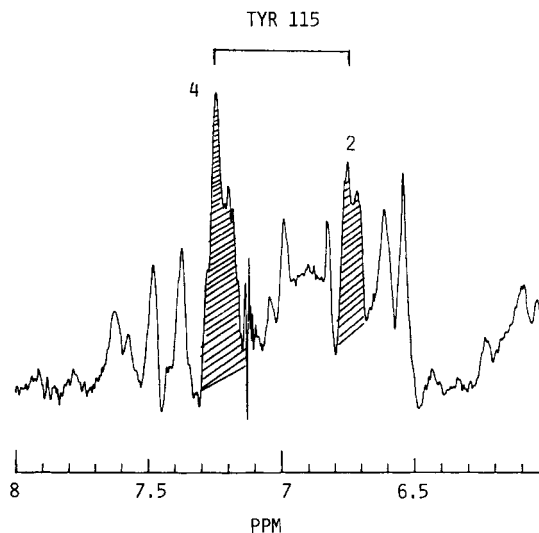

Fig. 5. Aromatic ^{1}H-NMR (500 MHz) difference spectrum, gene 32P* (wild-type) minus gene 32P* (Ser 115 mutant). Peaks 2 and 4 are the peaks in the difference spectrum of the unliganded protein that shift to positions marked 2 and 4 in the spectra of the various nucleotide complexes in Figure 3. (From [52]).

Protein Chemistry of g5P

Early protein chemistry studies laid the foundation for identifying the amino acids involved in the ssDNA interaction. The amino acid sequence of g5P [57–59] is shown in Figure 6. Application of the Chou-Fasman rules to this sequence predicted the protein secondary structure to be ~ 90% β-sheet [60], a prediction subsequently borne out by the X-ray work. Nitration of the protein with TNM was found to modify three Tyr residues in the unliganded protein, resulting in the abolition of ssDNA binding [60]. Prior complexation with DNA prevented the nitration of all three Tyr residues. Peptide mapping identified the nitrated tyrosines as residues 26, 41, and 56, which are underlined in Figure 6. Thus it was initially assumed that there were two buried Tyr residues (34 and 61) that were not available for nitration and three "free" surface Tyr that were potentially available for participation in nucleotide binding. The discrimination between surface and buried residues by TNM is a function often ascribed to this reagent, but subsequent structural data show that this distinction is not absolutely achieved, even for such a relatively small protein. The X-ray structure showed that three tyrosines are indeed located within a putative DNA binding groove [56], but that these three are residues 26, 34, and 41 (circled in Fig. 6), rather than 26, 56, and 41 as implied by nitration. Residue 56 is mostly buried with one edge exposed to solution, apparently accounting for its nitration. On the other hand, residue 34, exposed to solvent on one face, is relatively occluded at the bottom of the groove, which may account for its failure to nitrate.

Acetylation of the Lys residues in g5P with acetylimidazole was found to prevent DNA binding, and acetylimidazole treatment of g5p-DNA complexes caused dissociation of the complexes, results that strongly implied that at least some of the lysyl side chains participate in DNA binding [60]. Ultraviolet crosslinking of d(pT)$_4$ and fdDNA to g5P showed that Cys 33 lies close to the DNA-protein interface [61], while other workers obtained crosslinking of a residue in the 70–77 region after irradiation of intact fd-infected cells [62].

^{1}H-NMR studies of Gene 5 Protein-Deoxyoligonucleotide Complexes

NMR studies of g5P and its interaction with oligonucleotides [63–68] implicated tyrosine, phenylalanine, arginine, and lysine residues in the DNA binding mechanism. Experiments conducted at high field (500 MHz) were used to conclude that one phenylalanine and two or three tyrosine rings are involved in stacking interactions with nucleotide bases [69,70]. In the case of nucleotides possessing a 5' phosphate group, the presence of a preferential phosphate binding site at the 5' end of the DNA binding groove was indicated both by ^{31}P– [71] and ^{1}H-NMR [70] spectra. As indicated in Figure 7, difference spectra of complexes involving g5P and a series of adenylyl oligodeoxynucleotides revealed that resonances assignable to one phenylalanine and one tyrosine moved upfield on binding a mononucleotide [70]. These same

```
                 10                    20                         30
  M-I-K-V-E-I-K-P-S-Q-A-Q-F-T-T-R-S-G-V-S-R-Q-G-K-P⦿S-L-N-E-Q-L
              40                    50                       60
 -C⦿V-D-L-G-N-E⦿P-V-L-V-K-I-T-L-D-E-G-Q-P-A-Y-A-P-G-L-Y-T-V-H
           70                    80                   87
  -L-S-S-F-K-V-G-Q-F-G-S-L-M-I-D-R-L-R-L-V-P-A-K
```
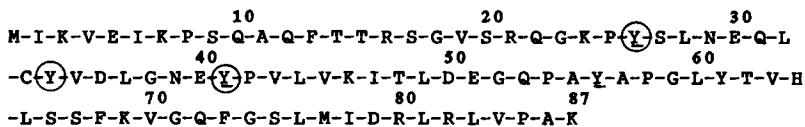
Fig. 6. Primary structure of gene 5 protein. The underlined Tyr residues are those nitrated by TNM, while the circled Tyr and Phe residues are those postulated to be involved in DNA binding.
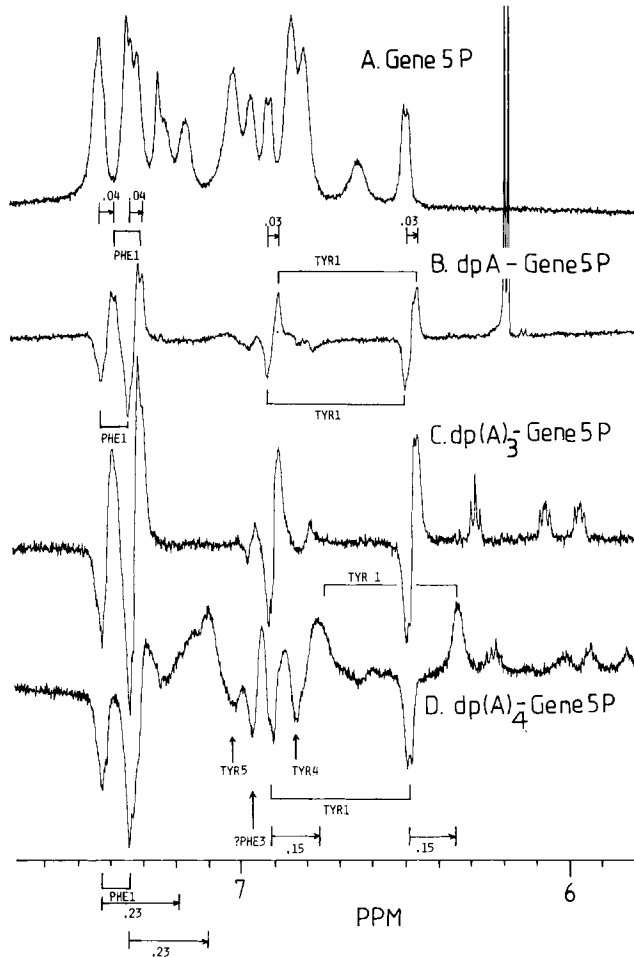
Fig. 7. ^{1}H-aromatic proton NMR difference spectra of g5P:oligonucleotide complexes minus g5P. A. Gene 5 protein alone. In this and the following figures, the dash is a minus sign, and the nucleotide symbol d(pA)$_n$ refers to the complex with gene 5 protein. B. Spectrum of the pA complex (1:1) minus that of the unliganded protein. C. Spectrum of the d(pA)$_3$ complex minus that of the unliganded protein. D. Spectrum of the d(pA)$_4$ complex minus that of the unliganded protein. (From [70]).

resonances moved much further upfield on the binding of d(pA)$_4$ and even further upfield on the binding of d(pA)$_8$. In addition to the resonances assignable to these two groups, the longer oligonucleotides also shifted signals that were apparently assignable to two additional Tyr (Fig. 7D). In reasonable, though not perfect, agreement, others have reported that binding of several adenylyl nucleotides results in significant upfield shifts of resonances belonging to one phenylalanine and two tyrosine residues [69,72]. From the first observations [63–68], it became apparent that the upfield shifts experienced by the g5P aromatic resonances were caused by planar (or nearly planar) stacking interactions with the nucleotide bases.

The X-ray Structure of g5P and the Advent of a Molecular Model of the Protein-DNA Interaction

Solution of the X-ray structure of unliganded g5P revealed an intimately associated dimer with a dyad axis, each monomer being composed of three major loops (Fig. 8A) [56,73]. The first, termed the "DNA binding loop," is composed of residues 15–32. This loop, which contains Tyr 26, Arg 16, Arg 21, and Lys 24, has larger than normal temperature factors, suggesting a flexible structure. It has been
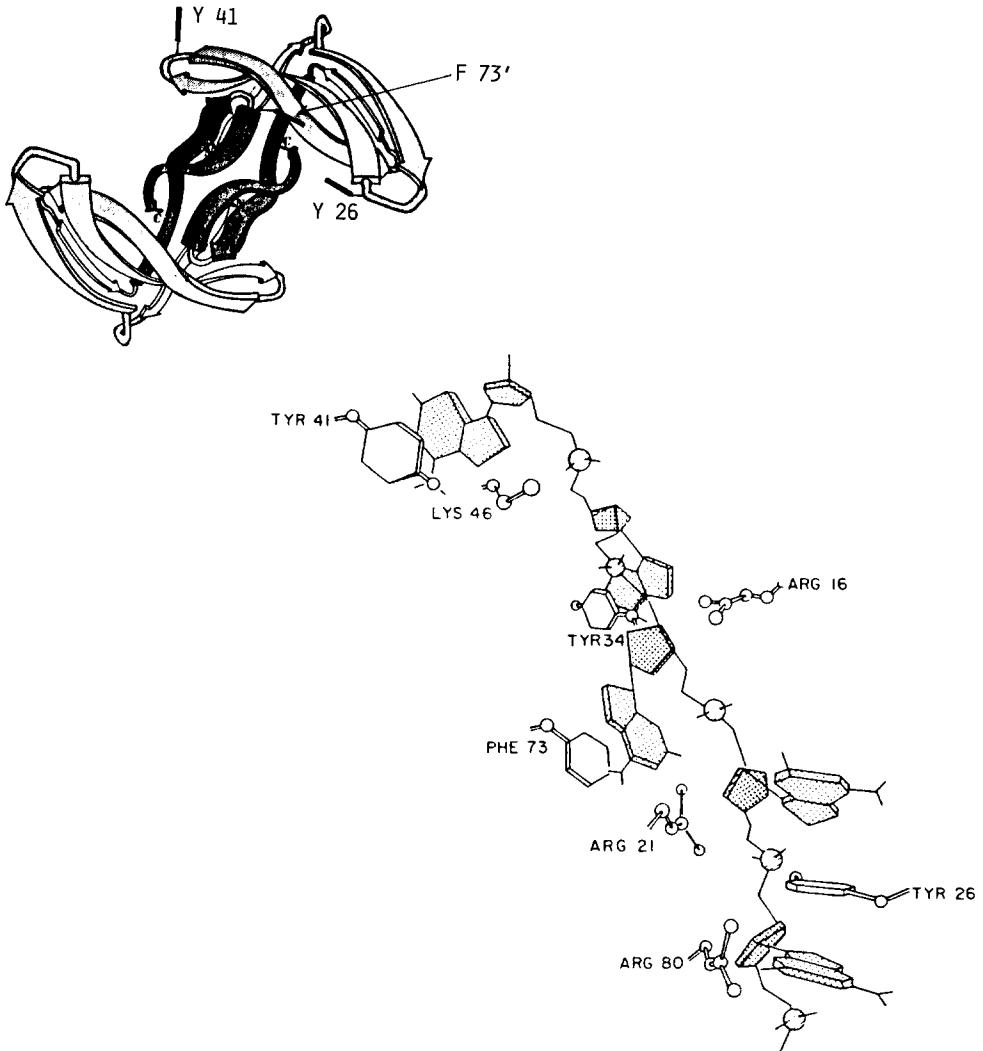

Fig. 8.   A. Richardson representation of the three-dimensional crystal structure of gene 5 protein (from ref. [56]). B. Model of the stacking of the bases of a fully extended pentanucleotide on the aromatic residues, Tyr 26, Phe 73', Tyr 34, and Tyr 41, located along the three-stranded β-sheet shown in Figure 6A (from ref. [74]).

proposed that this loop may undergo conformational changes on ligand binding. The second loop, composed of residues 38–49, is known as the "complex loop." The final ("dyad") loop (residues 62–82) forms a "hook" that interacts with the hook from the second monomer, forming the dimer.

Earlier attempts to grow satisfactory g5P:oligonucleotide cocrystals proved unsuccessful [74], though high quality g5P:d(pA)$_4$ crystals have been reported recently [75]. Consequently, Brayer and McPherson [74,76,77] have used the X-ray structure of the unliganded protein and solution data to generate a plausible model of the protein:DNA interaction. The essential features of their model are shown in Figure 8B. Five nucleotides, in fully extended conformation, are placed along the putative DNA binding channel, with bases stacked on Tyr 26, Phe 73', Tyr 34, and Tyr 41, with two of the five bases stacked on either side of Tyr 26. The phosphate backbone is complexed by four basic residues, two from the DNA binding loop and two from the main body of the protein. A better fit of the model could be obtained by moving residues in the DNA binding loop somewhat [74].

Recent NMR:Two-Dimensional NOESY Spectra of Gene 5 Protein and Oligonucleotide Complexes

In the light of the X-ray structure and proposed model for the g5P:DNA complex, we have recently examined the interaction of g5P with several oligonucleotides using two-dimensional (2D) ^{1}H nuclear Overhauser enhancement spectroscopy (NOESY) [78]. A satisfying amount of detailed information has been obtained using this method, despite the fact that g5P and its oligonucleotide complexes have rather large molecular weights as far as 2D NMR studies are concerned. Figure 9 shows the completely assigned aromatic region from a NOESY spectrum of unliganded g5P. The resonance assignments were based on crystallographic chemical modification and isotopic substitution data: for details of the assignment process, the reader should consult the original article [78]. A similar region from NOESY spectra of a 0.5:1 d(pA)$_8$:g5P mixture is shown in Figure 10. In this case, each d(pA)$_8$ strand must bind two g5P molecules cooperatively, forming a tetrameric protein complex containing two oppositely directed d(pA)$_8$ molecules [63].

The most striking finding from the NOESY spectra is that mixtures of g5P with (dpA)$_4$, d(A)$_4$, d(pT)$_4$, d(pA)$_8$, d(A)$_{12}$, and d(pT)$_{12}$ cause significant upfield movements of the signals from only two aromatics; Tyr 26 and Phe 73 [63] (G.C. King and J.E. Coleman, unpublished). In the case of both d(A)- and d(T)-containing oligonucleotides, the Phe 73 shift caused by a tetramer is close to doubled by an octamer and/or dodecamer. For example, 0.5:1 d(pA)$_8$:g5p and 0.3:1 d(pT)$_{12}$:g5P mixtures demonstrate Phe 73 shifts of approximately -0.52 and -0.32 ppm, respectively, while noncooperative complexation with d(pA)$_4$ and d(pT)$_4$ causes shifts of -0.23 and -0.15 ppm, respectively. The Tyr 26 upfield shifts tend to increase somewhat with larger oligonucleotides, but not to the same degree. Other aromatic resonances, including those of Phe 13, Tyr 34, and Tyr 41, show some small perturbations on oligonucleotide binding, but these are generally less than 0.05 ppm in magnitude [78].

All examined oligonucleotides also cause significant upfield shifts of resonances assigned to the methyl groups of Leu 28. This shift may well indicate the existence of a general hydrophobic interaction between a nucleotide base and this sidechain, or it may reflect a conformational change that brings the Tyr 26 ring current field into
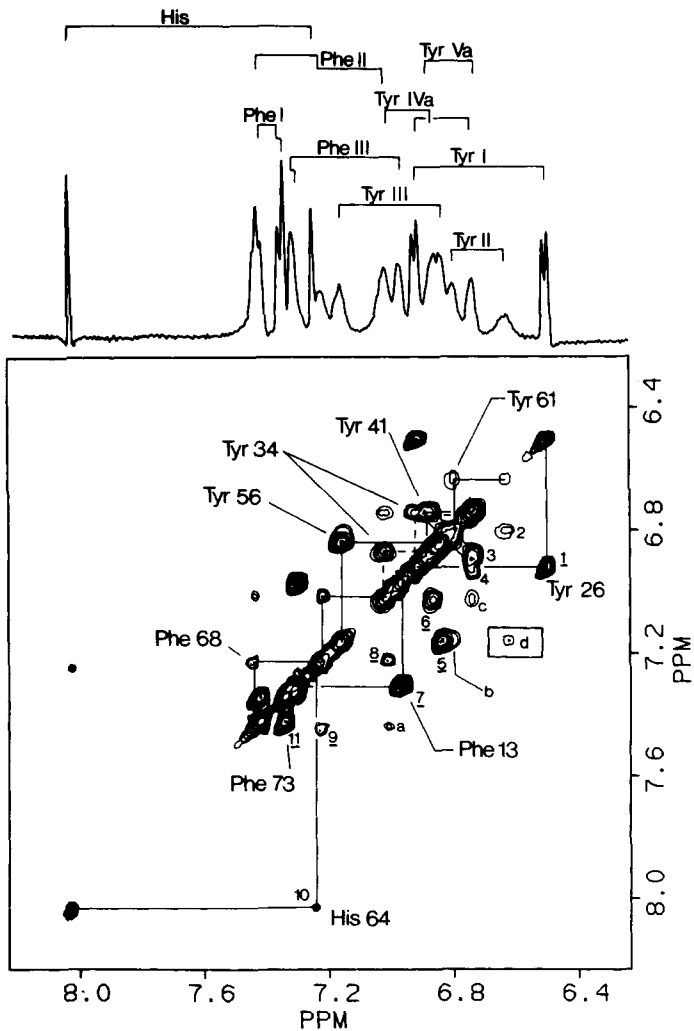
Fig. 9. Aromatic region from a 200-ms NOESY spectrum of native g5P, with sequence-specific resonance assignments [78]. Nearest-neighbor correlations are numbered 1-11, where the underlined cross-peaks reflect positions at which connectivities can be observed in COSY spectra. Longer-range correlations are designated a-d; peak d (inset) is contoured at a lower level. The one-dimensional ^{1}H aromatic spectrum and assignments to residue type are indicated at the top. (From [78]).

closer proximity with this residue. These possibilities have not been distinguished as yet, but it is apparent that Leu 28 has some role in the interaction.

The clear lack of significant shifts for the Tyr 34 and Tyr 41 signals under the present conditions is at variance with previous conclusions drawn from 1D spectra [70,69]. From close examination of the NOESY and 1D difference spectra, it appears that movement of the resonances for all five protons of Phe 73 through the center of the spectrum, coupled with small movements of the Tyr 34, Phe 13, and Tyr 41 resonances, introduces sufficient complexity to prevent unequivocal interpretation of the 1D difference spectrum. In the 2D spectrum, where the crosspeaks are readily resolved, the NOESY spectrum is much easier to interpret.
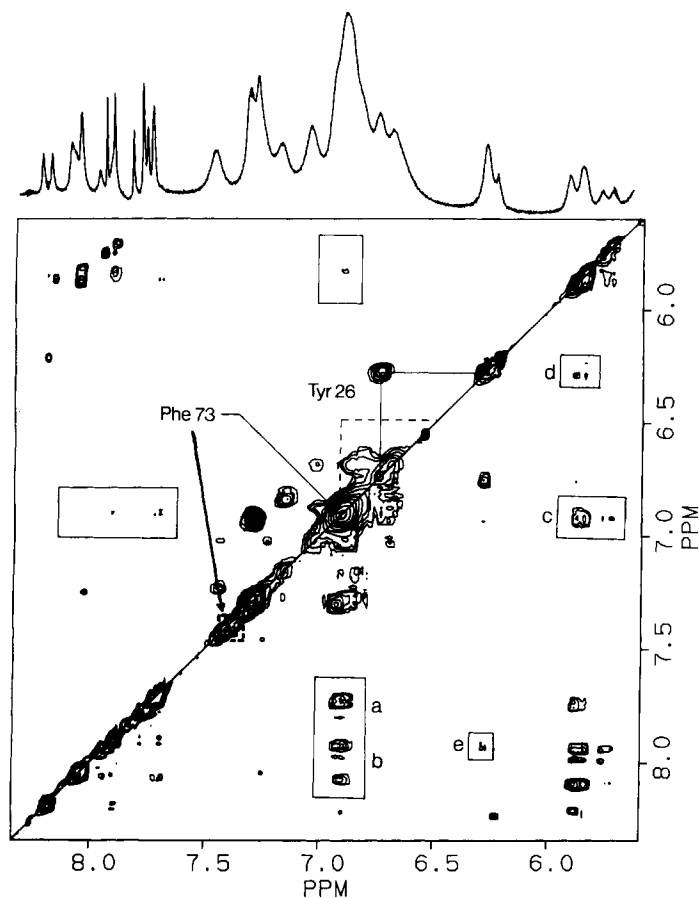
Fig. 10. Combined 100-ms and 750-ms NOESY spectra of a 0.5:1 mixture of d(pA)$_8$ and G5BP. The complexation-induced shift of the Phe 73 resonances is −0.6 ppm, while the shift of the Tyr 26 resonances is −0.27 ppm. Intermolecular NOE's between adenine H2, H8, and H1′ resonances and Phe 73 signals are indicated in the boxed areas a, b, and c, respectively. Intermolecular NOEs between the H1′ and H2 resonances on adenine residue and the (3,5)H signals of Tyr 26 are indicated in areas d and e (from ref. [78]).

Upfield shifts of protein resonances are suggestive evidence for stacking interactions. Direct intermolecular NOEs confirm and extend these findings. In the case of the d(pA)$_4$:g5P interaction, the nucleotide resonances can be specifically assigned with sequential procedures [78], allowing the identification of specific nucleotide-residue interactions. NOEs are observed between resonances of base 1 and Tyr 26 and between the resonances of bases 2 and 3 and Phe 73. As shown in Figure 10, similar NOEs are present in the d(pA)$_8$:g5P complex. While specific assignments of the base rings cannot be made from the spectrum of the d(pA)$_8$ complex (Fig. 10) because of overlap, there is at present no reason to believe that the protein:oligonucleotide interactions differ in nature from those of the d(pA)$_4$ complex.

The NOESY data obtained to date have led us to suggest a "flexible clamp" model for the g5P:oligonucleotide interaction. In this model, nucleotides are "clamped" by stacking with Tyr 26 on the flexible DNA binding loop and with Phe

73′ more rigidly held to the core of the protein. The NOE data for the d(pA)$_4$:g5P complex can be modeled quite well by rigid fitting of the nucleotide onto the protein surface, as shown in Figure 11. Consistent with the observation of very small shifts for the Tyr 41 resonances, the nucleotide path in this model does not take it close to this residue. Rather, the backbone follows a path dictated by the electropositive patch formed by the free basic sidechains of Arg 16, Arg 21, Lys 24, and Lys 46. The "clamp" may well envelope one or two bases, depending on the oligonucleotide conformation. It should be stressed that Figure 11 only shows that a model consistent with the NMR data is feasible. The precise disposition of the protein groups in the minimum energy complex will depend upon the extent to which the DNA binding loop moves on complexation, as will the precise oligonucleotide conformation. At this stage, we have chosen not to move protein groups in order to obtain a better fit, since further experiments may allow us to carry out these movements with additional experimental justification.

Implications of the 2D NMR Data for Modeling of the Gene 5 Protein-fd DNA Complex

NOESY data clearly demonstrate the primacy of Tyr 26, Leu 28, and Phe 73 in the interaction of g5P with oligonucleotide bases, which is basically due to the proximity of the electropositive patch with its affinity for nucleotide phosphates. In turn, these data suggest that there is a very much weaker interaction or none at all with Tyr 34 and Tyr 41. Whether this is a phenomenon restricted to oligonucleotide binding is an open question. However, if a nucleotide path similar to that of Figure 11 pertains to the complex of g5P with polynucleotides, then there are important
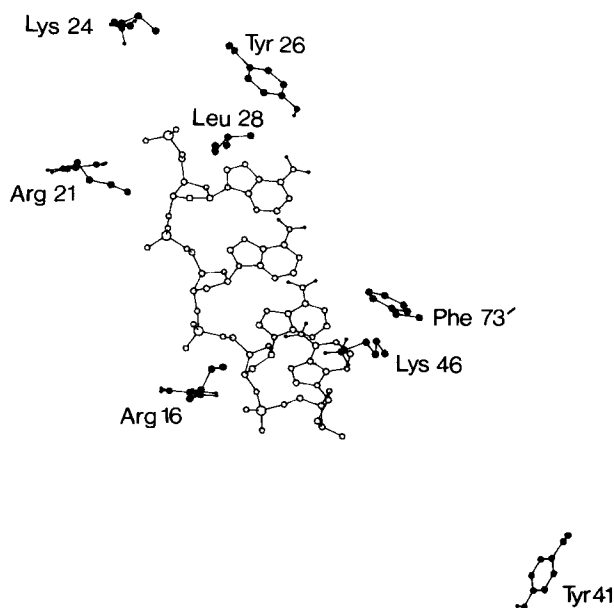

Fig. 11. Molecular fit of a helical tetranucleotide with the g5P surface. Nearby basic residues and the interacting hydrophobic sidechains identified by NMR arc shaded as is Tyr 41, which lies distant from the Tyr 26-Phe 73′ "clamp." The positions of the protein residues are taken directly from the display of the crystal structure on the counter graphics screen. (From [78]).

implications for the structure of the g5P:fd DNA superhelix. The crystallographic model depicted in Figure 8B generates a g5P:fdDNA model in which the DNA is wound on the outside of the protein superhelix [77]. This conclusion is contradicted by recent neutron scattering data, which indicate that the scattering radius of the DNA is smaller than that of the protein [79]. The latter imply packing of the DNA of the "inside" of the protein superhelix [79]. A shorter path across the protein surface could allow packing of the DNA inside the protein superhelix in the g5P:fdDNA complex, as well as some stacking of the bound nucleotides in a helical arrangement as some of the ^{31}P- and ^{1}H-NMR data suggest [70,71]. It might be that the "distal" nucleotide residues crossing each monomer (originally believed to stack on Tyr 34 and Tyr 41) actually have little interaction with the protein surface other than strategic electrostatic interactions with positively charged Lys or Arg residues, interactions which would be expected to have little effect on the ^{1}H-NMR spectrum.

SUMMARY

While it is inadvisable to generalize structural features to all single-stranded DNA binding proteins based on knowledge of the protein:nucleotide interactions observed for one member of the class, it does appear that interactions between base rings of the bound nucleotide and aromatic amino acid side chains may be the characteristic feature of the ssDNA binding proteins. The recent 2D ^{1}H-NMR data on the gene 5 protein:oligonucleotide complexes strongly suggest that this interaction can be characterized as stacking rather than as simply a component of a hydrophobic binding contact. Such stacking interactions may direct a more specific placement of the nucleotide lattice than the electrostatic attraction between phosphate backbone and basic sidechains that must account for the majority of the binding energy and is likely to be part of the interaction surface of all nucleic acid binding proteins.

General hydrophobic binding pockets may play a greater or lesser role in different proteins, but as yet there is no evidence for the absence of aromatic amino acid:nucleotide base interactions in any member of the class. Binding of the phosphate backbone to the protein and the interaction of the base rings with the protein surface may not be tightly coupled in the dynamic sense. For example, some of the NMR data suggest that the rotational or other motions of the aromatic amino acid and base rings in the complex occur on a faster time scale than the overall exchange of the nucleotide.

While proposals of general binding mechanisms are attractive, individual proteins or small subsets of a general class of protein may develop specific structural features that do not extend to all members of the group. Among the ssDNA binding proteins the presence of a structural zinc ion maintaining the DNA binding domain is unique to gene 32 protein, not a characteristic shared by SSB from E coli or gene 5 protein [54], although this feature may be shared by some retroviral single-stranded nucleotide binding proteins.

ACKNOWLEDGMENTS

REFERENCES

1. Alberts BM, Frey L: Nature 227:1313–1318, 1970.

2. Riva S, Cascino A, Geiduschek EP: J Mol Biol 54:85–102, 1970.
3. Curtis MJ, Alberts BM: J Mol Biol 102:793–816, 1976.
4. Huberman JA, Kornberg A, Alberts BM: J Mol Biol 62:39–52, 1971.
5. Jensen DE, Kelly RC, von Hippel PH: J Biol Chem 251:7215–7228, 1976.
6. Nossal NG, Peterlin BM: J Biol Chem 254:6032–6037, 1979.
7. Alberts BM, Barry J, Bedinger P, Burke RL, Hibner U, Liu C-C, Sheridan R: in Alberts BM, Fox CF (eds): "Mechanistic Studies on DNA Replication and Genetic Recombination." ICN-UCLA Symp Mol Cell Biol, Vol 19. New York: Academic Press, 1980, pp 449–471.
8. Sigal N, Delius H, Kornberg T, Gefter ML, Alberts BM: Proc Natl Acad Sci USA 69:3537–3541, 1972.
9. Burke RL, Alberts BM, Hosoda J: J Biol Chem 255:11484–11493, 1980.
10. Mosig G, Luder A, Garcia G, Dannenberg R, Bock S: *Cold Spring Harbor Symp Quant Biol* 43:501–515, 1979.
11. Formosa T, Burke RL, Alberts BM: Proc Natl Acad Sci USA 80:2442–2446, 1983.
12. Wu J-R, Yeh Y-C: J Virol 12:758–765, 1973.
13. Krisch HM, Bolle A, Epstein RH: J Mol Biol 88:89–104, 1974.
14. Gold L, O'Farrell PZ, Russel M: J Biol Chem 251:7251–7262, 1976.
15. Russel M, Gold L, Morrissett H, O'Farrell P: J Biol Chem 251:7263–7270, 1976.
16. Gold L, Lemaire G, Martin C, Morrissett H, O'Conner, P, O'Farrell P, Russel M, Shapiro R: In HJ Vogel (ed): "Nucleic Acid-Protein Recognition." New York: Academic Press, 1977, pp 91–113.
17. Krisch HM, Duvoisin RM, Allet B, Epstein RH: In Alberts BM, Fox CF (eds): "Mechanistic Studies on DNA Replication and Genetic Recombination." ICN-UCLA Symp Mol Cell Biol, Vol 19. New York: Academic Press, 1980, pp 517–526.
18. Krisch HM, Allet B: Proc Natl Acad Sci USA 79:4937–4941, 1982.
19. von Hippel PH, Kowalczykowski SC, Lonberg N, Newport JW, Paul LS, Stormo GD, Gold L: J Mol Biol 162:795–818, 1982.
20. Newport JW, Lonberg N, Kowalczykowski SC, von Hippel P: J Mol Biol 145:105–121, 1981.
21. Geider K, Kornberg A: J Biol Chem 249:3999, 1974.
22. Coleman JE, Oakley JL: CRC Crit Rev Biochem 7:247–289, 1980.
23. Williams KR, LoPresti M, Setoguchi M, Konigsberg WH: Proc Natl Acad Sci USA 1980. 77:4614–4617, 1980.
24. Williams KR, LoPresti MB, Setoguchi M: J Biol Chem 256:1754–1762, 1981.
25. Hosoda J, Moise H: J Biol Chem 253:7547–7455, 1978.
26. Delius H, Mantell NJ, Alberts B: J Mol Biol 67:341–350, 1972.
27. Carroll RB, Neet K, Goldthwait DA: J Mol Biol 91:275–291, 1975.
28. Anderson RA, Coleman JE: Biochemistry 14:5485–5491, 1975.
29. Hosoda J Takacs B, Black C: FEBS Lett 47:338–343, 1974.
30. Moise H, Hosoda J: Nature 259:455–458, 1976.
31. Williams KR, Konigsberg WH: J Biol Chem 253:2463–2470, 1978.
32. Tsugita A, Hosoda J: J Mol Biol 122:255–258, 1978.
33. Hosoda J, Burke RL, Moise H, Kubota I, Tsugita A: In Alberts BM, Fox CF (eds): "Mechanistic Studies on DNA Replication and Genetic Recombination." ICN-UCLA Symp Mol Cell Biol, Vol 19. New York: Academic Press, 1980, pp 505–513.
34. Spicer EK, Williams KR, Konigsberg WH: J Biol Chem 254:6433–6436, 1979.
35. Lonberg N, Kowalczykowski SC, Paul LS, von Hippel PH: J Mol Biol 145:123–138, 1981.
36. Greve J, Maestre M, Moise H, Hosoda J: Biochemistry 17:893–898, 1978.
37. Dunn JJ, Studier F: J Mol Biol 148:303–330, 1981.
38. Araki H, Ogawa H: Mol Gen Genet 183:66–73, 1981.
39. Williams KR, Spicer EK, LoPresti MB, Guggenheimer RA, Chase JW: J Biol Chem 258:3346–3355, 1983.
40. Chase JW, Merrill BM, Williams KR: Proc Natl Acad Sci USA 80:5480–5484, 1983.
41. Isackson PJ, Reeck GR: Biochim Biophys Acta 697:378–380, 1982.
42. Carballo M, Puigdomenech P, Tancredi T, Palav J: EMBO J 3:1255–1261, 1984.
43. Reeck GR, Isackson PJ, Teller DC: Nature 300:76–78, 1982.
44. Williams KR, Sillerud LO, Schafer DE, Konigsberg WH: J Biol Chem 254:6426–6432, 1979.
45. Williams KR, Konigsberg WH: In Chirikjian JG, Papas TS (eds): "Gene Amplification and Analysis," Vol. 2. North Holland, Amsterdam: Elsevier, 1981, pp 475–508.

46. Mosig G, Bock S: J Virol 17:756–761, 1976.
47. Breschkin A, Mosig G: J Mol Biol 112:279–294, 1977.
48. Williams KR, L'Italian JJ, Guggenheimer RA, Sillerud L, Spicer E, Chase J, Konigsberg W: In Elzinga M (ed): Comparative peptide mapping by HPLC Identification of single amino acid substitutions in temperature sensitive mutants. "Methods in Protein Sequence Analysis." Clifton, NJ: Humana Press, 1982, pp 499–507.
49. Williams KR, Konigsberg WH: In Mathews CK, Kutter EM, Mosig G, Berget PB (eds): "Bacteriophage T4: Washington, DC: Am Soc Microbiol, 1983, pp 82–89.
50. Doherty DH, Gauss P, Gold, L: Mol Gen Genet 188:77–90, 1982.
51. Prigodich RV, Shamoo Y, Williams KR, Chase JW, Konigsberg WH, Coleman JE: Biochemistry 25:3666–3672, 1986.
52. Prigodich RV, Casas-Finet J, Williams KR, Konigsberg W, Coleman JE: Biochemistry 23:522–529, 1984.
53. Giessner-Prettre C, Pullman B: Biochem Biophys Res Commun 70:578–581, 1976.
54. Giedroc DP, Keating KM, Williams KR, Konigsberg WH, Coleman JE: Proc Natl Acad Sci USA 83:1986.
55. Alberts B, Frey L, Delius H: J Mol Biol 68:139–152, 1972.
56. Brayer GD, McPherson A: J Mol Biol 169:565–596, 1983.
57. Nakashima Y, Dunker AK, Marvin DA, Konigsberg W: FEBS Lett 40:290–292, 1974.
58. Nakashima Y, Dunker AK, Marvin DA, Konigsberg W: FEBS Lett 43:125, 1974.
59. Cuypers T, van der Ouderaa F J, DeJong WW: Biochem Biophys Res Commun 59:557–563, 1974.
60. Anderson RA, Nakashima Y, Coleman JE: Biochemistry 14:907–917, 1975.
61. Paradiso PR, Nakashima Y, Konigsberg W: J Biol Chem 254:4739–4744, 1979.
62. Lica L, Ray DS: J Mol Biol 115:45–59, 1977.
63. Coleman JE, Anderson RA, Ratcliffe RG, Armitage IM: Biochemistry 15:5419–5430, 1976.
64. Garssen GJ, Hilbers CW, Schoenmakers JGG, von Boom JH: Eur J Biochem 81:453–463, 1977.
65. Coleman JE, Armitage IM: Biochemistry 17:5038–5045, 1978.
66. Garssen GJ, Kaptein R, Schoenmakers JGG, Hilbers CW: Proc Natl Acad Sci USA 75:5281–5285, 1978.
67. Garssen GJ, Tesser GI, Schoenmakers JGG, Hilbers CW: Biochim Biophys Acta 607:361–371, 1980.
68. Alma NCM, Hormsen BJM, Hilbers CW, van der Marel G, van Boom JH: FEBS Lett 135:15–20, 1981.
69. Alma NCM, Harmsen BJM, Hull WE, van der Marel G, van Boom JH, Hilbers CW: Biochemistry 20:4419–4428, 1981.
70. O'Connor TP, Coleman JE: Biochemistry 22:3375–3381, 1983.
71. O'Connor TP, Coleman JE: Biochemistry 21:848–854, 1982.
72. Alma NCM, Harmsen BJM, van Boom JH, van der Marel G, Hilbers CW: Eur J Biochem 122:319–326, 1982.
73. McPherson A, Jurnak FA, Wang AH, Molineux I, Rich A: J Mol Biol 134:379–400, 1979.
74. Brayer GD, McPherson A: Biochemistry 23:340–349, 1984.
75. McPherson A, Koszelak S, Axelrod H, Day J, Williams R, Robinson L, McGrath M, Cascio D: J Biol Chem 261:1969–1975, 1986.
76. Brayer GD, McPherson A: Eur J Biochem 150:287–296, 1985.
77. McPherson A, Brayer GD: In McPherson A, Jurnak F (eds): "Biological Macromolecules and Assemblies," Vol 2. New York: John Wiley, 1985, pp 323–392.
78. King GC, Coleman JE: Biochemistry (in press), 1986.
79. Gray DM, Gray CW, Carlson RD: Biochemistry 21:2702–2713, 1982.