^{1}H NMR (500 MHz) Identification of Aromatic Residues of Gene 32 Protein Involved in DNA Binding by Use of Protein Containing Perdeuterated Aromatic Residues and by Site-Directed Mutagenesis†

Richard V. Prigodich, Yousif Shamoo, Kenneth R. Williams, John W. Chase,‡ William H. Konigsberg, and Joseph E. Coleman*

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

Received November 15, 1985; Revised Manuscript Received February 24, 1986

ABSTRACT: Preparation of gene 32 protein containing perdeuterated tyrosyl and phenylalanyl residues has allowed the resolution of separate ^{1}H NMR signals for the Tyr and Phe residues of the protein by NMR difference spectra. Upfield shifts in the chemical shifts of a number of aromatic protons previously observed to accompany deoxyoligonucleotide complex formation with gene 32 protein [Prigodich, R. V., Casas-Finet, J., Williams, K. R., Konigsberg, W., & Coleman, J. E. (1984) *Biochemistry 23*, 522–529] can be assigned to five Tyr and two Phe residues that must form part of the DNA binding domain. Site-directed mutation of Tyr-115 to Ser-115 results in the disappearance of a set of 2,6 and 3,5 tyrosyl protons that are among those moved upfield by oligonucleotide complex formation. These findings suggest that the amino acid sequence from Tyr-73 to Tyr-115 which contains six of the eight Tyr residues of the protein forms part of the DNA binding surface.

This laboratory was the first to postulate a model for the binding locus of single-stranded DNA binding proteins consisting of partial intercalation of Tyr residues with the bases of the polynucleotide (Anderson et al., 1975; Coleman & Armitage, 1978; O'Connor & Coleman, 1983). The model was based on ^{1}H and ^{19}F NMR data obtained on oligonucleotide complexes of gene 5 protein of fd that show upfield chemical shifts of the aromatic protons (^{19}F) of Tyr (fluorotyrosyl) residues of the protein to accompany the formation of oligonucleotide complexes (Coleman & Armitage, 1978; O'Connor & Coleman, 1983). The ^{1}H NMR shifts for the gene 5 protein were confirmed by Hilbers and his colleagues (Garsen et al., 1977, 1980; Alma et al., 1981), who suggested that one Phe residue was also involved. The upfield-shifted protons were assigned to three Tyr residues and one Phe residue (O'Connor & Coleman, 1983). This model is also supported by the finding that nitration of the Tyr residues of gene 5 abolishes nucleotide binding (Anderson et al., 1975; Anderson & Coleman, 1975). The partial intercalation model is compatible with the crystal structure of the unliganded protein at 2.8 Å resolution which shows Tyr-26, Tyr-34, Phe-73, and Tyr-41 to lie along the putative DNA binding groove (McPherson et al., 1980).

Gene 5 is a small protein (a dimer of M_r 19 400) to which NMR techniques of probing structure are easily applied. NMR studies of nucleotide complex formation with other larger ssDNA binding proteins to determine if the above model represents a general binding mechanism have been hampered by protein size and the tendency of many single-stranded binding proteins to oligomerize in solution even in the absence of nucleotides. Gene 32 protein (M_r 33 487) from T4 is probably the most extensively studied single-stranded DNA binding protein, but NMR approaches to the solution structure of this protein were prevented by its oligomerization in solution (Coleman & Oakley, 1980). We have recently solved the problem by utilizing the discovery of Konigsberg and his colleagues that removal of the small N- and C-terminal domains of this protein prevents oligomerization, but does not significantly alter DNA binding (Williams et al., 1981). The resultant monomer (M_r 26 024), gene 32P*,[1] gives excellent high-resolution ^{1}H NMR spectra that show a number of aromatic resonances move upfield on oligonucleotide complex formation (Prigodich et al., 1984).

While these findings suggested a binding mechanism not dissimilar to that of a gene 5 protein, the spectra of gene 32P* are far more complex, and specific assignments of type and number of aromatic side chains involved were difficult to make. In order to more precisely map the DNA binding surface by ^{1}H NMR approaches, we have prepared gene 32 proteins containing perdeuterated Tyr and perdeuterated Phe, which greatly simplifies the aromatic proton spectra of the protein and allows the separation of signals from Tyr, Trp, and Phe residues. The assignment of specific Tyr residues as interacting with oligonucleotides has also been approached by ^{1}H NMR of gene 32 protein nucleotide complexes in which specific Tyr residues in the sequence have been altered by site-directed mutagenesis. The results of these ^{1}H NMR approaches are reported here.

Materials and Methods

Gene 32 Proteins. Homogeneous gene 32 protein was prepared as previously described (Bittner et al., 1979). *Escherichia coli* hosts auxotrophic for the amino acids tyrosine, phenylalanine, or tryptophan were kindly supplied from the American Culture Collection by Dr. Barbara Bachman. The

† This work was supported by NIH Grant GM21919 (to J.E.C.). The identification of a specific Tyr residue by site-directed mutagenesis is work to be submitted by Y.S. to the Graduate School of Yale University in partial fulfillment of the requirements for the Ph.D. R.V.P. is the recipient of National Institutes of Health Postdoctoral Fellowship AM07100. K.R.W. was supported by NSF Grant PCM8104118, and J.W.C. was supported by NIH Grants GM34697, GM23451, and CA13330.

‡ On the faculty of the Department of Molecular Biology, Albert Einstein College of Medicine.

[1] Gene 32P and gene 32P* refer to the gene 32 protein isolated from T4-infected *Escherichia coli* and the same protein with the N-terminal (1–21) and C-terminal (254–301) amino acid residues removed by limited proteolysis with trypsin.

0006-2960/86/0425-3666$01.50/0

gene 32P fragment, gene 32P*, was isolated from a tryptic digest and prepared according to a modified version of the procedure of Hosoda and Moise (1978) (Williams et al., 1981). The protein concentration was determined by using $\epsilon_{280} = 3.7 \times 10^4$ M^{-1} cm^{-1} (Jensen et al., 1976) and amino acid analysis. The protein samples were deuterium-exchanged by repeated dialysis against 10 mM $NaDCO_3$–0.3 M NaCl, pH 8. All samples contained 0.35–0.5 mM protein ± equimolar oligonucleotide, 15 mM NH_4DCO_3, and 30 mM NaCl, pD 8.4. The latter refers to the actual reading on the standard galss electrode. ^{31}P NMR spectra show the 5′-phosphate of the nucleotides to be present as the dianion under these conditions.

Nucleotides. $d(pA)_4$ and $d(pA)_6$ were purchased from P-L Biochemicals, Inc., and ApA was purchased from Sigma. The nucleotides were dissolved in D_2O, and the appropriate amount of nucleotide was determined from known extinction coefficients (Kelly et al., 1976). The deuterium-exchanged nucleotides were lyophilized, and then protein solution was added. All complexes were 1:1 with protein which yields >95% complex formation at the concentrations employed (Kelly et al., 1976).

^1H NMR Spectra. The spectra were obtained at 500 MHz on a Bruker WM-500 spectrometer at 303 K. The solvent, D_2O, served as a field-frequency lock. The samples were 0.4 mL in 5-mm tubes. Chemical shifts were measured in parts per million (ppm) downfield from a standard reference frequency (6684 Hz).

Construction of the Serine-115 Mutant of Gene 32P. The regulatory sequences and the structural gene (gene 32) for phage T4D gp32 were previously cloned (Krisch & Allet, 1982). Apparently due to the deleterious effects of gene 32 expression in *E. coli*, it has not been possible to clone the wild-type gene (Krisch & Selzer, 1981). The gene cloned by Krisch and Allet (1982), 32*am*A453, contains an amber codon (UAG) at the position of amino acid 116 in gene 32 which is tryptophan (UGG) in the wild-type protein. Since the synthesis of gene 32P is autogenously regulated at the level of translation, and sequences 5′ to the ATG initiation codon of gene 32 have been implicated in this regulation (Krisch & Allet, 1982), we reconstructed the 5′ regulatory region of the gene to permit expression. This will be the subject of a separate report (Y. Shamoo, K. Williams, and J. Chase, unpublished observations). We have utilized this construction containing the altered regulatory sequence to construct and express a mutant of gene 32P containing serine at the position of amino acid 115, normally tyrosine.

A derivative of M13mp8 was constructed by incorporating a DNA fragment containing gene 32 and obtained from pKX361 (kindly supplied by Dr. H. M. Krisch). A synthetic piece of DNA replaced the normal regulatory sequences located 5′ of the ATG initiation codon of gene 32. Single-stranded M13 DNA was isolated for site-directed oligonucleotide mutagenesis. A 21-base oligonucleotide was synthesized by using an Applied Biosystems instrument and containing two base mismatches. One of these changed the codon for tyrosine-115 (UAC) to serine (UCC), and the second changed the amber codon at 116 (UAG) back to tryptophan (UGG). Together these changes created a *Bst*N1 recognition sequence that served as an aid in identifying the new construction. The synthetic reactions using the oligonucleotide as a primer and subsequent plaque hybridization for identification of the serine-115 mutant DNA were performed as described by Mark et al. (1984). Following identification of the correct construction, M13 RFI DNA was isolated, and the *Eco*RI–*Bam*HI DNA fragment containing the altered gene 32 DNA sequence was isolated, and the overhanging ends were filled in utilizing the large fragment of *E. coli* DNA polymerase I. The resultant blunt-ended DNA fragment was ligated into *Hpa*I-cleaved pKC30 which allows temperature-dependent expression under λP_L regulation (Rosenberg et al., 1983).

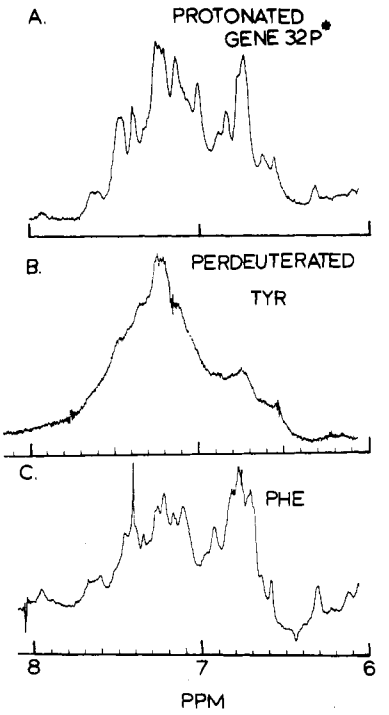

FIGURE 1: ^1H NMR (500 MHz) spectra of the aromatic protons of gene 32P* (A), of gene 32P* containing perdeuterated tyrosine (B), and of gene 32P* containing perdeuterated phenylalanine (C). Sample conditions in this and all following figures are as described under Materials and Methods.

RESULTS

*Isolated Tyrosine and Phenylalanine ^1H NMR Spectra of Gene 32P**. The complete ^1H NMR spectra of gene 32P* and its oligonucleotide complexes have been published (Prigodich et al., 1984). In this paper we are concentrating on the aromatic region, and hence, only the ^1H NMR of the aromatic protons of gene 32P* is repeated in Figure 1A. For the complete spectra the reader is referred to the previous publication (Prigodich et al., 1984).

The aromatic spectrum of gene 32P* represents the envelope of overlapping proton signals from 16 Phe, 8 Tyr, 5 Trp, and 3 His residues; hence, resolution of individual resonances is relatively low. The profile of this envelope changes significantly on formation of complexes with deoxyoligonucleotides (Prigodich et al., 1984), but any detailed analysis and assignment of individual resonances from such envelopes in the case of a protein as large as gene 32P* are not possible, hence, the focus on constructing difference spectra, especially from the simplified spectra generated by gene 32P* containing perdeuterated aromatic amino acids.

An ^1H NMR spectrum in the aromatic region of gene 32P* containing perdeuterated Tyr is shown in Figure 1B and the analogous spectrum of the protein containing perdeuterated Phe is shown in Figure 1C. From these spectra the general observations are as follows: the group of resonances from 7.3 to 7.5 ppm are Phe protons (remain in Figure 1B but are absent in Figure 1C), and in the absence of the Phe resonances many more features of the Tyr resonances resolve (Figure 1C). These spectra show that the expected incorporation of the

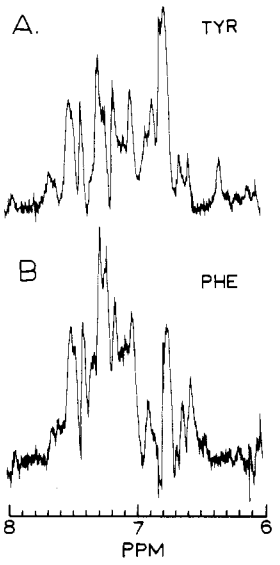

FIGURE 2: Aromatic ^1H NMR (500 MHz) difference spectra. (A) Gene 32P* minus gene 32P* (containing perdeuterated Tyr). This difference spectrum represents resonances due to tyrosine protons alone. (B) Gene 32P* minus gene 32P* (containing perdeuterated Phe). This difference spectrum represents resonances due to phenylalanine protons alone.

deuterated amino acid has taken place during protein synthesis.

The spectra of the perdeuterated proteins do not significantly enhance resolution, but the formation of various difference spectra using the ^1H spectra of both protonated and perdeuterated samples make it possible to identify individual groups of aromatic resonances much more precisely and make assignments of those signals that move on oligonucleotide complex formation. For example, by forming the difference spectra, gene 32P* – gene $32P^*_{Tyr}$[2] or gene 32P* – gene $32P^*_{Phe}$, one should generate the isolated ^1H spectra of the tyrosyl and phenylalanyl residues of the protein. These tyrosyl and phenylalanyl spectra are shown in Figure 2A,B and demonstrate the utility of the procedure. The integration of the Tyr signals shows that the isolated signal at 6.3 ppm contains only one proton. Hence, in contrast to gene 5 protein, some of the Tyr rings must be slowly rotating, and many of the 2,6- and 3,5-protons may occupy unique chemical environments on the chemical shift time scale.

Participation of Tyr Residues in Oligonucleotide Complex Formation with Gene 32P As Determined by Perdeuteration.* One can determine which of these tyrosyl protons are affected by oligonucleotide complex formation of gene 32P* by forming the difference spectra, gene 32P*–nucleotide minus gene $32P^*_{Tyr}$–nucleotide, and then superimposing the difference spectrum of the complex on the analogous difference spectra of the unliganded protein. A spectrum constructed by this procedure for the complex with ApA is shown in Figure 3A. The filled-in black signals represent new signals appearing in the nucleotide complexes, while the hatched areas indicate where these same signals have disappeared from the tyrosyl spectra of the unliganded protein.

In the previous study of the protonated protein we had identified the total number of aromatic ^1H resonances undergoing changes in chemical shift on nucleotide complex formation by the difference spectra generated by subtracting the spectrum of the unliganded protein from that of the oligonucleotide complex. Such a spectrum for the ApA complex

[2] Subscripts Tyr and Phe following gene 32P* refer to proteins containing perdeuterated tyrosine or phenylalanine, respectively.

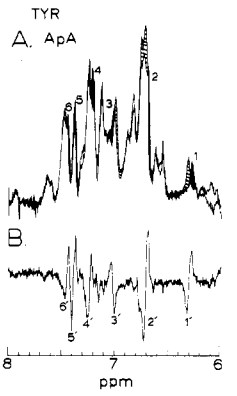

FIGURE 3: Aromatic ^1H NMR (500 MHz) difference spectra of the gene 32P*–ApA complex. (A) Spectrum of Tyr resonances from Figure 2A (envelope of hatched peaks) superimposed on the Tyr resonances of the gene 32P*–ApA complex generated in the same manner, i.e., gene 32P*–ApA complex minus gene 32P*–ApA complex containing perdeuterated Tyr as represented by the envelope over the filled in peaks. (B) Normal difference spectrum, gene 32P*–ApA complex minus gene 32P*, both proteins containing all protonated aromatic amino acids.

of gene 32P* is shown in Figure 3B. The negative peaks represent original resonance positions in the unliganded protein, while the positive peaks represent the new positions in the complex. In the case of the dinucleotide complex all the negative peaks (primed numbers, Figure 3B) can be assigned to regions of intensity disappearing from the protein on complex formation as determined by the perdeuteration method in Figure 3A, while all the corresponding positive peaks in the difference spectrum (Figure 3B) can be assigned to new signal intensity appearing in the complex (unprimed numbers, Figure 3A) as identified by the perdeuteration method. Hence, it can be concluded that the first dinucleotide binding to the gene 32P* DNA binding groove interacts only with Tyr among the aromatic residues. Integration of the signals suggests that the signals observed arise from two Tyr residues. For the ApA complex (Figure 3) the tyrosyl shifts are clearly small, ~0.1 ppm, involve approximately eight protons (two in peak 2) and are all upfield with the exception of resonance 3 which shifts ~0.1 ppm downfield.

When the tetranucleotide, $d(pA)_4$, is used for complex formation, the difference spectra generated from the combination of protonated and perdeuterated Tyr proteins show that all the Tyr protons shifted upfield by ApA are also shifted upfield by $d(pA)_4$, but $\Delta\delta$ increases from an average of ~0.1 to ~0.15 ppm (Figure 4A). Resonances 1, 2, and 5 are good examples (Figure 4A). Additional Tyr protons (peaks 7–9) are shifted by $d(pA)_4$, and one of these, peak 9, shifts downfield. While most of these shifting Tyr resonances can clearly be assigned to features in the standard aromatic difference spectrum [gene 32P*–$d(pA)_4$ minus gene 32P*] (Figure 4B), the overlap is now greater, and it is not so easy to conclude that all features of the difference spectra can be accounted for by shifts in Tyr resonances alone. In fact they cannot be as will be shown later.

When the oligonucleotide complex is formed with $d(pA)_6$, the nine Tyr signals shifted by $d(pA)_4$ are also shifted by the hexanucleotide (peaks 1–9, Figure 5A), and two additional shifted Tyr resonances become apparent (peaks 10 and 11, Figure 5A). Some of the upfield shifts have increased to ~0.2 ppm (e.g., peaks 1 and 4), while others (e.g., peaks 2, 5, 8, and 9) remain as they were in the $d(pA)_4$ complex. Once again all of the shifted Tyr resonances in Figure 5A can be assigned to features of the normal difference spectrum [gene 32P*–$d(pA)_6$ minus gene 32P*] in Figure 5B. Once again overlap

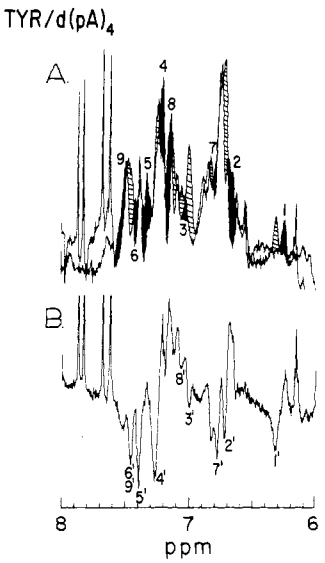

FIGURE 4: Aromatic ^{1}H NMR (500 MHz) difference spectra of gene 32P*–d(pA)$_4$ complex. (A) Spectrum of Tyr resonances from Figure 2A (envelope of hatched peaks) superimposed on the Tyr resonances of the gene 32P*–d(pA)$_4$ complex generated in the same manner, i.e., gene 32P*–d(pA)$_4$ complex minus gene 32P*–d(pA)$_4$ complex containing perdeuterated Tyr as represented by the envelope over the filled in peaks. (B) Normal difference spectrum, gene 32P*–d(pA)$_4$ complex minus gene 32P*, both proteins containing all protonated aromatic amino acids.

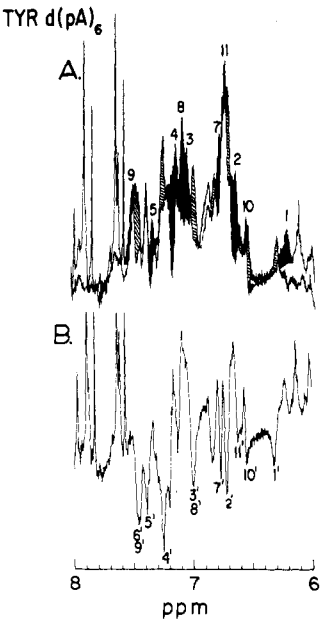

FIGURE 5: Aromatic ^{1}H NMR (500 MHz) difference spectra of gene 32P*–d(pA)$_6$ complex. (A) Spectrum of Tyr resonances from Figure 2A (envelope of hatched peaks) superimposed on the Tyr resonances of the gene 32P*–d(pA)$_6$ complex generated in the same manner, i.e., gene 32P*–d(pA)$_6$ complex minus gene 32P*–d(pA)$_6$ complex containing perdeuterated Tyr as represented by the envelope over the filled in peaks. (B) Normal difference spectrum, gene 32P*–d(pA)$_6$ complex minus gene 32P*, both proteins containing all protonated aromatic amino acids.

makes it difficult to determine if all features of the protonated difference spectrum are accounted for by the Tyr resonance shifts alone.

^{1}H Difference Spectra, Oligonucleotide Complex minus Gene 32P Containing Protonated Tyrosine vs. Deuterated Tyrosine.* Do shifts in Phe and Trp protons occur on oligonucleotide complex formation by gene 32P*? One way of answering this question is to form the standard difference

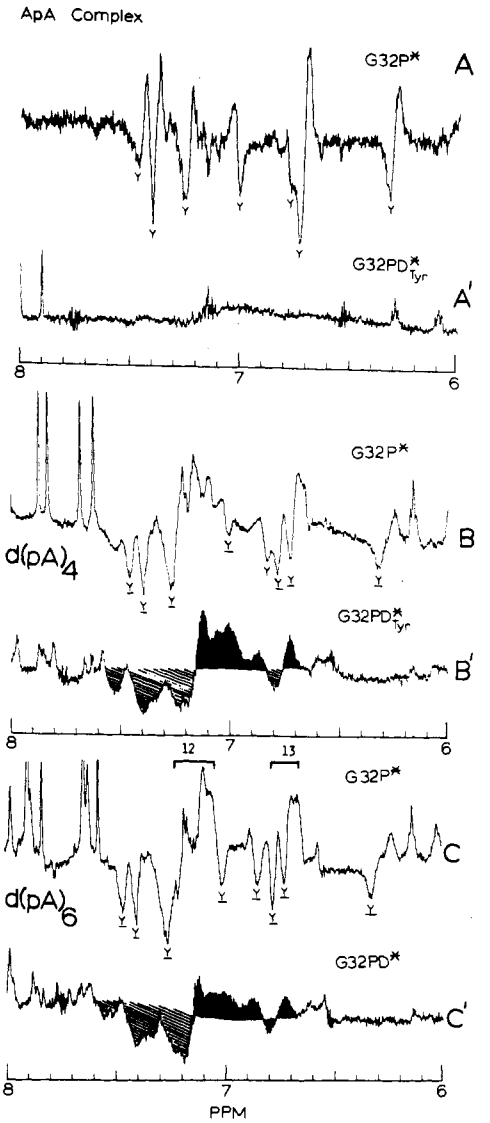

FIGURE 6: Aromatic ^{1}H NMR (500 MHz) difference spectra of gene 32P*–oligonucleotide complexes minus gene 32P* (uncomplexed). (A), (B), and (C) are the difference spectra constructed with the ApA, d(pA)$_4$, and d(pA)$_6$ complexes, respectively. The proteins contained all protonated aromatic amino acids. (A'), (B'), and (C') are the same difference spectra constructed by using proteins all of which contained perdeuterated tyrosine.

spectra (gene 32P*–oligonucleotide minus gene 32P*) but to compare this difference spectra between protonated protein in both complexed and uncomplexed form and perdeuterated tyrosyl protein in both complexed and uncomplexed form. The latter should show only shifted resonances assignable to Phe and Trp. Difference spectra constructed in this manner are compared in Figure 6A–C for the ApA, d(pA)$_4$, and d(pA)$_6$ complexes of gene 32P*. The difference spectrum of ApA with gene 32P*$_{Tyr}$ (Figure 6A') shows that no Phe or Trp resonances are perturbed by the dinucleotide complex formation. This confirms the conclusion from the construction of the Tyr difference spectrum alone (Figure 3) which shows that all the shifts on ApA complex formation can be explained by the interaction of ApA with Tyr residues (resonances marked Y in Figure 6A).

In contrast to ApA, formation of the d(pA)$_4$ complex causes upfield shifts of resonances assignable to Phe or Trp protons (Figure 6B'). The magnitude of the shifts in Figure 6B,C, while not measurable exactly because of overlap, must range from 0.1 to as much as 0.4 ppm. These Phe and Trp reso-

nances, 7.2–7.6 and 6.4–6.6 ppm, overlap features of the standard difference spectra already assigned to Tyr, clearly illustrating the difficulty of interpreting the spectra of the protonated protein alone. Areas of upfield shift (peaks 12 and 13) not previously assigned to Tyr, however, are now clearly attributable to the Phe or Trp protons. These Phe or Trp residues interact with the tetranucleotide and undergo no further shifts when the hexanucleotide interacts with gene 32P* (Figure 6C′).

Participation of Phe Residues in Oligonucleotide Complex Formation with Gene 32P As Determined by Perdeuteration.* Despite the obvious sensitivity of the various difference spectra to small changes in the nature of the oligonucleotide complex and to the magnitude of upfield shifts of specific aromatic resonances as illustrated in the spectra above, these difference spectra are the product of the subtraction of envelopes containing a very large number of resonances; hence, some caution must be exercised concerning the introduction of artifacts by failure to null resonances or the introduction of spurious nonspecific signals by not precisely matching the spectra from two different samples which by necessity were processed differently (e.g., complexed with nucleotide or perdeuterated). The examination of the Tyr protons by the perdeuterated difference spectra technique illustrated above turned out to be particularly easy because the main resolved features of the envelope of aromatic proton resonances of gene 32P* are due to Tyr resonances as made clear by comparing parts A and C of Figure 1. The Phe and Trp resonances, other than the group near 7.4 ppm, are spread over a large portion of the aromatic spectrum and not clearly resolved (Figure 1B). On the other hand, this feature makes it easy to subtract spectrum 1B from 1A and resolve the Tyr resonances. In contrast, subtracting spectrum 1C from 1A to isolate the Phe resonances requires the matching of many detailed features. While this can be monitored by the nulling of the single upfield Tyr resonance at 6.3 ppm, the assumption must be made that no other changes have occurred in the phasing, the base line or the precise condition of the protein between the two samples each of which has many detailed contours in the spectrum contributed by overlapping signals from tyrosyl protons. Because of this feature we have been more cautious in the approach to identifying the Phe residues.

An example of the isolated Phe proton spectrum generated by subtracting the spectrum of gene 32P* containing perdeuterated Phe (Figure 1C) from the spectrum of the totally protonated gene 32P* and nulling the Tyr resonance at 6.3 ppm is shown in Figure 2B. While we are not certain that we have removed all tyrosyl proton contributions to this spectrum, it is significant that complexation of ApA with gene 32P*, followed by generation of the analogous "Phe alone" difference spectrum, shows no significant changes in the Phe envelope (Figure 7A). Any difference seems to be accounted for by base-line problems between the two spectra. This confirms the conclusion that no Phe residues were perturbed in the ApA complex, arrived at independently by difference spectra generated by the perdeuterated Tyr protein (Figure 6A′). The lack of shift in the Phe spectrum (Figure 7A) is in marked contrast to the significant shifts in the Tyr difference spectrum induced by ApA (Figure 3A). Thus, while the chemical shift dispersion of the Phe protons is great and the resonances overlap the Tyr signals, it does appear possible to separate them by the perdeuterated difference spectral technique.

In contrast to ApA, $d(pA)_4$ complex formation induces a number of shifts in the Phe proton spectrum, peaks a–e in

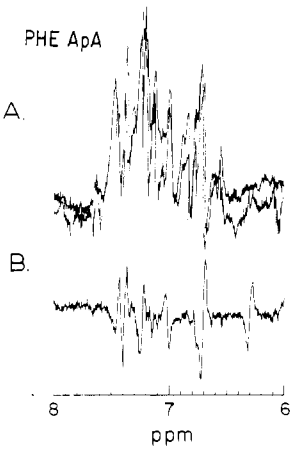

FIGURE 7: Aromatic ^1H NMR (500 MHz) difference spectra of the gene 32P*–ApA complex. (A) Spectrum of Phe resonances from Figure 2B superimposed on the Phe resonances of the gene 32P*–ApA complex generated in the same manner, i.e., gene 32P*–ApA complex minus gene 32P*–ApA complex containing perdeuterated Phe. Both Phe alone spectra are almost identical except for some variation in base line. (B) Normal difference spectrum as shown in Figure 3B.

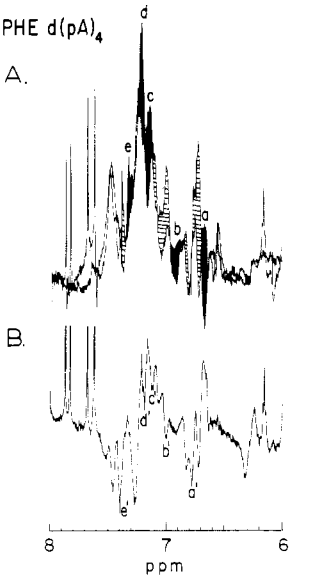

FIGURE 8: Aromatic ^1H NMR (500 MHz) difference spectra of gene 32P*–$d(pA)_4$ complex. (A) Spectrum of Phe resonances from Figure 2B (envelope of hatched peaks) superimposed on the Phe resonance of the gene 32P*–$d(pA)_4$ complex generated in the same manner; i.e., gene 32P*–$d(pA)_4$ complex minus gene 32P*–$d(pA)_4$ complex containing perdeuterated Phe as represented by the envelope over the filled-in peaks. (B) Normal difference spectrum as shown in Figure 4B. The letters indicate peaks assigned to Phe. Primed letters indicate the position of these peaks in the uncomplexed protein as do the primed numbers in the case of Tyr proton signals. See text for the complete description of the numbering system.

Figure 8A. These shifted signals are clustered in the region 6.9–7.3 ppm with an isolated signal near 6.7 ppm, the same region indicated for the oligonucleotide-shifted "Phe + Trp" resonances deduced from the spectrum of the perdeuterated tyrosyl protein (Figure 6), showing that a substantial portion of the shifted resonances in Figure 6B must represent Phe rather than Trp protons.

Assignment of Oligonucleotide-Shifted Tyrosyl Proton Resonances to Tyr 115 of Gene 32P.* In view of the apparent ring-current shifts induced in the ring protons of a number of Tyr residues of gene 32P* by oligonucleotide binding, the sequence of the protein from residues 72 to 116, beginning with Trp^{72}-Tyr^{73} and ending with Tyr^{115}–Trp^{116} and containing six

Tyr residues, seemed a region of the polypeptide chain that might be near the DNA-binding surface as we previously suggested (Prigodich et al., 1984). This sequence is

-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]-

and in order to assign the specific Tyr protons in the ^{1}H NMR spectrum to specific Tyr residues, we have initiated a series of site-directed mutations of these Tyr residues to assign specific proton resonances by generating the difference spectrum of the native minus the mutant protein. The identified resonances can then be compared to those previously identified as shifting on oligonucleotide complex formation. The first example tested was Tyr-115 → Ser-115. The difference spectrum of the native gene 32P* minus gene 32P*$_{Ser\text{-}115}$ is shown in Figure 9. The two major crosshatched resonances (absent in the mutant) are at 7.25 and 6.75 ppm and correspond to 2,6- and 3,5-protons of a tyrosyl represented by negative peaks 4′ and 2′ (peaks 4 and 2 in Figures 4A and 5A) which are present in all the oligonucleotide difference spectra of Figures 3–5 and clearly shift upfield on oligonucleotide complex formation. The chemical shifts of these two resonances also correspond exactly to the tyrosyl previously labeled Tyr-2, which we had identified as being involved in complex formation by NOE difference spectra for the gene 32P*–d-$(pA)_4$ complex (Prigodich et al., 1984).

In addition to these two major Tyr peaks in the mutant difference spectra, there are clearly several other difference peaks that are probably significant even if one allowed some noise due to failure to precisely match spectra. The aliphatic region has been nulled; hence, other aromatic side chain(s) must move on the Tyr-115 → Ser-115 change. There is nothing specific about these additional resonances, since they overlap both the Tyr and Phe resonances and could also be Trp (see Discussion).

Numbers of Tyrosyl and Phenylalanyl Protons of Gene 32P Shifted by Oligonucleotide Complex Formation.* By drawing in a normal line shape for the shifted resonances as determined by the perdeuteration difference spectra method (black regions in Figures 3–5 and 8) and using a cut and weigh method of integration, a rough estimate of the number of protons shifted in each oligonucleotide complex of gene 32P* can be made. By use of the most upfield-shifted tyrosyl 3,5-proton as reference, the results in the case of Tyr are 8, 16, and 20 protons shifted by complex formation with ApA, d-$(pA)_4$, and d$(pA)_6$, respectively. Hence, the dinucleotide appears to interact with two Tyr residues, the tetranucleotide with four Tyr residues, and the hexanucleotide with five Tyr residues. Similar integration of the shifted peaks in the Phe spectrum of the d$(pA)_4$ complex shows the tetranucleotide to induce shifts in 10 Phe protons. Hence, the tetranucleotide perturbs two Phe residues in addition to the four Tyr residues. As noted previously, the dinucleotide shifts no Phe protons and the hexanucleotide shifts no Phe protons in addition to those shifted by the tetranucleotide (Figure 6).

Discussion

The present more highly resolved difference spectra made possible with the gene 32P* proteins containing perdeuterated Tyr and Phe residues make clear that both Tyr and Phe residues are involved in oligonucleotide binding. The majority of the resonances from the ring protons of the Tyr and Phe residues involved shifts upfield on nucleotide complex formation (Figures 3–6), suggesting that the base rings on the incoming nucleotide overlap those of Tyr and Phe side chains in the binding groove. These ring-current shifts, however, have maximum $\Delta\delta$ values of 0.2–0.3 ppm which are substantially less than the ~1 ppm upfield shifts that might be expected from an intercalation model with ring to ring-stacking distances of ~3.4 Å (Giessner-Prettre & Pullman, 1976; Giessner-Prettre et al., 1976). A more probable model than the "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 distance and angle and can give a wide range of magnitudes as well as be both upfield or downfield.[3] Both phenomena are observed in Figures 3–5.

While there is still significant overlap of resonances in the perdeuterated difference spectra (Figures 3–5 and 8), it is considerably easier to estimate the number of protons involved by assuming reasonable line shapes. By use of a cut and weigh method, the shifted resonances assignable to tyrosyl protons appear to account for five Tyr residues, two of which are shifted by the dinucleotide, four by the tetranucleotide, and five by hexa- or longer nucleotides. Hence, the best estimate is that the DNA binding groove of gene 32 protein contains five Tyr residues and two Phe residues, the tyrosyl and phenylalanyl side chains being interleaved at the region of the binding surface occupied by residues 4–6 of a hexa- or longer nucleotide. In the earlier study on gene 5 protein a dinucleotide carrying a 5′-phosphate dianion binds preferentially at one end of the binding site (O'Connor & Coleman, 1983). This does

[3] The upfield shifts of aromatic proton resonances induced by tetranucleotide binding to gene 5P, similar in magnitude to those observed here for gene 32P*, are not large enough to be modeled by intercalation at a ring to ring distance of 3.4 Å. A model implying the latter type of intercalation for all nucleotide residues bound to gene 5P has been presented by McPherson and Brayer (1985) on the basis of the crystal structure of the unliganded protein combined with chemical modification and NMR data of the complexes in solution. While motion of the bases relative to the aromatic rings in the solution complex may modulate stacking distances, the NMR data on gene 5P obtained thus far do not prove van der Waals contact between the aromatic rings of the protein and the bases of the nucleotide. The shifts observed for the gene 5P–nucleotide complexes, however, are nucleus independent [$\Delta\delta(^1H) \simeq \Delta\delta(^{19}F)$] and decrease as expected on shifting the bases from A to T, supporting their assignment to ring-current shifts. Demonstration in the case of both gene 5P and gene 32P* of NOE's between H8 or H2 protons of adenine and the specific aromatic amino acids involved in complex formation would be a highly desirable confirmation of the base-aromatic ring stacking model. Employing standard conditions for signal irradiation in one-dimensional experiments, we have previously failed to demonstrate the expected weak NOE's in the complexes of both gene 5 and gene 32P*. Using 2D experiments and paying careful attention to the exchange conditions, we have been able recently to observe weak but specific H8-aromatic proton NOE's in gene 5P–oligonucleotide complexes (G. C. King and J. E. Coleman, unpublished observations). Similar examination of gene 32P* complexes should allow further clarification of the base positions relative to the aromatic side chains. In a previous publication (Prigodich et al., 1984) we briefly explored the dependence of the upfield shifts on the A to T substitution in gene 32P*–oligonucleotide complexes. While generally smaller for the T series, supporting the origin of the shifts as due to ring currents, precise characterization of individual shifts on the A to T substitution are better done on the series of mutants that allow complete assignment of the 2,6- and 3,5-protons of each residue. Hence, detailed studies await the completion of the series of site-directed mutations.

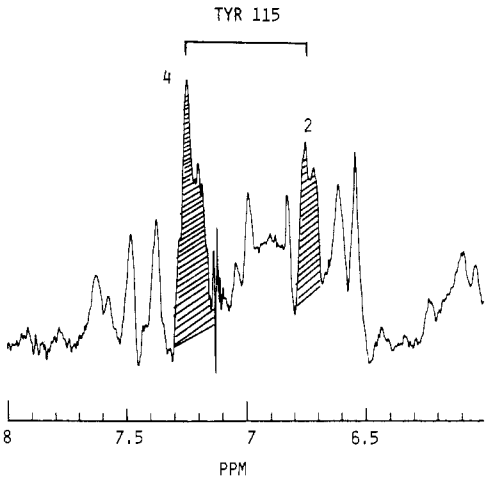

FIGURE 9: Aromatic ^{1}H NMR (500 MHz) difference spectrum, gene 32P* (wild-type) minus gene 32P* (Ser-115 mutant). Peaks 2 and 4 in the difference spectrum have the same chemical shifts as the peaks marked 2′ and 4′ in the previous spectra and assigned to Tyr. These peaks shift to positions marked 2 and 4 in the spectra of the various nucleotide complexes.

not appear to be the case with gene 32P* although ApA without the 5′-phosphate does appear to have a unique binding site which perturbs two Tyr residues (Figure 3). We are not sure where this is located along the groove.

One of the Tyr residues is clearly Tyr-115 as shown by the difference spectra of the mutant protein (Figure 9). The protons of Tyr-115 move upfield by ~0.15 ppm on formation of the complex with $d(pA)_6$ (data not shown).[3] This suggests the possibility that Tyr-73, -84, -92, -99, and -106 may include at least some of the other four Tyr residues involved in nucleotide binding. Their closeness to Tyr-115 in the sequence does not prove proximity in the three-dimensional structure, although the spacing 11, 8, 7, 14, and 7 residues would allow considerable lattitude for folding. Since there are only two more Tyr (137 and 186) in the whole sequence, at least two others in the sequence from Tyr-73 to Tyr-115 must be involved in nucleotide binding. The two Tyr residues outside this sequence do occur in Phe-containing sequences, Phe-Lys-Tyr^{137}-Arg-Phe and Phe-Ser-Asn-Tyr^{186}-Asp-Glu-Leu-Phe; hence, at least one of these Tyr residues may participate in nucleotide binding. In the present paper we have presented data from di-, tetra-, and hexanucleotide complexes. Our previous study extended to complexes with octa- and decanucleotides and concluded that the DNA binding groove of gene 32P* accommodated approximately eight nucleotide residues (Prigodich et al., 1984). From our present work it became apparent that all aromatic protons undergoing shifts on oligonucleotide binding had done so by the time the oligonucleotide had reached six residues. Extension to 8 and 10 residues simply increased the magnitude of the upfield shifts. Hence, the additional interactions between residues 7 and 8 of an octanucleotide and the gene 32P* must not involve additional aromatic side chains, but do alter the conformation all along the chain.

It has been difficult to prove the participation of Trp side chains in nucleotide binding. We have prepared a gene 32P* containing perdeuterated Trp, and the difference spectra did not convincingly demonstrate shifted Trp resonances on oligonucleotide binding. The difference spectra of the Phe residues alone constructed from the perdeuterated Phe protein (Figure 8) show that Phe resonances alone appear to account for most of the shifted resonance in Figure 6 (the latter would include both Phe and Trp). The difference spectra formed with gene 32P* containing perdeuterated Tyr and Phe residues provide much stronger evidence than previously that the gene 32P DNA binding surface contains five Tyr and two Phe residues. Thus, there are two single-stranded binding proteins, gene 5 and gene 32, that use interactions with aromatic side chains to form part of a single-stranded DNA binding surface. Electrostatic interactions between the sugar phosphate backbone and positively charged residues are another part (Anderson et al., 1975; Anderson & Coleman, 1975; Prigodich et al., 1984). While structural studies of more examples of single-stranded binding protein–nucleotide complexes are needed, this may represent a general binding mechanism for single-stranded DNA binding proteins.

Registry No. $d(pA)_4$, 15279-64-8; $d(pA)_6$, 24512-53-6; ApA, 2391-46-0; L-Tyr, 60-18-4; L-Phe, 63-91-2.

REFERENCES

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

Anderson, R. A., & Coleman, J. E. (1975) *Biochemistry 14*, 5485–5491.

Anderson, R. A., Nakashima, Y., & Coleman, J. E. (1975) *Biochemistry 14*, 907–917.

Bittner, M., Burke, R. L., & Alberts, B. M. (1979) *J. Biol. Chem. 254*, 9565–9572.

Coleman, J. E., & Armitage, I. M. (1978) *Biochemistry 17*, 5038–5045.

Coleman, J. E., & Oakley, J. L. (1980) *CRC Crit. Rev. Biochem. 7*, 247–289.

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

Giessner-Prettre, C., & Pullman, B. (1976) *Biochem. Biophys. Res. Commun. 70*, 578–581.

Giessner-Prettre, C., Pullman, B., Borer, P. N., Kan, L.-S., & Ts'o, P. O. P. (1976) *Biopolymers 15*, 2277–2286.

Hosada, J., & Moise, H. (1978) *J. Biol. Chem. 253*, 7547–7555.

Jensen, D. E., Kelly, R. C., & von Hippel, P. H. (1976) *J. Biol. Chem. 251*, 7215–7228.

Kelly, R., Jensen, D., & von Hippel, P. (1976) *J. Biol. Chem. 251*, 7240–7250.

Krisch, H. M., & Selzer, G. B. (1981) *J. Mol. Biol. 148*, 199–218.

Krisch, H. M., & Allet, B. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4937–4941.

Mark, C. F., Lu, S. D., Creasey, A. A., Yamamoto, R., & Lin, L. S. (1984) *Proc. Natl. Acad. Sci. U.S.A. 82*, 5662–5666.

McPherson, A., & Brayer, G. D. (1985) in *Biological Macromolecules and Assemblies* (McPherson, A., & Jurnak, F., Eds.) Vol. 2, pp 323–392, Wiley, New York, NY.

McPherson, A., Jurnak, F. A., Wang, A., Kolpak, F., Rich, A., Molineux, I., & Fitzgerald, P. (1980) *Biophys. J. 32*, 155–173.

O'Connor, T. P., & Coleman, J. E. (1983) *Biochemistry 22*, 3375–3381.

Prigodich, R. V., Casas-Finet, J., Williams, K. R., Konigsberg, W., & Coleman, J. E. (1984) *Biochemistry 23*, 522–529.

Rosenberg, M., Ho, Y., & Shatzman, A. (1983) *Methods Enzymol. 101*, 123–138.

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