Functional Residues and Mechanism of Binding of DNA "Melting" Proteins

Richard Arthur Anderson
Yale University.
FUNCTIONAL RESIDUES AND MECHANISM OF BINDING
OF DNA "MELTING" PROTEINS

A DISSERTATION
PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF
YALE UNIVERSITY
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
RICHARD ARTHUR ANDERSON
MAY 1976
ABSTRACT

FUNCTIONAL RESIDUES AND MECHANISM OF BINDING
OF DNA "MELTING" PROTEINS

Richard Arthur Anderson

Yale University, 1976

The binding of gene 5 protein from bacteriophage fd, gene 32 protein from bacteriophage T4, and the DNA-unwinding protein from E. coli to poly [d(A-T)], fd DNA, and poly (A) is accompanied by reversal in the signs of the ellipticity bands of the nucleic acid chromophores from 250 to 290 nm. Maximal CD changes occur at protein to nucleotide ratios of 1:4, 1:11, and 1:14 respectively. These CD changes have been used to assay alterations in binding of the proteins to DNA accompanying changes in the ionic environment and subsequent to chemical modification of the proteins. Divalent cations dissociate the protein-fd DNA complexes at less than 0.1 M, while monovalent cations are less effective. Dissociation of the complexes is more difficult in the order (E. coli protein > gene 32 protein>gene 5 protein). All cations are more effective in dissociating the complexes with poly [d(A-T)]. Far ultraviolet CD shows the secondary structure of the E. coli protein and gene 32 protein to be similar with ~20% α, 20% β, and 60% random coil, while CD of gene 5 protein is dominated by a large peak at 228 nm due to interacting tyrosyl chromophores. Tetranitromethane nitrates 3 of the 5 tyrosyl residues of gene 5 protein and reduces the binding affinity of the protein for fd DNA by ~100-fold. The same reaction nitrates 5 of the 9 tyrosyls of gene 32 protein, without alteration of secondary
structure and prevents DNA binding. The *E. coli* protein does not react with tetranitromethane. Prior complexing with an excess of tetranucleotides or fd DNA prevents nitration of the gene 5 protein; fd DNA complex formation also protects gene 32 protein from nitration. The $^{19}$F NMR spectrum of gene 5 protein labelled with m-fluorotyrosine shows 3 surface and 2 buried tyrosyls consistent with the nitration results. Up-field chemical shift of the NMR resonances of the 3 surface fluorotyrosyls of gene 5 protein upon complex formation with tetranucleotides is consistent with intercalation between nucleotide bases. Intercalation may explain the increased inter-nucleotide spacing calculated from electron micrographs of the gene 5 and gene 32 protein-DNA complexes as opposed to the relatively condensed structure representing the DNA complex of the *E. coli* protein which lacks reactive surface tyrosyls. Acetylation of the gene 5 protein with N-acetyl-imidazole prevents complex formation, acetylating $\varepsilon$-NH$_2$ groups of the 6 lysyl residues of the protein and phenolic OH groups of 3 tyrosyl residues (the same tyrosyls which are nitrated). Removal of O-acetyl groups with hydroxylamine does not restore binding to DNA. Acetylimidazole acetylates surface lysyl residues in the gene 5 protein-fd DNA complex and dissociates it. The CD of acetylated and nitrated gene 5 proteins is not significantly altered. In contrast maleic anhydride reacts with the 7 amino groups of the protein and changes the secondary structure to one similar to that present in 6 M guanidine HCl. The single SH group of native gene 5 protein does not react with Ellman's reagent, but it reacts rapidly with one Hg$^{2+}$ ion which unfolds the protein; fd DNA prevents reaction with Hg$^{2+}$. Electrostatic forces may be as important as hydrogen bonding in maintaining the native structure.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
of the gene 5 protein. Trypsin treatment of gene 32 protein cleaves
~50 amino acids, but the DNA-binding properties of the native protein are
retained, while CD suggests a slight conformational change accompanying
cleavage. This cleavage occurs spontaneously in extracts of T4-
infected bacteria. A model of gene 5 protein structure is presented
based on a secondary structure showing large regions of β-structure as
predicted by the method of Chou and Fasman with tertiary folding
dictated by the interactions of the tyrosyl, lysyl and sulfhydryl
residues with a tetranucleotide. Lysyl groups, exposed in both the
free protein and the DNA complex appear to be of prime importance in
DNA binding through electrostatic interactions with phosphate groups
of DNA. Three tyrosyls in a "stacked" structure probably intercalate
between DNA bases. Preliminary results of NMR studies designed to
test aspects of the model are described.
ACKNOWLEDGEMENTS

I cannot adequately express my gratitude to Dr. Joseph E. Coleman for his help and direction, both in thought and deed, with this work. For sharing many insights and a calmer perspective in many trying times, I wish to thank Dr. Jan Chlebowski.

I would like to thank Dr. Y. Nakashima for teaching me about peptide analysis and Dr. G. Bourgignon for performing electron microscopy.

I wish to express my appreciation to Dr. W. Konigsberg and Dr. D. Crothers, members of my research committee, for their comments and criticisms; and also to Dr. Konigsberg for the use of his amino acid analyzer.

The fluorine nuclear magnetic resonance spectra were done in conjunction with W. Hull and Dr. B. Sykes at Harvard University. The phosphorus and proton nuclear magnetic resonance results were obtained in collaboration with Dr. I. Armitage.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>vi</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>ABBREVIATIONS</th>
<th>viii</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>CHAPTER I: INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteriophage T4 Gene 32 Protein</td>
<td>2</td>
</tr>
<tr>
<td>E. coli DNA-Unwinding Protein</td>
<td>7</td>
</tr>
<tr>
<td>Bacteriophage fd Gene 5 Protein</td>
<td>12</td>
</tr>
<tr>
<td>Other DNA-Binding Proteins</td>
<td>18</td>
</tr>
<tr>
<td>Approach to the Study</td>
<td>22</td>
</tr>
</tbody>
</table>

| CHAPTER II: MATERIALS AND METHODS | 24 |

| Bacterial and Viral Strains | 24 |
| Chemicals | 24 |
| Media and Buffers | 25 |
| Preparation of DNA-Cellulose | 26 |
| Preparation and Purification of Gene 5 Protein | 26 |
| Preparation of m-Fluoro-tyrosyl Gene 5 Protein | 28 |
| Preparation and Purification of Gene 32 Protein | 28 |
| Preparation and Purification of E. coli DNA-Unwinding Protein | 30 |
| Polyacrylamide Gel Electrophoresis | 31 |
| Purification of fd DNA | 31 |
| Preparation of Oligonucleotides | 33 |
| Chemical Modifications of Proteins | 34 |
| Amino Acid and Peptide Analyses | 35 |
| Absorption Spectra | 36 |
| Circular Dichroism | 36 |
| Magnetic Circular Dichroism | 36 |
| 19F Nuclear Magnetic Resonance Spectra | 36 |
| Mercury Analyses | 36 |

| CHAPTER III: RESULTS | 37 |

| CD of Gene 5 Protein-Nucleic Acid Complexes | 37 |
| Effects of Cations on Gene 5 Protein-DNA Complexes | 39 |
| Effects of Chemical Modification of Gene 5 Protein on the Formation of DNA Complexes | 41 |
| Reaction of Gene 5 Protein with N-Acetylimidazole | 41 |
| Reaction of Gene 5 Protein with Tetranitromethane | 45 |
| Characterization of Nitrated Gene 5 Protein | 45 |
| Protection of Gene 5 Protein Against Nitration | 52 |
Localization of the Nitrated Tyrosyl Residues in the Primary Structure of Gene 5 Protein ................................... 53
Characterization of Crosslinked Gene 5 Nitrated Protein Oligomers .................................................... 55
Reduction of Nitrated Gene 5 Protein ............................................................................................................ 56
NMR Spectrum of Fluorotyrosyl Gene 5 Protein ....................................................................................... 56
Production of Additional Fluorine-Labelled Proteins ............................................................................... 61
Maleylation of Amino Groups in Gene 5 Protein ......................................................................................... 61
Chemical Modification of the Single Sulfhydryl of Gene 5 Protein .......................................................... 62
Conformation of the Chemically Modified Gene 5 Proteins ................................................................... 65
Structure of the E. coli DNA-Unwinding Protein and the Gene 32 Protein ................................................. 68
CD Assay of the Binding of the Gene 32 and E. coli Proteins to DNA ....................................................... 72
Effect of Cations on Nucleic Acid Complexes with E. coli DNA-Unwinding Protein and Gene 32 Protein ... 76
Nitration of Gene 32 Protein and the E. coli DNA-Unwinding Protein ..................................................... 80
Enzymatic Cleavage of Gene 32 Protein ....................................................................................................... 82
CHAPTER IV: DISCUSSION ......................................................................................................................... 87
Three-Dimensional Model of the Gene 5 Protein-Tetranucleotide Complex ........................................... 101
BIBLIOGRAPHY ........................................................................................................................................ 110
APPENDIX ................................................................................................................................................ 122
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Amino Acid Analyses of Nitrated Gene 5 Protein</td>
<td>47</td>
</tr>
<tr>
<td>II</td>
<td>Amino Acid Analyses of Nitrated Gene 5 Protein Proteolytic Fragments</td>
<td>54</td>
</tr>
<tr>
<td>III</td>
<td>Amino Acid Analyses of Gene 32 Protein and <em>E. coli</em> DNA-Unwinding Protein</td>
<td>70</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Polyacrylamide Gel: <em>E. coli</em> DNA-Unwinding Protein</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>CD Spectra of Gene 5 Protein-DNA Complex Formation</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>CD Spectra of Gene 5 Protein-Nucleic Acid Complex Formation</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>CD Spectra of Gene 5 Protein-DNA Complex Dissociation by Cations</td>
<td>42</td>
</tr>
<tr>
<td>5</td>
<td>Absorption Spectra of Acetylated Gene 5 Protein</td>
<td>44</td>
</tr>
<tr>
<td>6</td>
<td>Polyacrylamide Gel: Nitrated Gene 5 Protein</td>
<td>46</td>
</tr>
<tr>
<td>7</td>
<td>Absorption Spectra of Nitro- and Amino-tyrosyl Gene 5 Proteins</td>
<td>49</td>
</tr>
<tr>
<td>8</td>
<td>Binding of Nitro- and Amino-tyrosyl Gene 5 Proteins to DNA</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>Spectrophotometric Hydrogen Ion Titration of Native and Nitrated Gene 5 Protein</td>
<td>51</td>
</tr>
<tr>
<td>10</td>
<td>$^{19}$F NMR Spectrum of Fluorotyrosine Gene 5 Protein</td>
<td>58</td>
</tr>
<tr>
<td>11</td>
<td>$^{19}$F NMR Spectrum of Fluorotyrosine Gene 5 Protein Plus fd DNA</td>
<td>59</td>
</tr>
<tr>
<td>12</td>
<td>$^{19}$F NMR Spectrum of Fluorotyrosine Gene 5 Protein Plus Tetranucleotides</td>
<td>60</td>
</tr>
<tr>
<td>13</td>
<td>Absorption Spectra of Maleylated Gene 5 Protein</td>
<td>63</td>
</tr>
<tr>
<td>14</td>
<td>CD Spectra of Modified Gene 5 Proteins</td>
<td>66</td>
</tr>
<tr>
<td>15</td>
<td>CD Spectrum of Aminotyrosyl Gene 5 Protein</td>
<td>67</td>
</tr>
<tr>
<td>16</td>
<td>Absorption Spectrum of the <em>E. coli</em> DNA-Unwinding Protein</td>
<td>69</td>
</tr>
</tbody>
</table>
Figure 17. CD Spectra of Gene 32 Protein and the E. coli DNA-Unwinding Protein ........................................... 71

Figure 18. CD Spectra of Gene 32 Protein-DNA Complex Formation ....................................................... 74

Figure 19. CD Spectra of E. coli DNA-Unwinding Protein-DNA Complex Formation ........................................... 75

Figure 20. CD Spectra of Gene 32 Protein-Poly (A) Complex Formation ............................................................... 77

Figure 21. CD Spectra of Gene 32 Protein-Poly [d(A·T)] Complex Dissociation by Mg2+ .................................... 78

Figure 22. Dissociation of Gene 32 Protein and E. coli DNA-Unwinding Protein-Nucleic Acid Complexes by Cations ........................................ 79

Figure 23. Absorption Spectra of Nitrated Gene 32 Protein ................................................................. 81

Figure 24. Polyacrylamide Gel: Modified Gene 32 Proteins ................................................................. 83

Figure 25. CD Spectra of Trypsin Cleaved Gene 32 Protein ......................................................................... 85

Figure 26. Primary Structure of Gene 5 Protein: Chou-Fasman Analysis ......................................................... 102

Figure 27. Secondary Structure of Gene 5 Protein: A Model ................................................................. 104

Figure 28. Tertiary Structure of Gene 5 Protein: A Speculative Model ........................................................ 106
ABBREVIATIONS

CD circular dichroism
\( \text{C(NO}_2\text{)}_4 \) tetranitromethane (TNM)
EDTA ethylenediamine tetra-acetic acid
MCD magnetic circular dichroism
NMR nuclear magnetic resonance
RF replicative form
Tris tris (hydroxymethyl) amino methane
SDS sodium dodecyl sulfate
CHAPTER I

INTRODUCTION

Since the introduction of DNA-cellulose affinity chromatography by Alberts et al. (1968), a number of proteins exhibiting specific affinity for single- or double-stranded nucleic acids have been isolated from viral, bacterial, and eukaryotic sources. One class of DNA-binding proteins revealed by this technique has become known as "unwinding" proteins because their tight, cooperative, and preferential binding to single-stranded nucleic acids results in a decrease in the melting temperature of double-stranded nucleotide polymers mixed with the purified proteins. These unwinding proteins bind to single-stranded nucleic acids without regard to base sequence, though there is preferential interaction with the more easily melted A + T-rich regions of double-stranded polymers. As a class they apparently have a stoichiometric role in various phases of cell nucleotide metabolism as opposed to the enzymatic roles of nucleotide polymerases and nucleases and the regulatory roles of proteins such as the bacteriophage lambda and lac-operon repressors. The specific roles of the "unwinding" proteins in the cell apparently vary from system to system yet they are present widely, if not universally, in both eukaryotic and prokaryotic organisms.

Three "unwinding" proteins—the gene 32 protein from bacteriophage T4, the gene 5 protein from the filamentous bacteriophage fd, and the DNA-unwinding protein found in uninfected E. coli—have been the subject of many genetic and functional studies. They are easily purified in
reasonably large quantities making them particularly attractive for the physico-chemical studies which will be described in the body of this presentation. Some of the differences and similarities seen in the physico-chemical and structural aspects of these three proteins correlate with properties of these proteins previously observed using electron microscopy and assays of biological function. The results from these latter techniques are discussed in this introduction.

**Bacteriophage T4 Gene 32 Protein**

The gene 32 protein from phage T4 was the first "unwinding" protein studied by Alberts and co-workers (1968); studies of its in vivo role and in vitro properties have progressed in parallel although the specifics of gene 32 protein's function remain obscure. Epstein et al. (1963) have shown that gene 32 amber mutants fail to synthesize phage DNA; this was confirmed by Kozinski and Felgenhauer (1967) who demonstrated that a gene 32 amber mutant underwent only a single round of replication. DNA replication ceases rapidly upon shifting a culture infected with a T4 mutant carrying a temperature-sensitive gene 32 product to a non-permissive temperature (Alberts et al., 1968; Alberts, 1970; Alberts and Frey, 1970). Tomizawa et al. (1966) observed that the formation of "joint" molecules of DNA (branched structures considered to be intermediates in recombination (Broker and Lehman, 1971)) did not occur in *E. coli* infected with T4 phage containing an amber mutation in gene 32. Thus the gene 32 product appears to be necessary throughout the infectious cycle for both DNA replication and recombination.

Gene dosage experiments demonstrating a positive correlation of both the amount of DNA synthesis and burst size with the ratio of native to amber gene in multiple-infected *E. coli* cells indicates a stoichiometric role for the gene 32 protein (Snustad, 1968; Sinha and Snustad, 1971). This is consistent with the large number of copies of the
protein, approximately 10,000 per cell, made during a normal infection (Alberts, 1970).

The gene 32 protein has been found to bind tightly to all single-stranded DNA's tested but not to single-stranded RNA's (tested at 4° in a sucrose gradient) (Alberts and Frey, 1970). The protein apparently does not bind to T4 DNA in vitro though it has been shown to bind to A + T-rich regions of certain double-stranded DNA's including that from phage lambda (Alberts and Frey, 1970; Delius et al., 1972, Hosoda et al., 1974). This latter property is quite reproducible and the gene 32 protein binding sites coincide remarkably well with regions of the DNA sensitive to limited alkali denaturation; "denaturation maps" of double-stranded, circular SV 40, and polyoma DNA as well as phage lambda DNA have been constructed using gene 32 protein (Delius et al., 1972; Mulder and Delius, 1972; Morrow and Berg, 1972; Yaniv et al., 1974; Monjardino and James, 1975).

The DNA polymerase specified by T4 gene 43 is required for phage DNA synthesis in vivo (DeWaard et al., 1965; Warner and Hobbs, 1967). Huberman et al. (1971) have found that the in vitro activity of this T4 DNA polymerase is stimulated 5-10 fold by the addition of gene 32 protein to the reaction mix when a single-stranded template is used. The purified T4 polymerase alone has in fact been found to require a single-stranded template in vitro; it cannot use double-stranded DNA, such as T4 DNA, for a template even with the introduction of internal "nicks" in the nucleic acid (Goulian et al., 1968; Nossol, 1974). Nossol (1974) has found that the addition of stoichiometric amounts of gene 32 protein to the reaction mix allows the T4 polymerase to use nicked, double-stranded T7 DNA as a template and primer. Thus the gene
32 protein could function to make the DNA a more acceptable template for the polymerase by facilitating strand displacement in the case of double-stranded DNA and by straightening out "hairpin" and other irregular secondary structures in the case of single-stranded templates. The product of T4 polymerase activity in the presence of gene 32 protein is still not a faithful duplication of a double-stranded template, however; the polymerase tends to slip off and begin copying the primer strand as well as the displaced strand to produce randomly branched structures which can be seen with electron microscopy (Nossol, 1974). Perhaps another of the 19 T4 gene functions required for normal DNA replication acts in vivo to prevent this infidelity (Epstein et al., 1963; Nossol, 1974). In contrast to the stimulation of the synthetic function of the T4 DNA polymerase by gene 32 protein the 3' → 5' exonuclease activity of the enzyme is completely inhibited in the presence of stoichiometric amounts of gene 32 protein (Huang and Lehman, 1972).

Huberman et al. (1971) have demonstrated that a weak complex is formed between the gene 32 protein and the T4 DNA polymerase even in the absence of DNA. They saw no significant level of association between E. coli DNA polymerase (pol I) and the gene 32 protein under the same conditions. This parallels the lack of stimulation of the activity of the E. coli polymerase by the gene 32 protein. Thus it can be suggested that the stimulation of the polymerization reaction may not be purely a result of "unwinding" protein-template interaction but may involve some type of protein-protein binding in a larger replication complex.

The role of gene 32 protein in genetic recombination may also be related to its ability to hold single-stranded DNA in an extended
conformation free from the complications of secondary structure. Gene 32 protein promotes both denaturation, lowering the melting temperature of the synthetic double-stranded polynucleotide poly d[A-T] by 40°C, and renaturation of DNA \textit{in vitro} (Alberts and Frey, 1970). Alberts and Frey (1970) showed that the \textit{in vitro} renaturation rate of T4 DNA single-strands at 37° is increased at least 100-fold by the addition of saturating amounts of gene 32 protein (1000-fold increase in the presence of 40 mM Mg^{2+}). Though the melting temperature lowering is seen with all of the DNA-unwinding proteins tested (see below), the increase in renaturation rate is a more specific property of gene 32 protein; it is not seen with the gene 5 protein of bacteriophage fd (discussed in the next section) for example (Alberts et al., 1972). Wackernagel and Radding (1974) have extended these observations by demonstrating that gene 32 protein promotes the \textit{in vitro} formation of the heteroduplex joint molecules between appropriate DNA segments which are considered to be intermediates in the \textit{in vivo} recombination process.

Mosig and Breachkin (1975) have isolated a temperature-sensitive mutant of gene 32 of T4 in which the protein can function to some degree in replication and recombination at the non-permissive temperature but phage progeny production is apparently blocked at the ligase step of recombination. They conclude that in addition to its other functions the gene 32 protein has some specific mode of interaction with the \textit{E. coli} DNA ligase. Krisch \textit{et al.} (1974) have noted the overproduction of non-functional gene 32 peptides during infections with nonsense and temperature sensitive mutants under non-permissive conditions. Based on this they propose that the gene 32 protein regulates its own synthesis.
The native gene 32 protein is a single polypeptide chain with a molecular weight of 35,000 and an axial ratio for an equivalent prolate ellipsoid of about 4 (Alberts et al., 1968). One can estimate a length for the protein of about 120 Å from this data; this is much longer than necessary to cover the 10 single stranded nucleotides which is the apparent number of nucleotide bases per protein monomer in the DNA-protein complex (Alberts et al., 1968). Alberts and Frey (1970) note the probable importance of electrostatic forces in the formation of the complex of gene 32 protein with poly-anionic DNA in view of its sensitivity to dissociation by increasing salt concentration (0.6 M NaCl for complete disruption) but also point out that the protein carries a net negative charge at pH 7.

Many of the observed functional properties of gene 32 protein can be correlated to its ability to preferentially bind tightly and cooperatively to single-stranded nucleic acids. The dissociation constant for this binding of the protein to single-stranded fd DNA has been estimated to be less than $10^{-9}$ M (Alberts and Frey, 1970). The cooperative nature of the interaction has been demonstrated both by the large increase in binding affinity with increasing protein concentration demonstrable with sucrose gradient centrifugation (Alberts and Frey, 1970) and by the clustering of bound protein molecules seen in electron micrographs of mixtures of gene 32 protein with excess single-stranded DNA (Delius et al., 1972). In relation to this Carrol et al. (1972, 1975) have found that gene 32 protein has a pronounced tendency toward self-aggregation. Apparently two separate processes, dimer and higher oligomer formation, with different sensitivities to salt, pH, and temperature occur in solution. The formation of higher oligomers is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
decreased but the amount of dimer formation remains the same in the presence of ten residue or longer nucleotide oligomers; thus the presence of nucleic acid does not obviously increase protein-protein interactions (Carroll et al., 1975).

**E. coli DNA-Unwinding Protein**

The in vivo role of the DNA-unwinding protein isolated from extracts of uninfected *E. coli* is unknown; no mutants with defective production of the protein have been found (Sigal et al., 1972; Weiner et al., 1975). Much work has been done with this protein using in vitro systems, however. Unwinding protein isolated using DNA-cellulose affinity chromatography can form complexes with single-stranded DNA's in vitro which have been visualized using electron microscopy (Sigal et al., 1972). When an excess of DNA is used, strong clustering of the DNA-bound protein molecules is seen with some DNA strands appearing fully saturated while others show no protein bound. This is similar to the phenomena seen by Delius et al. (1972) with the T4 gene 32 protein and indicates the highly cooperative nature of the binding interaction (Sigal et al., 1972). DNA molecules fully saturated with *E. coli* unwinding protein show an approximately 35% decrease in contour length as compared to uncomplexed DNA (Sigal et al., 1972).

Like the gene 32 protein the *E. coli* DNA-unwinding protein binds preferentially to single-stranded nucleic acids with no detectable binding to double-stranded polymers; the addition of the protein to solutions of the latter, however, results in a lowering of their melting temperatures (Sigal et al., 1972). "Denaturation maps" of A + T-rich regions of double-stranded DNA's equivalent to those seen with alkali denaturation and with gene 32 protein binding can be produced.
with the *E. coli* unwinding protein (Sigal *et al.*, 1972; Reed *et al.*, 1975). Binding of the *E. coli* protein to poly (rI) and poly (rU), synthetic polyribonucleotides, has been detected using gel exclusion chromatography; there was no binding to poly (rA), phage R 17 RNA, or poly (dA) detected, however (Sigal *et al.*, 1972; Weiner *et al.*, 1975).

Estimates of the number of copies of the unwinding protein in rapidly dividing *E. coli* cells are in the range of from 300, determined by titration using specific antibodies to the protein by Weiner *et al.* (1975), to 800 (Sigal *et al.*, 1972). This number of copies per bacterium is sufficient to cover large regions of DNA in the six replication forks thought to be active in such cells (Sigal *et al.*, 1972). The protein may function in vivo in a stoichiometric role as a structural element keeping the DNA template strands aligned (Sigal *et al.*, 1972) or a regulatory factor determining the number of initiation points in *E. coli* chromosome replication (Weiner *et al.*, 1975).

It has been suggested that the *E. coli* DNA unwinding protein has a number of functions based on results obtained using *in vitro* DNA replication systems (Weiner *et al.*, 1975; Molineux *et al.*, 1974; Geider and Kornberg, 1974). In the simple case in which the unwinding protein is added to a mixture of template (usually double-stranded DNA with long single-stranded gaps), enzyme, and nucleotide triphosphates, the polymerizing activity of *E. coli* DNA polymerase II is found to be increased while that of *E. coli* DNA polymerases I and III are decreased or unaffected (Molineux *et al.*, 1974; Weiner *et al.*, 1975). Molineux *et al.* (1974) found that the stimulation of DNA polymerase II reached a maximum at a 1:8 molar ratio of protein to nucleotide base and was independent of polymerase concentration; from this they suggested that the primary
stimulatory effect was manifested by the unwinding protein on the template rather than on the polymerase.

Weiner et al. (1975) found a similar increase in polymerase II activity in the presence of the *E. coli* DNA-unwinding protein but this was dramatically overshadowed by the stimulation of the *E. coli* DNA polymerase III holoenzyme (composed of two polymerase III peptides plus 2 copolymerase III* subunits) which they observed (Wickner et al., 1973; Wickner and Kornberg, 1974a). The DNA polymerase III holoenzyme stimulation is optimal at the 1:8 ratio as above, both the rate and extent of synthesis are increased, and the stimulation is specific for the *E. coli* protein as it was not seen with gene 5 protein or gene 32 protein addition (Weiner et al., 1975).

Geider and Kornberg (1974) have reconstructed a system to convert single-stranded filamentous phage DNA to a covalently closed circular duplex form (replicative form). The *E. coli* DNA unwinding protein is required in this system along with five other *E. coli* proteins including RNA polymerase, DNA polymerase III*, copolymerase III*, DNA polymerase I, and DNA ligase. The unwinding protein appears to have at least two roles in this system in limiting and directing the RNA primer initiation to a single unique site and in extension of the RNA-primed template. Spermidine can replace the unwinding protein in supporting the rate and extent of DNA synthesis but does not lead to the production of full length complementary strands initiated at a unique location. Gene 5 protein not only does not replace the *E. coli* unwinding protein but efficiently inhibits the reaction as would be predicted from in vivo results (see below).

The stimulatory effects of the unwinding protein are produced only
in the narrow concentration range of one protein monomer per 5-10 template nucleotide bases. In a related experiment Weiner et al. (1975) show that antibody to the E. coli DNA unwinding protein blocks filamentous phage replication in crude extracts of infected E. coli.

Molineux et al. (1974) have found that in the presence of the unwinding protein DNA polymerase II can also produce a full size complement to virai-strand fd DNA in vitro but the product has a significant number of randomly inserted ribonucleotides. This result is interesting in view of the lack of a known role for DNA polymerase II in vivo (Campbell et al., 1972; Hirota et al., 1972; Tait and Smith, 1974; Tait et al., 1974). DNA polymerase III peptide (present in more active forms such as polymerase III* and holoenzyme) has been genetically implicated in E. coli chromosomal replication (Tait and Smith, 1974; Gefter et al., 1972; Wechsler and Gross, 1971).

Molineux and Gefter (1974), in examining the effects of the E. coli DNA unwinding protein on the exonuclease activities associated with the three known E. coli DNA polymerases, found evidence for the formation of a 1:1 complex between the unwinding protein and DNA polymerase II. The association is disrupted by 0.1 M KCl. There was no binding to polymerases I or III. Correspondingly the exonuclease functions of polymerases I and III were inhibited by the presence of unwinding protein; while the polymerase II exonuclease was fully active with saturating and higher concentrations of protein though inhibited by low concentrations (1:20 protein to base ratios). An interpretation consistent with those observations is that prior complexing with unwinding protein increases the affinity of DNA polymerase II for single-stranded DNA. DNA polymerase III does not bind to single-stranded DNA.
either in the presence or absence of the unwinding protein (Molineux and Gefter, 1974).

The effect of saturating amounts of \textit{E. coli} unwinding protein on transcription of single-stranded DNA templates by \textit{E. coli} RNA polymerase is a decrease both in rate and in length of product for an overall 50-70\% drop in activity (Weiner et al., 1975; Molineux et al., 1974). With a double-stranded template there is no change in transcription activity (Molineux et al., 1974). Wickner and Kornberg (1974b) have purified a form of \textit{E. coli} RNA polymerase from uninfected cells which can, if saturating levels of the \textit{E. coli} DNA unwinding protein are present, distinguish between M13 filamentous phage and \textit{\phi}X174 phage single-stranded DNA in the RNA priming event necessary for initiation of DNA synthesis (Wickner et al., 1972a). The \textit{\phi}X174 DNA priming is suppressed. However, the unwinding protein is necessary for the conversion of \textit{\phi}X174 DNA to its replicative form in the reconstituted system of Wickner and Hurwitz (1974). It is known that the filamentous phage and \textit{\phi}X174 replication mechanisms differ at many stages (Schekman et al., 1972; Wickner et al., 1972b); it is interesting that the unwinding protein requirement is common but apparently has different functional manifestations. The unwinding protein is also necessary for replicative form synthesis in extracts of \textit{E. coli} dnaG and dnaB mutants infected with the \textit{\phi}X174-like phage G4 (Weiner et al., 1975).

Work on the physical chemistry of the unwinding protein and its complex with DNA is at an early stage. Molecular weight determinations have yielded values of 22,000 (Sigal et al., 1972) and 18,500 (Weiner et al., 1975) daltons per polypeptide with the native form of the protein being a tetramer of four identical subunits (Weiner et al., 1975;
Molineux et al., 1974). A bonding stoichiometry of 1 protein monomer per 8 nucleotide bases has been determined by gel filtration (Weiner et al., 1975), by absorption spectroscopy utilizing an inhibition of the hypochromic shift usually seen when Mg\(^{2+}\) is added to single-stranded DNA, and by sucrose gradient sedimentation (Sigal et al., 1972). A nitrocellulose filter assay gave a stoichiometry of 1 protein monomer per \(\sim 17\) nucleotide bases (Weiner et al., 1975). Other interesting physical properties of the unwinding protein include a marked tendency to aggregate at concentrations greater than 1 mg/ml (Molineux et al., 1974; Weiner et al., 1975). The protein is remarkably heat stable; 90% of it survives boiling for 8 minutes (Weiner et al., 1975). Dialysis versus 0.04 M potassium phosphate (pH 7) reversibly precipitates the protein (Sigal et al., 1972). Amino acid analysis is unremarkable except for failure to find any cysteine residues in the protein (Weiner et al., 1975).

**Bacteriophage fd Gene 5 Protein**

The production of single-stranded viral DNA by replicating F-specific filamentous bacteriophages including f1, fd, and M13 has been shown to require the participation of two phage-coded proteins, the products of phage genes 2 and 5 (see Marvin and Hohn, 1969, for review). During the initial stages of phage infection the single-stranded DNA of the infecting phage is converted to a double-stranded replicative form (RF) by host cell enzymes. The gene 2 product, probably a nuclease which specifically nicks the viral strand of the RF, is necessary for the production of about 200 progeny RF (Pratt and Erdahl, 1968; Tseng and Marvin, 1972; Mazur and Model, 1973; Findanian and Ray, 1972; Hohn et al., 1971). Conversion of DNA synthesis from the synthesis of
RF to the synthesis of new single-stranded viral DNA requires the product of gene 5. Experiments with phage carrying a temperature-sensitive mutation in gene 5 have shown that single-stranded viral DNA synthesized at a permissive temperature is quickly converted to a double-stranded form at the non-permissive temperature (Salstrom and Pratt, 1971). After infection with a phage carrying a temperature-sensitive mutation in gene 2 neither double-stranded nor single-stranded DNA is synthesized at the non-permissive temperature. If the temperature is lowered to the permissive temperature 10 minutes after infection, only single-stranded DNA is synthesized; the single-stranded synthesis is coincident with the accumulation of large amounts of gene 5 protein (Mazur and Model, 1973). In normal phage infection, gene 5 protein accumulates in as many as 10^5 copies per cell (Alberts et al., 1972; Oey and Knippers, 1972).

In vivo complexes of gene 5 protein with fd DNA have been isolated from infected cells (Webster and Cashman, 1973; Pratt et al., 1974) and appear in the electron microscopy as cigar-shaped rods thick enough to contain two protein-covered strands of DNA (Pratt et al., 1974). With the exception of some branching, gene 5 protein-fd DNA complexes made by mixing purified protein and DNA in vitro look similar in the electron microscope to the in vivo complexes (Alberts et al., 1972). The amount of branching seen in vitro decreases with low protein to DNA ratios suggesting that artifactual multiple seeding points for the cooperative binding interaction occur in vitro (Alberts et al., 1972). In both complexes there is one mole of protein for every 4-5 moles of DNA base.

Thus it appears that a function of gene 5 protein is stoichiometric
involving complex formation with single-stranded viral DNA and thereby preventing its use as a template for synthesis of the complementary strand.

A plausible model for this function involves the gene 5 protein binding to viral-strand DNA as it comes off a "rolling circle" DNA replication structure (Gilbert and Dressler, 1968; Dressler, 1970) thus inhibiting formation of the enzymatic complex necessary for initiation of DNA synthesis (Oey and Knippers, 1972). It is thought that the gene 5 protein fd DNA complex moves to the cell membrane where the gene 5 proteins are replaced by the virion coat proteins as mature virions are extruded from the intact bacterium. The gene 5 protein is then apparently recycled into complexes with newly replicated DNA molecules (Webster and Cashman, 1973). The gene 5 protein does not form a part of the filamentous phage particle (Marvin and Hohn, 1969); its amino acid sequence (Nakashima et al., 1974a,b) shows no significant homology with that of the coat protein (Asbeck et al., 1969; Nakashima and Konigsberg, 1974).

Studies using a temperature-sensitive dnaB^- strain of E. coli suggest an additional positive role of gene 5 protein in the asymmetric displacement synthesis of single-stranded DNA (Staudenbauer and Hofschneider, 1973). The E. coli dnaB product is required for double-stranded, but not for single-stranded DNA synthesis (Olsen et al., 1972; Staudenbauer and Hofschneider, 1972). Infection of dnaB^- E. coli with wild-type filamentous phage at a temperature which does not permit dnaB gene product to function, results in single-stranded synthesis. However, if phage with a temperature-sensitive gene 5 product is used, single-stranded synthesis is stopped on shifting to the non-permissive
temperature. This implies a positive role for gene 5 protein in DNA replication by filamentous bacteriophage; perhaps akin to that suggested by Oey and Knippers (1972) in which the gene 5 protein enhances the rate of removal of the newly replicated viral DNA in a 5' to 3' manner from the complementary template strand thereby more efficiently exposing the 3' end as a primer for the DNA polymerase.

In vitro the addition of stoichiometric amounts of gene 5 protein to a reconstituted RF DNA synthesizing system halts DNA synthesis (Geider and Kornberg, 1974). This Geider and Kornberg (1974) system employs a form of E. coli DNA polymerase III which is the polymerase found by Staudenbauer et al. (1973) to be involved in filamentous phage RF replication in vivo. It has also been demonstrated that the gene 5 protein inhibits in vitro DNA synthesis by DNA polymerases I and II from E. coli and the DNA polymerase from bacteriophage T7 using a variety of single- and double-stranded template DNA's (Oey and Knippers, 1972). This inhibitory effect was greatest at 37°C while almost no effect was seen at 5°C, a temperature at which gene 5 protein would be less able to "melt" regions of DNA double-strandedness and secondary structure and thus not be able to bind to the nucleic acid. The presence of gene 5 protein inhibits the action of the single-strand specific exonuclease I and completely blocks the activity of the double-strand specific exonuclease III (Oey and Knippers, 1972). Probably this involves two different mechanisms; in the first case blocking progression of the enzyme along the single-strand and in the latter case partially denaturing regions of the DNA substrate thus effectively inactivating the requisite double-stranded substrate.

The cooperative nature of the gene 5 protein-nucleic acid
interaction is graphically illustrated by the clustering of protein molecules in discrete patches along the DNA strand seen in electron micrographs taken of mixtures of protein with a large excess of DNA (Alberts et al., 1972). Using sucrose gradient centrifugation with varying ratios of DNA and gene 5 protein Alberts et al. (1972) estimate that the affinity for protein binding to a site contiguous to a previously bound gene 5 protein is sixty times greater than for an isolated site on the DNA. Oey and Knippers (1972) found that gene 5 protein exists primarily as dimers under low salt conditions but as monomers when the KCl concentration is greater than 0.2 M. In line with this Rasched and Pohl (1974) found that chemical crosslinking of gene 5 protein in solution produced mainly dimers. However complexing with short oligonucleotides (as small as tetra-deoxynucleotides) lead to the formation of large amounts of higher protein oligomers (up to eight) under the same cross-linking conditions (Rasched and Pohl, 1974). Possible explanations include protein conformation changes upon binding to DNA or simply exposure of "sticky" surfaces previously blocked by the isolated protein dimer structures. Day (1973) has found changes in the circular dichroism (CD) spectrum of gene 5 protein upon binding to DNA which may reflect the same phenomena (see below). The possible relationship of these observations to the cooperative DNA binding interaction remains to be explored. The previously mentioned Carroll et al. (1975) find no increase in protein-protein interaction with T4 gene 32 protein in the presence of oligonucleotides.

Gene 5 protein is a single polypeptide chain of MW 9689 containing 87 amino acid residues. The sequence of the protein from fd bacteriophage has recently been determined by Nakashima et al. (1974a,b); it
contains 7 acidic residues, 11 basic residues including 6 lysines, 1
cysteine and 5 tyrosines and no tryptophan. The sequence of the gene 5
protein from the closely related M13 phage has been found to be identi-
cal (Cuypers et al., 1974). The pure protein has been shown to bind
tightly and cooperatively to single-stranded DNA without regard to base
sequence, but not to double-stranded DNA (Oey and Knippers, 1972;
Alberts et al., 1972). Using the observation that DNA complexed to
gene 5 protein is fully hyperchromic it can be demonstrated that the
preferential binding to single-stranded DNA lowers the Tm of a variety
of double-stranded DNA's and synthetic deoxyribonucleotide polymers
by an average of 40° (Alberts et al., 1972; Oey and Knippers, 1972).
Thus gene 5 protein induces the melting of poly[d(A-T)] at room
temperature.

The DNA from the filamentous bacteriophages, in particular fd, is
convenient to use in many DNA replication and binding protein studies
because of its availability in large quantities as unique, covalently-
closed, single-stranded circles (Marvin and Schaller, 1966; Schaller,
1969). In general the nucleotide base composition of fd DNA expressed
in mole per cent - 34.1 T, 21.7 C, 24.4 A, 19.9 G - is not consistent
with Watson-Crick base pairings (Marvin and Hohn, 1969). Schaller et
al. (1969) have however found evidence that a small fraction of the
DNA, approximately 40 nucleotide residues per chain, exists in a double-
stranded form resistant to single-strand specific nuclease; this is
interpreted as a single polynucleotide chain folding back on itself.
One might speculate that this segment might have some roles in the
initiation of transcription and replication (Geider and Kornberg, 1974).
Precise physical chemical measurements on fd DNA indicate a molecular
weight of \(1.90 \times 10^6\) daltons corresponding to \(5740 \pm 210\) nucleotides per molecule (Berkowitz and Day, 1974; Newman et al., 1974).

**Other DNA-Binding Proteins**

Another distinct DNA-binding protein with properties quite similar to those of the aforementioned proteins has been isolated from *E. coli* cells infected with bacteriophage T7. The reports of Reuben and Gefter (1973, 1974) and Scherzinger et al. (1973) agree remarkably well on the properties of the protein though their reports of the molecular weight, 31,000 and 25,000 daltons respectively, are significantly different. The T7 protein binds to single-stranded but not double-stranded DNA and lowers the melting temperature of poly [d(A-T)]. The activity of the T7-induced DNA polymerase using double-stranded DNA with single-stranded gaps as a template-primer is stimulated 5-fold by addition of the T7 DNA-binding protein. Even greater stimulation is observed using a single-stranded DNA template at low temperatures where the T7 polymerase is usually unable to effectively use a single-stranded template. The addition of the T7 DNA-binding protein to reactions catalyzed by *E. coli* DNA polymerases I, II, or III or the T4 DNA polymerase had no effect. In contrast to the intra-system specificity generally found in these systems, the *E. coli* DNA-unwinding protein can (in vitro) stimulate the T7 DNA polymerase to the same extent as the T7 protein. If this is true in vivo it may explain why no conditional-lethal phage mutant (and thus gene assignment) has been found for the T7 DNA-binding protein.

A relatively small (MW 14,000) protein coded for by gene D of φX 174 has been found to be associated with the single-stranded DNA of this phage prior to final coat protein packaging (Weisbecker and Sinsheimer, 1973, 1974).
Although published results are still preliminary, the gene D protein is apparently present in large amounts in \( \Phi X 174 \) infected E. coli and may be analogous to the gene 5 protein in the filamentous phage system. Two proteins of MW 72,000 and 45-48,000 which bind to single-but not double-stranded DNA or to RNA can consistently be isolated in large quantities—there are about \( 10^6-10^7 \) copies per infected cell—from African Green Monkey Kidney and human KB tissue culture cells infected with adenovirus types 2 or 5 (Van der Vliet and Levine, 1973; Van der Vliet et al., 1975; Shanmugam et al., 1975). They are not coat proteins and have no obvious enzymatic function; the larger protein predominates when actively growing cells are infected and the smaller when confluent monolayers are infected (Van der Vliet and Levine, 1973). An adenovirus mutant in which viral DNA synthesis is defective has been isolated in which neither DNA-binding protein can be detected after infections at the non-permissive temperature; this evidence along with DNA-cellulose thermo-elution profiles, which imply that the proteins themselves are temperature-sensitive, and peptide mapping experiments imply that the 48,000 MW protein is a physiological or artifactual digestion product of the 72,000 MW species (Van der Vliet et al., 1975). The observations that these proteins are made early in infection before the start of viral DNA synthesis and are isolated in the same nuclear membrane fraction as newly synthesized viral DNA suggest that these proteins (protein?) are involved in viral DNA replication (Yamashita and Green, 1974; Shanmugam et al., 1975). Thus there is much to suggest a similarity between adenovirus DNA-binding proteins and the T4 gene 32 protein.

Single-stranded DNA-binding proteins have also been isolated as
endogenous components of a number of different types of eukaryotic cells. Hotta and Stern (1971a) have isolated a 35,000 MW non-histone protein from the nuclei of meiotic Lilium cells which binds strongly to single-stranded but not native DNA and increases the rate of renaturation of denatured Lilium DNA. The protein is not found in somatic Lilium cells. A 20,000 MW protein with properties quite similar to those of the prokaryotic DNA-unwinding proteins has been isolated from mitotic cells of the fungus Ustilago maydis by Banks and Spanos (1975). Each mitotic cell contains about 2-3 x 10^5 copies of the protein. The Ustilago protein "melts" poly [d(A-T)] and stimulates the renaturation of phage T7 DNA. It stimulates the initial rate of incorporation by a Ustilago DNA polymerase but inhibits degradation of single- and double-stranded DNA by an Ustilago deoxyribonuclease and by E. coli exonuclease III respectively.

Mammalian sources of "DNA-unwinding" proteins include spermatocytes (Hotta and Stern, 1971b) and calf thymus (Herrick and Alberts, 1973). This latter protein melts double-stranded RNA as well as DNA. The calf thymus protein stimulates the high molecular weight calf thymus DNA polymerase (Herrick and Alberts, 1973). A protein, known as P8, which binds preferentially to single-stranded DNA has been isolated from a variety of human and mouse fibroblast tissue culture cells (Salas and Green, 1971; Tsai and Green, 1973). The large amount of P8 found in fibroblasts (2-3% of total soluble protein) suggests a stoichiometric or structural role for the protein (Tsai and Green, 1973). The amount of P8 is increased in growing cells and in polyoma virus-transformed cells (Tsai and Green, 1973; Melero et al., 1975). Melero et al. (1975) find that P8 is synthesized at a maximal rate prior to DNA synthesis.
in virus-transformed cells and Salas and Green (1971) show that the protein markedly accumulates in thymidine arrested cells (pre S-phase); these results do not necessarily imply any connection of P8 production with DNA synthesis as the protein is made even in resting cells (Tsai and Green, 1973). The localization of P8 protein primarily in cytoplasmic fraction of cell extracts and its relatively weak binding to DNA cellulose (it is found in the 0.2 M salt fraction of DNA cellulose columns as compared to the 0.6-2 M salt fractions for the prokaryotic DNA-unwinding proteins) suggests that P8 may not be the mammalian equivalent of gene 32 protein (Tsai and Green, 1973). In vitro characterization of the P8 protein is still in progress (Melero et al., 1975).

The DNA-unwinding proteins discussed here should not be confused with the "untwisting" or ω proteins isolated from a variety of sources including E. coli, mouse tissue culture cells, Drosophila eggs, and SV-40 infected cells, which remove superhelical turns from covalently closed circular DNA's (Wang, 1971; Champoux and Dulbecco, 1972; Baase and Wang, 1974; Sen and Levine, 1974). The mechanism of ω action is thought to be enzymatic although the substrate DNA's are still covalently closed after the ω protein has functioned to remove the superhelical turns and no transiently nicked intermediates can be isolated (Baase and Wang, 1974). The ω proteins could act as the swivel points postulated in a number of models for DNA replication.

Histones, enzymes, control factors, and many other types of proteins have been found to bind specifically to nucleic acids (see Von Hippel and McGhee, 1972, for review). Among the more interesting of the proteins which interact with DNA are the repressor and activator
proteins, such as the well known lac operon repressor, which by virtue of very specific, high affinity binding to unique sites control gene expression in prokaryotic systems. Evidence is accumulating that similar mechanisms operate in eukaryotic cells; steroid binding proteins which apparently derive a highly specific affinity for chromosomal DNA sites upon binding particular steroid hormones are beginning to be purified and investigated (Yamamoto, 1974; Yamamoto and Alberts, 1974; O'Malley and Means, 1974). The investigation of these systems, however, is far from the stage at which the physico-chemical basis of such unique DNA sequence-protein interaction can be investigated.

Much investigation in recent years has centered on the mechanisms of action of DNA-dependent RNA polymerases (see Chamberlin, 1974, for review). In many prokaryotic systems the RNA polymerase must apparently recognize a specific DNA sequence before binding tightly to the template chromosome and beginning RNA synthesis. The tight binding step may well involve local "melting" or denaturation of the native DNA initiation site.

Knowledge of physico-chemical aspects of the interaction of "unwinding proteins" with DNA should illuminate aspects of DNA replication as well as protein structure and function relationships. In addition it is hoped that insights obtained through the investigation of the currently more accessible "unwinding protein" - DNA systems will aid in elucidating approaches to the study of the more complex protein-nucleic acid interactions.

Approach to the Study

This investigation of structural aspects of the interaction of the T4 gene 32 protein, the fd gene 5 protein, and the E. coli DNA-
unwinding protein with a variety of nucleic acid polymers has involved
the use of a variety of spectroscopic probes and chemical modification
techniques. Day (1973) has examined the absorption and circular dich-
roism spectra of gene 5 protein and its complex with fd DNA and found
characteristic changes in the spectra of the bound molecules versus
their uncombined states. He also noted that the near ultraviolet CD
of native gene 5 protein appears to reflect primarily the ellipticity
of the tyrosyl chromophores whose dissymmetry changes significantly
on complex formation (Day, 1973). It has been demonstrated in this
laboratory that large changes in the CD of double-stranded nucleic acid
polymers as well as of single-stranded native DNA's accompany the
formation of complexes not only with the gene 5 protein but also with
the gene 32 and E. coli DNA-binding proteins. The present study
has utilized the CD changes occurring in the polymers as well as changes
occurring in the CD spectrum of the protein on complex formation to
monitor the interaction of gene 5 protein with DNA as a function of
cation concentration and species, pH, and chemical modification of
specific amino acid side chains of the protein. Nuclear magnetic
resonance (NMR) and absorption spectroscopy have proven useful in
confirming and extending the findings of involvement of specific amino
acid residues in DNA binding for this series of proteins.
CHAPTER II

MATERIALS AND METHODS

Bacterial and Viral Strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Phenotype</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli K37</td>
<td>Hfr, su&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Dr. M. Oeschger</td>
<td></td>
</tr>
<tr>
<td>E. coli AT2741</td>
<td>Hfr, tyr&lt;sup&gt;-&lt;/sup&gt;, thi&lt;sup&gt;-&lt;/sup&gt;, rel&lt;sup&gt;-&lt;/sup&gt;, -</td>
<td>Dr. B. Low</td>
<td>Taylor &amp; Trotter, 1967</td>
</tr>
<tr>
<td>E. coli KA197</td>
<td>Hfr, phe&lt;sup&gt;-&lt;/sup&gt;, thi&lt;sup&gt;-&lt;/sup&gt;, rel&lt;sup&gt;-&lt;/sup&gt;, -</td>
<td>E. coli Genetic Stock Center</td>
<td></td>
</tr>
<tr>
<td>E. coli BW46</td>
<td>xonA&lt;sup&gt;-&lt;/sup&gt;, su&lt;sup&gt;-&lt;/sup&gt;</td>
<td>Dr. E. Niles</td>
<td>Kushner et al, 1972</td>
</tr>
<tr>
<td>E. coli KL239</td>
<td>su&lt;sub&gt;I&lt;/sub&gt;&lt;sup&gt;+&lt;/sup&gt;, pro&lt;sup&gt;-&lt;/sup&gt;, met&lt;sup&gt;-&lt;/sup&gt;, trp&lt;sup&gt;-&lt;/sup&gt;, his&lt;sup&gt;-&lt;/sup&gt;, str&lt;sup&gt;-&lt;/sup&gt;, mal&lt;sup&gt;-&lt;/sup&gt;, lac&lt;sup&gt;-&lt;/sup&gt;</td>
<td>Dr. B. Low</td>
<td></td>
</tr>
<tr>
<td>fd bacteriophage</td>
<td>wild type</td>
<td>Dr. D. Marvin</td>
<td></td>
</tr>
<tr>
<td>T4 bacteriophage</td>
<td>am33, am55</td>
<td>Dr. E. Niles</td>
<td>Gold et al., 1973</td>
</tr>
</tbody>
</table>

Chemicals

Poly[d(A·T)], calf thymus DNA, tetratinomethane, N-acetylimidazole, and maleic anhydride were obtained from Sigma Chemical Co. (St. Louis, Mo.). Poly(A) and oligo[d(pA)<sub>4</sub>] were obtained from P.L. Biochemicals, Inc. (Milwaukee, Wis.). Guanidine-HCl and ammonium sulfate were Schwarz-Mann (Orangeburg, N.Y.) ultra-pure and enzyme grade respectively. 1-[<sup>14</sup>C]-Acetylimidazole was obtained from Calatonic (Los Angeles, Ca.). Trypsin (TPCK), chymotrypsin (CKI), pancreatic deoxyribonuclease I (D), and carboxypeptidase A (COADFP) came from Worthington Biochemical Corp.
m-Fluoro-DL-tyrosine was purchased from Aldrich Chem. Co. (Cedar Knolls, N.J.), polyethylene glycol (Carbowax 6000) from Union Carbide (New York, N.Y.), and cesium chloride from Kerr-McGee Chem. Co. (West Chicago, Ill.). All other chemicals were reagent grade.

**Media and Buffers**

All cultures were generally grown in the following media unless specified:

**Broth**

- 16 gm Difco Bacto-tryptone
- 10 gm Difco Yeast extract
- 5 gm NaCl

made to 1 liter with water and autoclaved 90 minutes.

**Plaque Assay**

Agar mix contained 10 gm/1 Difco Bacto-tryptone and 5 gm/1 NaCl with 1% Difco Bacto-agar powder for bottom agar. Top ("soft") agar was 0.7% agar for fd and 0.4% agar for T4 plaque assays. The plaque assays were done using standard dilution techniques and 8-12 hr incubation of plates at 37°.

The initial stages of all DNA-binding protein purifications involving DNA-cellulose chromatography employed the following buffer:
Preparation of DNA-Cellulose

Single-stranded DNA-cellulose was prepared by the technique of Alberts and Herrick (1971) using Whatman CF 11 cellulose powder and calf thymus DNA (which seemed to give better results than salmon sperm DNA). After equilibrating overnight in 10 mM K₂HPO₄ and 1 mM EDTA at a concentration of 2 mg/ml, the DNA solution was heated to 100° for 15 minutes and then rapidly quenched by pouring into large, silicon-coated, iced petri dishes. Tris-HCl, pH 7.4, was added to 20 mM. Cellulose powder was then added very slowly with mixing until a reasonably dry, lumpy mix was obtained. This was dried under gauze at room temperature for several days before grinding gently to a powder followed by extended lyophilization. The powder could be stored for several months as a dry powder. A minimum of 1 mg of DNA was tightly bound per gm of wet cellulose using this procedure. In general a 20 x 2 cm column equilibrated in standard buffer was used for protein preparations. A column could be re-used for 2-3 weeks without significant loss of DNA.

Preparation and Purification of Gene 5 Protein

Twenty liters of E. coli K37 were grown to a cell density of 2 x 10⁸ cells/ml under conditions of high aeration in broth medium at 37°. The culture was inoculated with wild-type fd at a multiplicity of 50. After 6 hr the culture usually entered plateau phase and the bacteria were harvested by continuous-flow centrifugation. The culture yielded 200 gm of cells which were frozen in 50 gm aliquots. Gene 5 protein was extracted by a modification of a procedure suggested by E.A. Anderson and A.K. Dunker. All steps were carried out at 4° unless otherwise stated. Fifty to 100 gm of cells were thawed and suspended in a minimal
volume of standard buffer without glycerol and sonicated for 30 min maintaining the temperature at less than 15°. After addition of MgCl$_2$ to 20 mM, CaCl$_2$ to 4 mM, and deoxyribonuclease I (Worthington, D) to 20 μg/ml, the sonicate was incubated at room temperature for 2 hr. Sodium chloride was added to 0.8 M and the extract incubated further for 1 hr. Cell debris was removed by centrifugation at 500 x g for 10 min. Ammonium sulfate was added to the supernatant in a concentration of 25 mg/100 ml of solution. Precipitate was allowed to form for 1 hr and collected by centrifugation at 10,000 x g for 1 hr. The supernatant was decanted and the pellet washed gently with standard buffer without glycerol. The pellet was resuspended in a minimal volume of standard buffer without glycerol. The deoxyribonuclease treatment and the ammonium sulfate precipitation steps were repeated and the pellet resuspended in 20 ml of standard buffer without glycerol or NaCl. The solution was clarified by centrifugation for 90 minutes at 30,000 rpm in a Spinco 30 rotor. The extract was made 10% in glycerol and 20 mM in Na$_2$EDTA and applied to a DNA-cellulose column as described above. The column was eluted with standard buffer containing NaCl in steps of 0.05 M, 0.15 M, 0.4 M, 0.8 M, and 2.0 M NaCl. The major peak of gene 5 protein, monitored by optical density at 276 nm, elutes with the 0.8 M NaCl step though prolonged washing with the 0.4 M NaCl buffer will leach off significant amounts of protein. Polyacrylamide gel electrophoresis (described below) can be used after this and subsequent steps to monitor purity. The fractions containing gene 5 protein were pooled and precipitated with ammonium sulfate as described above, redissolved in standard buffer and passed through a 70 x 2 cm Sephadex G-75 (fine) column equilibrated with 10 mM Tris HCl, pH 8, in order to remove high
molecular weight contaminants. Yields of gene 5 protein approached 1 mg of protein per gm of cells, as determined by optical density at 276 nm, 
\[ \text{OD}_{276} = 0.73 \] (Day, 1973). Protein solutions were concentrated by ultra-
filtration using an Amicon UM2 membrane.

Preparation of m-Fluoro-tyrosyl Gene 5 Protein

*E. coli*, strain AT 2741, were grown to early exponential phase in a synthetic medium containing low tyrosine levels (4 \( \mu \)g/ml); M9 media (Miller, 1972) supplemented with 20 \( \mu \)g/ml of the 19 other common amino acids. m-Fluoro-D,L-tyrosine, 30 \( \mu \)g/ml, was added at early log phase followed 10 minutes later by the addition of wild-type fd at a multiplicity of 100. The culture was harvested 6 hr later and the gene 5 protein purified in the usual manner.

Preparation and Purification of Gene 32 Protein

Stocks of the T4 bacteriophage double amber mutant (33\(^{-}\), 55\(^{-}\)) which, in a non-permissive host overproduces gene 32 protein and is lysis defective, were prepared by infecting early log phase cultures growing at 37\(^{\circ}\) of the permissive *E. coli* KL 239 strain at a multiplicity of \( \approx 1 \), letting the infection proceed for 6-8 hr until lysis occurred, and then disrupting the remaining cells with a few drops of chloroform and shaking. These phage grow slowly with a poor yield; when infecting at low multiplicities care must be taken to watch for what are probably revertants with changed plaque morphology. The double mutant produces a small, clear plaque. To produce gene 32 protein, 20 l of non-permissive *E. coli* BW46 were grown up in broth media at 34\(^{\circ}\) to a density of about \( 1 \times 10^9 \) cells/ml and infected with T4 (33\(^{-}\), 55\(^{-}\)) at a multiplicity of 5. The cells were harvested 1 hr after infection using continuous-flow
centrifugation. The yield varied from 120-150 gm of cells; they can be stored frozen at -15° for 4-6 months without significant loss of protein. The gene 32 protein was purified essentially by the methods of Alberts and Frey (1970) with the addition of a final purification using phosphocellulose chromatography as described by Wackernagel and Radding (1974). Fifty grams of cells were thawed, sonicated, treated with DNase, and cleared of cell debris in the same manner as for the gene 5 protein isolation. The homogenate was then dialyzed for 4 hr each versus two 3 l changes of standard buffer lacking glycerol but with 10 mM EDTA and then overnight against 3 l of complete standard buffer made 5 mM in EDTA. The solution was clarified by centrifugation for 90 minutes at 30,000 rpm before loading onto a single-stranded DNA-cellulose column. After washing with new 0.05 M NaCl standard buffer, steps with 0.15 M, 0.5 M, 0.8 M, and 2 M NaCl were run through the column. Material eluting at the 0.8 M NaCl step contained the gene 32 protein as monitored by optical density at 280 nm and acrylamide gel electrophoresis. The fractions showing significant optical density were pooled, dialyzed against standard buffer without NaCl, and loaded onto a 45 x 0.9 cm column of DEAE-cellulose (Whatman DE52) equilibrated in the standard buffer without NaCl. After washing with one column volume of buffer, a 400 ml linear gradient of 0-0.5 M NaCl in standard buffer was started. The gene 32 protein elutes with approximately 0.2 M NaCl. The gene 32 protein-containing fractions, monitored by gel electrophoresis, were pooled and thoroughly dialyzed against a buffer containing 20 mM potassium phosphate (pH 6.5), 1 mM Na₂ EDTA, 1 mM β-mercaptoethanol, and 10% glycerol. The sample was then applied to a 40 x 0.9 cm column of phosphocellulose (Whatman, PL1) equilibrated with
the phosphate buffer. Gene 32 protein showing a single band on polyacrylamide gel electrophoresis was recovered in the wash from this column. After dialysis versus 10 mM Tris-HCl, pH 8, the purified protein can be stored at 4° or at -15°. Protein solutions were concentrated using ultrafiltration with an Amicon UM10 membrane. Concentrations were determined using $E_{280}^{0.1\%} = 1.1$ (Alberts and Frey, 1970) or by amino acid analysis. Yields of gene 32 protein were on the order of 0.6 mg per gm of cells.

**Preparation and Purification of E. coli DNA-Unwinding Protein**

This protein was isolated by the methods of Molineux et al. (1974). *E. coli* BW46 cells were harvested in late log phase. 300 gm of cells were lysed in standard buffer with 2.0 M NaCl but lacking glycerol by sonication in the manner of the gene 5 protein preparation. After debris was removed by centrifugation at 5000 rpm for 10 minutes (Sorvall), the supernatant was made 10% in polyethylene glycol (adding 1/2 volume of a 30% solution in 2 M NaCl). The solution was left to stand at 0° from 45 minutes before sedimenting out the precipitate (Sorvall GSA rotor, 12,000 rpm for 30 minutes). This supernatant was dialyzed versus 3 changes of standard buffer plus 5 mM EDTA but without glycerol and against a final change with 10% glycerol added for 4 hr each. Centrifugation at 30,000 rpm for 90 minutes was used to clarify the solution before loading onto a denatured DNA-cellulose column. Steps of 0.05 M, 0.6 M, 1.0 M, and 2 M NaCl in standard buffer were washed through the column. DNA-unwinding protein appeared in most of the fractions eluted with greater than 0.6 M salt as revealed by acrylamide gel electrophoresis but the majority of it came off late in the 2.0 M NaCl wash. The high salt fractions containing the unwinding protein were combined.
and dialyzed versus standard buffer without salt. The sample was then applied to a 20 x 2 cm DEAE-Sephadex A50 column and eluted with a 400 ml 0-0.7 M NaCl gradient. The unwinding protein elutes at approximately 0.3 M NaCl; this sample shows one band upon polyacrylamide gel electrophoresis (Figure 1). This protein was stored in the standard buffer. Concentrations were determined by amino acid analysis. It was concentrated by ultrafiltration using an Amicon UM10 membrane. Yield was 1 mg from 100 gm of cells.

Polyacrylamide Gel Electrophoresis

10% Polyacrylamide, 0.1% sodium dodecyl sulfate (SDS) slab gels produced using the procedures of Weber and Osborn (1969) and run using an apparatus similar to that described by Studier (1973) were used to monitor protein purity. The only modification upon the Weber and Osborn recipe involved the incorporation of 5 M urea into the gels. Standards run in each gel included bovine serum albumin (68,000 MW), pepsin (35,000 MW), trypsin (23,300 MW), and lysozyme (14,300 MW).

Purification of fd DNA

Cells were removed by centrifugation from a standard culture of E. coli infected with fd as described above. Polyethylene glycol and NaCl, 2% w/v, 0.5 M, was added to the supernatant. The precipitated phage were allowed to settle for 2-3 days at 4° (Wiseman et al., 1972). The precipitate was collected at 8,000 x g and the phage extracted 3x with cold H₂O. Cesium chloride was added to a density of 1.3 gm/ml and the fd were banded by ultracentrifugation in a Spinco 30 rotor at 27,000 x g for 24 hr at 4°. After dialysis to remove CsCl, the fd DNA was extracted with redistilled phenol by the methods of Marvin and...
Figure 1. Polyacrylamide gel electrophoresis of purified *E. coli* DNA-unwinding protein: standard proteins (as in Materials and Methods) are shown at the left; the *E. coli* protein sample shows a single band that migrates very slightly faster than trypsin (MW 23,300).
Schaller (1966). The concentration of fd DNA was determined using the extinction coefficient $\varepsilon(P) = 7370 \text{ cm}^{-1} \text{ M}^{-1}$ (259 nm, in 0.95 M NaCl, 0.015 M sodium citrate (pH 8)) as determined by Berkowitz and Day (1974). Electron microscopy revealed that greater than 95% of the DNA was in the form of intact covalent circles (performed by Dr. G. Bourguignon).

**Preparation of Oligonucleotides**

Short, 5'-phosphate terminated oligonucleotides of defined length but mixed base composition were separated from a pancreatic deoxyribonuclease I limit digest of salmon sperm DNA using procedures developed by Tomlinson and Tener (1963). One gram of DNA was hydrated overnight in 100 ml of 15 mM magnesium acetate (pH 7) at 4°C. One mg deoxyribonuclease was added and over the next 3 hr pH was monitored and kept at pH 7 with the addition of 0.1 N NaOH. The increase in optical density at 271 nm was also followed. When no further changes in these parameters were noted, the mixture was heated to 100°C for a few minutes and the resulting white precipitate was removed by filtration through an 0.45 μm millipore filter in a Nalge filter unit. The filtrate was diluted to 200 ml with distilled water and applied to an 82 x 4 cm DEAE-cellulose (Whatman DE52) column which had been washed thoroughly with 4 M NaCl followed by distilled water until the eluant was salt-free. After washing the digest into the column with a small amount of distilled water, the column was eluted with a 15 l, linear 0-0.3 M NaCl gradient in 0.003 M Tris-HCl (pH 7.8) and 7 M urea. Fractions were collected and monitored by optical density at 271 nm where all deoxyribonucleotides have extinction coefficients of 9100 ± 10% cm$^{-1}$ M$^{-1}$. Correlation of optical density peaks with oligonucleotide length was made on the basis of the salt concentration at which they eluted - the results agreed quite well.
with the original findings of Tomlinson and Tener (1963). Material from
the peaks rechromatograph to the same position and appear in the eluant
where the assignments predict in relation to known standards when they
are co-chromatographed. This procedure gave a yield of approximately
3000 OD$_{271}$ units each of tri-, tetra-, and pentanucleotides with lesser
amounts of other oligonucleotides through septanucleotides. To remove
the salt and urea about 300 OD$_{271}$ units of material was diluted 4-fold
with distilled water and applied to a 40 x 2 cm DEAE-cellulose column
previously equilibrated with ammonium bicarbonate and washed with water.
The oligonucleotide sample was washed into the column with a large amount
of water. The column was then washed with 0.02 M ammonium bicarbonate
(pH 8.4) until no chloride ion was detectable in the effluent using the
silver nitrate test (acidify the sample with 1-2 drops concentrated HNO$_3$,
add 2-3 drops of 1% AgNO$_3$, a white precipitate indicates the presence
of Cl$^-$). The oligonucleotide was then eluted with 2 M ammonium bicar­
bonate. Repeated lyophilization and resuspension in water was used to
remove the ammonium bicarbonate. Oligonucleotide concentration was
determined using $\varepsilon$(P) = 9100 at 271 nm.

Chemical Modifications of Proteins

The various modifications were carried out as described in the
literature and details will be given in the Results section. Reactions
were terminated by dialysis of the reaction mixture or by gel filtration
of the reaction mixture over a column of Sephadex G-25 (medium) in 10
mM Tris-HCl pH 8. The modified proteins were concentrated by ultra-
filtration using an Amicon membrane appropriate for the molecular weight
of the proteins.
Amino Acid and Peptide Analyses

Amino acid analyses were performed by the method of Spackman et al. (1958) on a Beckman 120B amino acid analyzer. Samples were hydrolyzed in 6 N HCl for 22 and 72 hr at 110°C. Half-cystine was determined as cysteic acid by the method of Moore (1963). Tryptophan content was determined using magnetic circular dichroism as per McFarland and Coleman (1972). Peptide analysis of nitrated gene 5 protein: Nitrated gene 5 protein, 10 mg/ml, in 0.05 M NH₄HCO₃, pH 8.5, was digested with chymotrypsin at 0.1 mg/ml for 1 hr at 25°C after which another aliquot of chymotrypsin (final ratio of gene 5 protein: chymotrypsin of 50:1, w/w) was added and incubation continued for another hr. The digestion was stopped by boiling for 2 min. The peptides were initially separated using a 100 x 0.9 cm Sephadex G-50 (fine) column eluted with 0.2 M ammonium hydroxide. The yellow material obtained in the first peak from this elution was digested with trypsin using the same procedure as with chymotrypsin, applied to a 50 x 0.9 cm Sephadex G-50 (fine) column and eluted with a 1:1 mixture of 0.4 M ammonium hydroxide:n-propanol. The material in the major peak from this chromatography was applied to Whatman 20 chromatography paper and subjected to high voltage electrophoresis in 0.1 M sodium bicarbonate, pH 9.2. Other peptides from the original chymotryptic digest were separated and purified by means of electrophoresis in 10% pyridine-acetate buffer at pH 6.5. Yellow spots were detected by exposing the dried paper to ammonia fumes and other spots were visualized with fluorescamine (Bohlen et al., 1973). All spots were eluted with 0.2 M NH₄OH. Amino terminal residues were determined by dansylation and thin layer chromatography on polyamide plates as described by Gray (1972). Carboxyl terminal residues were determined

Absorption Spectra were measured on a Cary 15 spectrophotometer. Circular Dichroism was measured on a Cary 61 spectropolarimeter. CD is expressed in terms of molecular ellipticity, \([\theta] = 2.303 \left(\frac{4500}{\pi}\right)(\epsilon_L - \epsilon_R)\) in units of deg cm\(^2\)/dmol. Magnetic Circular Dichroism (MCD) was performed on a Cary 61 spectropolarimeter equipped with a superconducting magnet producing a field of 44.7 kilo gauss in the sample chamber (McFarland and Coleman, 1972).

\(^{19}\text{F}\) Nuclear Magnetic Resonance (NMR) Spectra were obtained on a Varian XL-100-15 NMR spectrometer operating at 94.1 MHz for \(^{19}\text{F}\), and locked on the \(^2\text{H}\) resonance of the solvent (\(^2\text{H}_2\text{O}\)). The \(^{19}\text{F}\) chemical shifts were determined relative to an external capillary of CF\(_3\)COOH. Data obtained by Fourier transform were processed as described by Sykes et al. (1974).

Mercury Analyses were performed by atomic absorption spectroscopy (Duckworth and Coleman, 1970).
CHAPTER III

RESULTS

CD of Gene 5 Protein-Nucleic Acid Complexes

Circular dichroism is a convenient method of monitoring changes in DNA conformation on the binding of gene 5 protein, since the major ellipticity changes in the DNA occur in the region of 260 nm and gene 5 protein shows very little ellipticity above 240 nm (Day, 1973, see below). The addition of aliquots of gene 5 protein to a constant amount of poly[d(A-T)] (ε_{260} = 6.7 \times 10^3, Green and Mahler, 1971) shifts the positive ellipticity at 260 nm to negative values (Figure 2). A plot of the ellipticity change at 260 nm as a function of the amount of gene 5 protein present shows that no further spectral change occurs after approximately 0.25 moles of protein per mole of nucleotide has been added (Figure 2, insert). This ratio of one protein monomer to four nucleotides at the maximum ellipticity change is also observed with fd DNA (Figure 2, insert). The maximum ellipticity change occurs at 270 nm for fd DNA. The CD spectra for fd DNA (see Figure 4) are identical to those published by Day (1973). The CD assay can detect a change of 10% in the degree of binding of gene 5 protein with no ambiguity. The concentration of protein used, 5 \times 10^{-6} \text{ M}, and the similar concentrations of DNA present (calculated in terms of the molarity of gene 5 protein binding sites), coupled with the lack of further change in the CD spectrum of the complex above a protein to nucleotide ratio of 1:4 (Figure 2) suggest that the dissociation constant,
Figure 2. Circular dichroism of poly [d(A-T)], 3.5 x 10^{-5} M, in 2.85 ml of 10 mM Tris HCl-1 mM Na$_2$EDTA, pH 8, 25°, during titration with 25 μl aliquots of gene 5 protein (5p), 2.8 x 10^{-4} M. --- poly [d(A-T)] alone; ------ (1-8) poly d(AT) + 5p; protein:DNA ratio = 1:14(1), 1:7(2), 1:5(3), 1:3.5(4), 1:2.4(6), 1:2(7), 1:1.4(8). Insert: Change in DNA ellipticity as a function of the molar ratio of 5p:DNA bases. (●) poly [d(A-T)] which shows the maximum change in ellipticity at 260 nm; (○) fd DNA which shows the maximum ellipticity change at 270 nm.
Kₐ of the native gene 5 protein-fd DNA complex must be less than 5 × 10⁻⁸ M.

Significant changes in the CD of d(pA)₄ in the 260 nm region are induced by the addition of gene 5 protein (Figure 3). Similar changes are induced in the CD spectrum of the random sequence tetramer isolated by the Tomlinson and Tener (1963) techniques. A protein to nucleotide ratio of greater than 1:1 is required to produce maximal ellipticity changes, commensurate with a much reduced affinity for the tetranucleotide. An estimate of Kₐ for the gene 5 protein-tetranucleotide complex pictured in Figure 3 is 3 × 10⁻⁶ M, at least 2 orders of magnitude greater than the constant estimated for the complex with intact fd DNA. The difference may be a measure of the increase in stability of the complex conferred by protein-protein interactions between adjacent molecules of gene 5 protein on the DNA. Cooperativity of binding of gene 5 protein to fd DNA has been suggested by other observations (Alberts et al., 1972). Gene 5 protein shows no demonstrable affinity for trinucleotides. However, the protein has a significant affinity for single-stranded RNA as shown by the induction of large changes in the CD spectrum of poly(A) (Figure 3). As in the case of DNA the maximal CD change occurs at a protein to ribonucleotide molar ratio of 1:4.

Effects of Cations on Gene 5 Protein-DNA Complexes

The addition of various cations to the gene 5 protein-fd DNA complex reverses the CD of the complex to that observed for the DNA alone under the same salt concentrations, suggesting that the complex has dissociated. The complete CD spectra during the dissociation of the gene 5 protein-fd DNA complex by Na⁺ are shown in Figure 4. The
Figure 3. Circular dichroism changes on the binding of gene 5 protein (5p) to d(pA)₄ and poly(A). (——--) Poly(A), 2.4 x 10⁻⁵ M, in 2.5 ml; (-----) same sample of poly(A) plus 50 μl of 5p, 30 x 10⁻⁵ M, to give a protein:base ratio of 1:4; (······) d(pA)₄, 2.1 x 10⁻⁵ M, in 2.5 ml; (···) d(pA)₄ plus 5p at a molar ratio of 3.5 protein molecules per tetranucleotide. Conditions: 10 mM Tris HCl-10 mM Na₂EDTA, pH 7, 25°.
protein shows a prominent CD maximum at 228 nm apparently due to the
dissymmetry of the tyrosyl chromophores. On complex formation with
fd DNA the magnitude of this peak falls significantly (Day, 1973). On
dissociation of the gene 5 protein-fd DNA complex with Na⁺ the ellip-
ticity peak of the protein at 228 nm returns to its normal magnitude
(Figure 4, left inset).

Similar effects of other cations on the gene 5 protein complexes
with both poly [d(A*T)] and fd DNA are shown in the right inset to
Figure 4 by plotting the ellipticity change of the complex against the
salt concentration. The chloride salts of K⁺, Li⁺, and Cs⁺ dissociate
the fd DNA-gene 5 protein complex at the same concentration as Na⁺.
Divalent cations, Mg²⁺ (Figure 4) and Ca²⁺ (superimposable on Mg²⁺
results), are much more effective than monovalent cations in reversing
the binding of gene 5 protein to poly [d(A-T)]. This is consistent with
the known effectiveness of divalent metal ions in stabilizing the
double-helical conformation.

Effect of Chemical Modification of Gene 5 Protein on the Formation of
DNA Complexes

The CD data summarized above suggest that the alteration in the
CD of poly [d(A-T)] or fd DNA induced by gene 5 protein appears to be a
sensitive means of detecting complex formation between the protein and
DNA. We have used this CD assay to monitor the effect of several
chemical modifications of gene 5 protein on its ability to form complexes
with DNA.

Reaction of Gene 5 Protein with N-Acetylimidazole

N-Acetylimidazole has been shown to selectively O-acetylate
Figure 4. Circular dichroism changes during the titration of the gene 5 protein (5p)-fd DNA complex with NaCl. (1) no NaCl, (2) 0.05 M, (3) 0.10 M, (4) 0.19 M, (5) 0.28 M, (6) 0.38 M, (7) 0.42 M, (8) 0.46 M, (9) 0.51 M, and (10-12 superimposed) 0.55-0.80 M NaCl. (---------) fd DNA alone at the same concentration as in spectrum (12), 3.9 x 10^{-5} M. Conditions: molar ratio of 5p:fd DNA = 1:3, 10 mM Tris HCl-1 mM Na2EDTA, pH 8. Ellipticity is plotted in degrees due to the concentration changes accompanying the salt additions. The insert at the left shows the restoration of the magnitude of the ellipticity maximum of the protein at 228 nm as the complex dissociates (the numbers correspond to the spectra in the main figure). The insert at the right summarizes the salt titration data by plotting the % of the maximum ellipticity change (at 260 nm for the 5p-poly [d(A-T)] complex and at 270 nm for the 5p-fd DNA complex) as a function of the molarity of the added cation; ( ○ ), Mg^{++}-d(AT); ( ● ) Mg^{++}-fd DNA; ( ■ ) Na^{+}-d(AT); ( □ ) K^{+}-d(AT); ( ▲ ) Na^{+}-fd DNA.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
tyrosyl phenolic hydroxyl groups and to N-acetylate the ε-amino groups of lysyl residues in intact proteins (Riordan et al., 1965). Gene 5 protein, $3 \times 10^{-4} \text{M}$, in 0.02 M barbital buffer, pH 7.5, was reacted with a 100-fold molar excess of N-acetyl imidazole for 30 minutes, 25°. Reaction for longer times resulted in precipitation of the protein. Monitoring of the absorption spectrum (Figure 5) of the modified protein at 276 nm before ($\varepsilon_{276} = 3500 \text{ M}^{-1} \text{cm}^{-1}$) and after ($\varepsilon_{276} = 7300 \text{ M}^{-1} \text{cm}^{-1}$) treatment with 1 M $\text{NH}_2\text{OH}$ showed that 3.3 tyrosyl residues of the protein had been O-acetylated with N-acetylimidazole and removed with hydroxylamine (Simpson, et al., 1963). Acetylation of the protein with $[^{14}\text{C}]$-N-acetylimidazole resulted in the incorporation of 3200 cpm/100 nmole of protein. Treatment of the $[^{14}\text{C}]$-labelled protein with 1 M $\text{NH}_2\text{OH}$ removed 1140 cpm/100 nmole from the protein. The optical density increase at 276 nm accompanying the removal of the $[^{14}\text{C}]$-acetyl groups showed that 3.4 tyrosyl residues were deacylated. This suggests that the 2060 cpm/100 nmole remaining with the protein following deacylation represent 5.96 N-acetyl groups per protein monomer, probably on the ε-amino groups of the 6 lysyl residues of the protein. Neither the acetylated protein nor the acetylated protein treated with hydroxylamine alter the CD spectrum of poly [d(A-T)] or fd DNA when incubated with the polymers in large excess. Thus the dissociation constant must be increased by at least 4 orders of magnitude by acetylation of the lysyl residues of the protein.

Formation of the complex between gene 5 protein and fd DNA does not protect the protein from acetylation. Acetylation of the complex under the conditions described above, followed by passage of the product through a G-25 Sephadex column to remove the reagent results in a CD
Figure 5. Absorption spectra of acetylated and hydroxylamine treated gene 5 protein, 0.01 M Tris-HCl, pH 8: (-----) fully acetylated gene 5 protein; (------) same sample (corrected for concentration change) after addition of hydroxylamine to 1.0 M.
spectrum for the products identical to that for uncomplexed fd DNA. The unmodified complex passed through the same column retains the CD spectrum of the gene 5 protein-fd DNA complex (Figure 4, curve 1). Reaction of fd DNA with N-acetylimidazole does not affect its ability to complex with native gene 5 protein. While both lysyl and tyrosyl residues are involved in the reaction with N-acetylimidazole, it may be primarily the lysyl residues that are accessible to the reagent in the complex (see below).

Reaction of Gene 5 Protein with Tetranitromethane (C(NO₂)₄)

C(NO₂)₄ specifically nitrates tyrosyl residues of proteins at the 3-position of the ring (Sokolovsky et al., 1966). Gene-5 protein, 8 x 10⁻⁴ M, in 0.05 M Tris-HCl, 0.15 M NaCl, pH 8, was reacted for 20 min at room temperature with a 64-fold molar excess of C(NO₂)₄. Reaction for longer times or with a greater molar excess of C(NO₂)₄ leads to precipitation of the protein. Acrylamide gel electrophoresis (Figure 6) indicates that more than 50% of the nitrated gene 5 protein is present as dimers or higher oligomers, necessitating the isolation of a monomeric fraction using Sephadex G-100 gel chromatography in 0.2 M KCl, 10 mM Tris-HCl, pH 8.

Characterization of Nitrated Gene 5 Protein

Amino acid analysis of nitrated gene 5 protein monomers shows the presence of 3 mononitrotyrosyl residues out of the total of 5 tyrosyl residues in the protein. Two of the tyrosyl residues are unmodified. The mole ratio of the other amino acids in the protein was not altered (Table I). No material appears in the amino acid in the position of cysteic acid, the expected product of the reaction of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 6. 10% polyacrylamide gel showing (from left to right) purified native gene 5 protein; nitrated gene 5 protein; separated monomers of nitrated gene 5 protein; gene 5 protein protected from nitration by prior binding to fd DNA. Position of marker proteins are indicated at the right by their molecular weights.
<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Expected from Sequence Data</th>
<th>Total Nitrated Protein</th>
<th>Nitrated Monomers</th>
<th>DNA-Protected Protein after C(NO₂)₄ Reaction</th>
<th>Tetranucleotide Protected Protein after C(NO₂)₄ Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr</td>
<td>5</td>
<td>1.8</td>
<td>1.7</td>
<td>4.6</td>
<td>4.3</td>
</tr>
<tr>
<td>NO₂-tyr</td>
<td>-</td>
<td>2.4</td>
<td>2.9</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Asp</td>
<td>5</td>
<td>5.3</td>
<td>5.2</td>
<td>5.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Thr</td>
<td>4</td>
<td>3.8</td>
<td>3.9</td>
<td>3.7</td>
<td>3.9</td>
</tr>
<tr>
<td>Ser</td>
<td>7</td>
<td>6.4</td>
<td>6.6</td>
<td>6.4</td>
<td>6.6</td>
</tr>
<tr>
<td>Glu</td>
<td>10</td>
<td>9.5</td>
<td>9.8</td>
<td>9.9</td>
<td>10.0</td>
</tr>
<tr>
<td>Pro</td>
<td>6</td>
<td>6.5</td>
<td>6.9</td>
<td>6.0</td>
<td>6.5</td>
</tr>
<tr>
<td>Gly</td>
<td>7</td>
<td>7.3</td>
<td>7.3</td>
<td>6.7</td>
<td>18.5b</td>
</tr>
<tr>
<td>Ala</td>
<td>4</td>
<td>4.0</td>
<td>4.1</td>
<td>4.0</td>
<td>4.6</td>
</tr>
<tr>
<td>Val</td>
<td>8</td>
<td>8.1</td>
<td>8.1</td>
<td>7.2</td>
<td>7.2</td>
</tr>
<tr>
<td>Met</td>
<td>2</td>
<td>2.1</td>
<td>1.9</td>
<td>1.6</td>
<td>1.8</td>
</tr>
<tr>
<td>Ile</td>
<td>4</td>
<td>4.0</td>
<td>3.8</td>
<td>4.0</td>
<td>3.5</td>
</tr>
<tr>
<td>Leu</td>
<td>10</td>
<td>9.9</td>
<td>10.0</td>
<td>10.2</td>
<td>9.4</td>
</tr>
<tr>
<td>Phe</td>
<td>3</td>
<td>3.2</td>
<td>3.0</td>
<td>2.9</td>
<td>2.9</td>
</tr>
<tr>
<td>His</td>
<td>1</td>
<td>0.9</td>
<td>1.0</td>
<td>1.3</td>
<td>1.2</td>
</tr>
<tr>
<td>Lys</td>
<td>6</td>
<td>5.7</td>
<td>5.7</td>
<td>6.4</td>
<td>5.4</td>
</tr>
<tr>
<td>Arg</td>
<td>4</td>
<td>3.8</td>
<td>3.9</td>
<td>3.1</td>
<td>3.7</td>
</tr>
<tr>
<td>TOTAL</td>
<td>87</td>
<td>86.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Half-cystine not regularly determined.

\(^b\)Hydrolyzed sample contained nucleotides which decompose to glycine.
cysteine with C(NO$_2$)$_4$ (Sokolovsky et al., 1966), unless the modified protein is oxidized with performic acid before hydrolysis. Using a series of amino acid analyses correlated with spectral data (Figure 7) to determine the concentration of nitrated protein, extinction coefficients for the nitrated protein at pH 8 have been calculated: $\varepsilon_{428} = 10.2 \times 10^3 \text{M}^{-1} \text{cm}^{-1}$, $\varepsilon_{381} = 6.2 \times 10^3 \text{M}^{-1} \text{cm}^{-1}$, and $\varepsilon_{276} = 16.9 \times 10^3 \text{M}^{-1} \text{cm}^{-1}$. The extinction at 381 nm, the isosbestic point for the hydrogen ion titration of the hydroxyl group of 3-nitrotyrosine (Sokolovsky et al., 1966) is consistent with the presence of 3 nitrated tyrosyl residues per protein monomer, in agreement with the amino acid analysis. Amino acid analysis indicates that all 5 tyrosines are nitrated if the protein is denatured and reacted in 6 M guanidine-HCl.

Addition of nitrated gene 5 protein monomers to fd DNA produces little change in the 270 nm ellipticity band of fd DNA until protein to nucleotide molar ratios significantly greater than the 1:4 seen for the native protein are obtained (Figure 8). A 75% change was seen at a ratio of 1:1. This suggests that the nitrated monomers retain a slight affinity for DNA. The dissociation constant estimated from the change in the 270 nm ellipticity band as a function of the nitrated monomer concentration is not less than $6 \times 10^{-6}$ M, a 100-fold increase over the maximum dissociation constant estimated for the native protein.

Spectrophotometric hydrogen ion titration curves of native gene 5 protein show that the molar extinction coefficient at 293 nm increases over a broad pH range from 8 to 13 (Figure 9). The final absorption indicates the presence of 5 ionized tyrosyl residues. Since there are no tryptophan residues in the molecule the interpretation of the spectral data is quite straightforward. Taking into account a slight
Figure 7. Absorption spectra of native, nitrated, and reduced gene 5 proteins, all spectra taken in 0.01 M Tris-HCl, pH 8: (---) native gene 5 protein, $2.3 \times 10^{-5}$ M; (-----) nitrotyrosine gene 5 protein monomers, $3.3 \times 10^{-5}$ M; (-----) aminotyrosine gene 5 protein monomers, $3.7 \times 10^{-5}$ M.
Figure 8. Change in fd DNA ellipticity as a function of the molar ratios of nitro- and aminotyrosine gene 5 protein to DNA bases (R = base/protein): (■) nitrotyrosine gene 5 protein; (□) aminotyrosine gene 5 protein; (●) native gene 5 protein for comparison (from Figure 2, insert).
Figure 9. Spectrophotometric hydrogen ion titration of native gene 5 protein (5p) and nitrated gene 5 protein. Aliquots of 1 M NaOH (1 M HCl for the reverse titrations) sufficient to change the pH by 0.5 units were added to protein samples using a glass syringe driven by a micrometer. Conditions: 0.01 mM Tris-HCl, 0.15 M NaCl, 0.001 M Na₂EDTA. Above pH 11.5 and below pH 5.5 the protein solutions began to show irreversible turbidity. The data shown are corrected for concentration changes and light scattering artifacts (Beaven and Holiday, 1952). (○), native 5p, 4.5 x 10⁻⁵ M, titration followed at 293 nm; (■) back titration of the same sample. Similar data were obtained at 283 nm. (△) nitrated 5p, 2.2 x 10⁻⁵ M; (●) back titration of the same sample. The titration was followed at 428 nm. (-----) Smooth curves manually drawn to fit the titration points. (---------) Theoretical curves drawn by assuming the 3 tyrosyl hydroxyl groups of native 5p have pKₐ = 9.95 and 2 tyrosyl hydroxyl groups have pKₐ = 11.7. ΔE₂93 assumed to accompany tyrosyl ionization was 2.5 x 10³ M⁻¹ cm⁻¹.
change in slope of the line through the points above pH 11, two sigmoid titration curves can be fitted to the data corresponding to 3 tyrosyl residues with apparent pKₐ's of 10.0 and 2 tyrosyl residues with apparent pKₐ's of 11.7 (Figure 9).

Spectrophotometric hydrogen ion titration of the nitrated gene 5 protein at 428 nm indicates that 3 nitrotyrosyl residues ionize with apparent pKₐ's of 7 (Figure 9), close to the pKₐ value reported for the hydroxyl group of free nitrotyrosine (Sokolovsky et al., 1966).

Protection of Gene 5 Protein Against Nitration

If gene 5 protein is complexed with excess fd DNA before reaction with C(NO₂)₄, no tyrosyl residues are modified as judged by the lack of change in the absorption spectrum and amino-acid analysis indicating less than 0.1 modified residue per protein (Table I). The CD of the product after removal of the reagent by gel filtration is identical to that of the gene 5 protein-fd DNA complex; the CD undergoes the usual changes when the complex is dissociated with Mg²⁺ as illustrated in Figure 4. Acrylamide gel electrophoresis of the "protected" gene 5 protein shows no evidence of crosslinking (Figure 6). Nor is there any apparent crosslinking of protein to DNA as the DNA separated from the reacted complex has normal absorption and CD spectra and complexes with a normal stoichiometry of native gene 5 protein. There is no reaction when C(NO₂)₄ is added to a solution of fd DNA, as monitored by absorption changes at 350 nm. This is in contrast to results found with calf thymus DNA by Hugli and Stein (1971). The nitration of proteins which do not specifically bind to DNA, such as bovine serum albumin, is not altered by the presence of DNA.
Gene 5 protein can also be protected from nitration by complexing with a tetranucleotide; however, a large excess of nucleotide must be used due to the higher dissociation constant of the protein-tetranucleotide complex. Using a gene 5 protein concentration of $1 \times 10^{-4}$ M under the standard nitration conditions (100-fold excess of $C(NO_2)_4$) a 100-fold excess of nucleotide (phosphate) will prevent nitration of tyrosyl residues as monitored by amino acid analysis (Table II) and prevent protein-protein crosslinking. A 16-fold nucleotide (phosphate) excess does not hinder tyrosyl nitration nor crosslinking.

Gene 5 protein which has previously been acetylated with N-acetyl-imidazole is protected from reaction with $C(NO_2)_4$. In an experiment in which 1 M hydroxylamine could remove 3.6 acetyl's from the acetylated protein before treatment with $C(NO_2)_4$, 3.2 acetyl's could be removed after the nitration reaction. Spectrally, 0.5 nitro-tyrosyl residue per protein was formed; this was confirmed by amino acid analysis. Acrylamide gel electrophoresis indicated that nearly all of the doubly reacted protein migrated in the position of monomers with only a faint band at the dimer position.

**Location of the Nitrated Tyrosyl Residues in the Primary Structure of Gene 5 Protein**

Two peptides (C2 and C4) isolated from a chymotryptic digest of nitrated gene 5 protein contained all three nitrated tyrosyl residues (Table II). Peptide C2, in the nomenclature of Nakashima et al. (1974a, b), extending from residue 14 through 26 (Figure 8) contained only one tyrosine, number 26, which is nitrated. Peptide C4, residues 35 to 61, has two of its 3 tyrosines modified. This peptide was further digested with trypsin yielding 2 major fragments, C4T1 and C4T2, with
### TABLE II
AMINO ACID ANALYSIS DATA FOR NITRATED GENE 5 PROTEIN PROTEOLYTIC FRAGMENTS

<table>
<thead>
<tr>
<th>Amino Acid&lt;sup&gt;a,b&lt;/sup&gt;</th>
<th>C2</th>
<th>C4</th>
<th>C4T1</th>
<th>C4T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr</td>
<td>- (1)</td>
<td>0.9 (3)</td>
<td>- (1)</td>
<td>0.5 (1)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>NO&lt;sub&gt;2&lt;/sub&gt;-tyr</td>
<td>0.7</td>
<td>1.9</td>
<td>1.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Asp</td>
<td>-</td>
<td>2.8 (3)</td>
<td>1.7 (2)</td>
<td>1.0 (1)</td>
</tr>
<tr>
<td>Thr</td>
<td>1.5 (2)</td>
<td>1.1 (1)</td>
<td>-</td>
<td>0.9 (1)</td>
</tr>
<tr>
<td>Ser</td>
<td>1.7 (2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glu</td>
<td>1.1 (1)</td>
<td>3.6 (3)</td>
<td>1.3 (1)</td>
<td>2.3 (2)</td>
</tr>
<tr>
<td>Pro</td>
<td>0.8 (1)</td>
<td>2.8 (3)</td>
<td>1.3 (1)</td>
<td>1.8 (2)</td>
</tr>
<tr>
<td>Gly</td>
<td>2.1 (2)</td>
<td>2.6 (3)</td>
<td>0.5 (1)</td>
<td>2.0 (2)</td>
</tr>
<tr>
<td>Ala</td>
<td>-</td>
<td>1.5 (2)</td>
<td>0</td>
<td>1.9 (2)</td>
</tr>
<tr>
<td>Val</td>
<td>1.1 (1)</td>
<td>2.6 (3)</td>
<td>2.4 (3)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td>Met</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ile</td>
<td>-</td>
<td>0.8 (1)</td>
<td>-</td>
<td>0.9 (1)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Leu</td>
<td>-</td>
<td>3.8 (4)</td>
<td>1.8 (2)</td>
<td>2.0 (2)</td>
</tr>
<tr>
<td>Phe</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>His</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lys</td>
<td>1.1 (1)</td>
<td>0.9 (1)</td>
<td>0.8 (1)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td>Arg</td>
<td>1.7 (2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Recovery<sup>e</sup> | 40% (S,E) | 70% (S) | 27% (S,E) | 30% (S,E) |

<sup>a</sup>Half-cystine not regularly determined.
<sup>b</sup>Numbers in parentheses refer to expected number of residues based on Nakashima et al. (1974a,b).
<sup>c</sup>N-terminal residue.
<sup>d</sup>C-terminal residue.
<sup>e</sup>Percent recovery after purification procedures: gel filtration on Sephadex (S); electrophoresis on paper (E).
compositions and amino and carboxyl termini corresponding to residues 35 through 46 and 47 through 61 respectively. The former contained one nitrated tyrosine, number 41. A carboxypeptidase A digestion of C4T2, which contains one unmodified and one nitro-tyrosyl, yielded equimolar amounts of unmodified tyrosine (tyr-61) and leucine (leu-60). Thus residue 56 is the third modified tyrosyl. It is concluded that tyrosyl residues at positions 26, 41, and 56 are specifically nitrated by reaction of the protein with C(NO₂)₄.

Characterization of Crosslinked Gene 5 Nitrated Protein Oligomers

Using concentrations determined by amino acid analysis, the extinction coefficients of the oligomeric fraction of nitrated gene 5 protein are found to be lower than those of the monomers: ε₄₂₈ = 8.2 x 10³ M⁻¹ cm⁻¹, ε₃₈₁ = 5.6 x 10³ M⁻¹ cm⁻¹, and ε₂₇₈ = 14.6 x 10³ M⁻¹ cm⁻¹. Again the amino acid analysis shows approximately 3 moles of nitro-tyrosine per mole of protein and the normal molar ratios of other residues (Table I). The oligomeric proteins produce no change in the CD spectra when added to fd DNA even in great excess. The CD spectrum of the oligomeric proteins show a significantly decreased ellipticity at 228 nm (not shown).

The lack of measurable cross-linking of gene 5 protein complexed to DNA by reaction with C(NO₂)₄ is consistent with the conclusions of other investigators (Vincent et al., 1970; Williams and Lowe, 1971; Bruice et al., 1968) suggesting that the nitrated tyrosyls themselves are involved in the bridging reaction (possibly by potentiating a free radical mechanism involving an adjacent aromatic carbon). As noted above the oligomeric proteins have approximately a full complement of nitrated tyrosyl residues. This conclusion also explains the changes
in the tyrosine regions of the CD and absorption spectra of the cross-linked proteins. Studies which use C(NO\textsubscript{2})\textsubscript{4} for its crosslinking properties, for example with histones (Martinson and McCarthy, 1975), would seem to be quite prone to artifacts as complex formation would probably protect the tyrosyls of interest and prevent specific crosslinking at the regions of actual molecular contact.

Reduction of Nitrated Gene 5 Protein

The nitro groups of nitrated gene 5 protein (3 x 10\textsuperscript{-5} M) were reduced to amino groups by the addition of a 25-fold molar excess of sodium hydrosulfite in 0.05 M Tris-HCl, pH 8 using the technique of Sokolovsky et al. (1967). As shown in Figure 7 the absorbance at 428 nm disappeared in less than 5 min, and the absorption maximum at 276 nm shifted to 284 nm characteristic of amino-tyrosine. Sodium hydrosulfite was removed from the protein by dialysis and the protein tested for complex formation with fd DNA; the affinity of this protein for fd DNA was not restored to that of native gene 5 protein but appeared to remain the same as that of the nitrated gene 5 protein (Figure 8).

NMR Spectrum of Fluorotyrosyl Gene 5 Protein

A fluorotyrosyl derivative of gene 5 protein was prepared by infecting a tyrosine auxotroph of \textit{E. coli} growing on minimal media supplemented with m-fluorotyrosine with fd bacteriophage. The yields of active gene 5 protein were about 15% of those normally obtained. However, no difference between the DNA-binding properties of the native and fluorotyrosyl gene 5 protein could be detected by the CD assay for complex formation as depicted in Figures 2 and 3 for the native protein. Both the CD and absorption spectra of the fluorotyrosyl protein at pH 8
were identical to the normal protein except for a slight shoulder at the long wavelength side of both spectra reflecting a slightly larger proportion of ionized fluorotyrosyl residues at this pH. This would be expected from the moderate decrease in pK$_a$ of the hydroxyl group induced by the m-fluoro substitution (Sykes et al., 1974). A $^{19}$F NMR spectrum of the fluorotyrosyl-substituted protein is shown in Figure 10. Five resonances corresponding to the 5 tyrosyl residues of the protein are resolved. Three of the resonances (3, 4 and 5 in Figure 10) are grouped around the resonance position of $^{19}$F in free m-fluorotyrosine and close to the resonance position shown by the fluorines in denatured fluorotyrosyl-substituted alkaline phosphatase (Hull and Sykes, 1974). The other two $^{19}$F resonances (1 and 2 in Figure 10) are shifted down-field as would be expected of fluorines on tyrosyl residues buried in the interior of the protein (Sykes et al., 1974). This result is consistent with the accessibility of three of the tyrosyl residues to modification by N-acetylimidazole and C(NO$_2$)$_4$.

In an initial attempt to look for changes in the environment of the tyrosyl residues when the gene 5 protein forms a complex with nucleic acids, intact fd DNA was added to the protein NMR sample. The resulting spectra (Figure 11) shows only markedly broadened resonances as would be expected if the mobility of the protein was restricted by complexing with a very large supra-molecular structure. Even more broadened resonances were produced by a poly [d(A-T)]-protein complex. The broadening problem was overcome by using random sequence tetradexoxyribonucleotides to form complexes. Figure 12B and 12C demonstrate the distinct up-field shift of approximately 0.3 PPM each for two of the exposed tyrosyls when the fluorotyrosine labelled protein is bound to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 10. $^{19}$NMR spectrum of $7.3 \times 10^{-4}$ M of gene 5 protein containing $m$-fluorotyrosyl residues. Conditions: 10 mM Tris-HCl, pH 8, in H$_2$O, 25°. (See Materials and Methods and Hull and Sykes (1974) for further details.)
RESONANCE POSITION OF m$_{19}$F-TYROSINE

PPM
Figure 11. $^{19}$F NMR spectra of fd DNA addition to fluorotyrosyl gene 5 protein: Top - fluorotyrosyl gene 5 protein alone as in Figure 10; Bottom - addition of equal volume (3-fold molar excess) of fd DNA to $7.3 \times 10^{-4}$ M fluorotyrosyl gene 5 protein, $1 \times 10^{-2}$ M Tris-HCl (pH 8), $5 \times 10^{-4}$ M Na$_2$EDTA. (See Materials and Methods for other conditions.)
\[ \text{F-NMR SPECTRUM} \]

\[ \text{F-Tyrosyl 5P} \]

\[ \text{RESONANCE POSITION OF m}^{19}\text{F-TYROSINE} \]

\[ + \text{ fd DNA} \]

\[ \text{PPM} \]

55.0 56.2 58.2 58.9
Figure 12. $^{19}$F NMR spectra of tetranucleotide addition to fluorotyrosyl gene 5 protein, $1 \times 10^{-2}$ M Tris-HCl, pH 8: a) fluorotyrosyl gene 5 protein at $7.3 \times 10^{-4}$ M (0.250 mls); b) addition of 0.060 ml of tetranucleotide ($3.42 \times 10^{-2}$ M in H$_2$O) to sample in (a) to give an 8:1 nucleotide:protein molar ratio; c) addition of further 0.090 ml of tetranucleotides to give a 20:1 nucleotide:protein molar ratio. (See Materials and Methods for other conditions.)
a tetranucleotide. An up-field shift of the magnitude observed here could be explained by the exposed tyrosyls "stacking" with the nucleotide bases resulting in the fluorine nuclei becoming "shielded" by the nucleotide base ring current induced magnetic fields (Giessner-Prettre and Pullman, 1970). The "internal" tyrosyls apparently experience little change in magnetic environment upon protein-DNA interaction. Changes in the magnetic environment of the aromatic residues of gene 5 protein on the binding of a tetranucleotide have been confirmed by proton NMR (see Appendix).

Production of Additional Fluorine-labelled Proteins

An attempt to produce fluoro-phenylalanine labelled gene 5 protein by infecting an *E. coli* phenylalanine auxotroph (KA 197) with fd in the presence of p-fluorophenylalanine failed when the bacteria stopped growing soon after addition of the fluoro-amino acid. There was no net phage production.

The *E. coli* tyrosine auxotroph (AT 2471) used to make the fluoro-gene 5 protein was shown to be sensitive to T4 infection but was not lysed by the T4 (am33, am55) gene 32 protein overproducer mutant. An attempt to produce fluoro-tyrosyl gene 32 protein was unsuccessful however; the gene 32 protein could be found in the eluant of the extract from a DNA-cellulose column.

Maleylation of Amino Groups in Gene 5 Protein

The reaction of proteins with maleic anhydride introduces negatively charged maleyl groups at $\alpha$ and $\varepsilon$ amino moieties (Freedman et al., 1968). Gene 5 protein, $1.8 \times 10^{-4}$ M, in 0.15 M borate buffer, 0.15 M NaCl, pH 9, was reacted with a 40-fold molar excess (calculated per lysyl
residue present) of solid maleic anhydride for 5 min at 25°, followed by treatment with 1 M NH₂OH to remove any maleylated hydroxyl groups. The value of n, the number of maleyl groups introduced was calculated from the formula (Freedman et al., 1968) (see Figure 13):

\[
n = \frac{A_{280}^{\text{5p}} - A_{250}^{\text{5p}}}{A_{250}^{\text{mal-NH}} - A_{280}^{\text{mal-NH}}}
\]

Using the values from Figure 13 in equation 1: \( A_{280} = 0.142, A_{250} = 0.410, \epsilon_{280}^{5p} = 6900, \epsilon_{250}^{5p} = 3150, \epsilon_{280}^{\text{mal-NH}} = 310, \epsilon_{250}^{\text{mal-NH}} = 3360 \text{ M}^{-1} \text{ cm}^{-1} \);

\( n = 6.8 \) moles of maleyl groups per mole of protein. Apparently the ε-amino groups of all 6 lysyl residues of the protein and the α-amino groups of the N-terminal methionyl residue are maleylated. In an amino acid analysis of the maleylated protein, 40% of the cysteine appeared in the position of carboxymethyl cysteine - a possible side reaction product of the maleylation reaction. While this is a significant amount - though some may have been formed artifactually during hydrolysis - the modification of the cysteine residue is not a major factor in the alteration of gene 5 protein's structural and functional properties by maleylation. There is no binding of the maleylated protein to fd DNA as determined by the CD assay.

Chemical Modification of the Single Sulfhydryl Group of Gene 5 Protein

Ellman's reagent, 5,5'-dithiobis(2-nitrobenzoic acid) (Ellman, 1959), is known to react specifically with accessible protein sulfhydryl groups producing a characteristic absorbance at 412 nm. The reagent does not react with native gene 5 protein. In contrast Ellman's reagent reacts rapidly with 1 mole of sulfhydryl group per mole of protein when the protein is first denatured in 6 M guanidine HCl.
Figure 13. Absorption spectra of maleylated gene 5 protein, 0.01 M Tris-HCl (pH 8), 0.001 M Na₂EDTA, 0.15 M NaCl: (-----) maleylated gene 5 protein, 1.6 x 10⁻⁵ M; (-----) unreacted gene 5 protein, treated in parallel except for exposure to maleic anhydride.
Dialysis of native gene 5 protein against equimolar (50 μM) mercuric chloride in 0.05 M Tris-HCl, 0.15 M NaCl, pH 8 results in the binding of 1 mole of Hg(II) per mole of protein as determined by atomic absorption. The mercurated protein does not form a complex with fd DNA and undergoes major conformational changes (see below). Prior formation of the complex with fd DNA, however, completely prevents the binding of Hg(II) to the protein. Gel filtration of the mercury-treated gene 5 protein-fd DNA complex on a 30 x 1 cm column of G-25 Sephadex completely removes the Hg(II). The CD of the excluded material is the same as that of the native gene 5 protein-fd DNA complex.

An intriguing phenomenon possibly related to the tight binding of mercury to the sulfhydryl of unprotected gene 5 protein was observed upon binding the native protein to a mercurated analogue of poly [d(A·T)] which has a deoxy-5-mercuriuridine replacing the thymidine (Dale et al., 1975; Dale and Ward, 1975). The CD of the poly [d(A·U-Hg)] changed in the expected manner upon the addition of gene 5 protein up to a 1:4 protein to nucleotide molar ratio. However the CD of the polymer and of the protein 228 nm peak could not be returned to that expected for the dissociated molecules even in the presence of 200 mM Mg²⁺ (see Figure 4). This irreversibility might possibly involve a mercury-sulfhydryl interaction, but it has not yet been investigated further.

Dialysis of gene 5 protein (3 x 10⁻⁴ M) against 10⁻² M iodoacetate or 3-bromopyruvate for 24 hr does not result in any change in the absorption or circular dichroism spectra or in the DNA-binding properties of the treated proteins.
Conformation of the Chemically Modified Gene 5 Proteins

The CD spectrum of the chemically modified species of gene 5 protein can be taken as a measure of the integrity of the protein secondary structure. Since the large positive ellipticity band at 228 nm in the native protein is probably a reflection of the particular environment of the tyrosyl chromophores conferred by the tertiary folding of the protein (Day, 1973), the CD in this case may also be considered to reflect the intactness of the tertiary folding.

The CD spectra of the several modified proteins are summarized in Figure 14. The CD spectrum of unmodified gene 5 protein dissolved in 6 M guanidine HCl is that expected for a polypeptide in the random coil conformation. (The CD spectrum of the protein in 6 M urea however looks like that of the native gene 5 protein.) A similar spectrum is obtained for gene 5 protein reacted with maleic anhydride; the 228 nm band has disappeared and the CD spectrum is dominated by a large negative ellipticity band at 205 nm (Figure 14). Another similar spectrum (not shown) is observed for the Hg(II) protein. The CD spectrum of fully acetylated gene 5 protein is very slightly altered from that of the native protein. A slight shift is reversed on treatment of the protein with 1 M NH$_2$OH.

Nitration only slightly changes the contours of the CD spectrum of gene 5 protein. The ellipticity band at 228 nm is shifted 3 nm to longer wavelength and the long wavelength side of this band is broadened (Figure 14). The nitrated protein also shows increased ellipticity in the region around 280 nm consistent with the increased absorption of nitrotyrosine in this region. Reduction of the nitro groups to amino groups shifts the band back to 228 nm (Figure 15), consistent with the
Figure 14. Circular dichroism of native and chemically modified gene 5 protein (5p). Conditions: protein, $2 \times 10^{-5}$ M, in 0.01 M Tris-HCl, pH 8. Ellipticity at wavelengths greater than 240 nm is shown magnified 10x. (-----) native 5p; (•••••) fully acetylated 5p; (-----) nitrated 5p; (-----) maleyated 5p.
CD OF CHEMICALLY MODIFIED fd 5p

NATIVE 5p
ACETYLATED
NITRATED
MALEYLATED

$[\theta] \times 10^8$ deg cm$^2$/dmole

$\lambda$, nm

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 15. Circular dichroism of aminotyrosine gene 5 protein, $3.8 \times 10^{-5}$ M in 0.01 M Tris-HCl, pH 8.
increase in pK of the phenolic hydroxyls induced by reduction. Thus the nitratated protein does not appear to exhibit major changes in secondary or tertiary structure.

Structure of the E. coli DNA-Unwinding Protein and the Gene 32 Protein

The absorption spectrum of the *E. coli* DNA-unwinding protein in Figure 16 ($\lambda_{\text{max}} = 280$ nm, $\lambda_{\text{min}} = 255$ nm, $280/260 = 1.75$) is remarkable only because there is a consistent, large discrepancy in the protein concentration estimated using a rough extinction coefficient calculated (Wetlaufer, 1962) from the amino acid composition (Table III) in comparison to that determined directly by amino acid analysis corroborated by magnetic circular dichroism measurements of tryptophan content. Concentration values obtained using the amino acid analyzer were used throughout this work. Even at concentration values of less than 1 mg/ml the *E. coli* protein precipitated from solution when refrigerated (4°C) thus further complicating manipulations.

The circular dichroism from 320 to 190 nm of the *E. coli* unwinding protein and gene 32 protein are shown in Figure 17. The general conformation of the polypeptide backbone must be similar in both proteins, since the far ultraviolet CD spectra of the two proteins show similar band positions and mean residue ellipticities. Both have prominent minima at 208 nm and maxima of approximately equal amplitude at 192-194 nm. Secondary minima near 222 nm are less prominent. Mean residue ellipticities at representative wavelengths are nearly identical for the two proteins: $[\theta_m]_{208} = -9900$, $[\theta_m]_{217} = -7650$, $[\theta_m]_{222} = -6050$ for the gene-32 protein; and $[\theta_m]_{208} = -9700$, $[\theta_m]_{217} = -7300$, $[\theta_m]_{222} = -4850$ for the *E. coli* unwinding protein. Application of the graphical method of Greenfield and Fasman (1969) to fit these CD spectra between
Figure 16. Absorption spectrum of *E. coli* DNA-unwinding protein, 1.6 x $10^{-5}$ M in $1 \times 10^{-2}$ M Tris-HCl (pH 8), $1 \times 10^{-4}$ M dithiothreitol, 0.015 M NaClO₄, 10% glycerol.
<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Native</th>
<th>Nitrated</th>
<th>fd DNA &quot;Protected&quot;</th>
<th>Trypsin Treated</th>
<th>AA Residue Difference from Native</th>
<th>E. coli &quot;unwinding&quot; protein</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gene 32 Protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E. coli &quot;unwinding&quot; protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Native</td>
</tr>
<tr>
<td>Tyr</td>
<td>8.6</td>
<td>3.6</td>
<td>8.3</td>
<td>9.4</td>
<td>0</td>
<td>5.2</td>
</tr>
<tr>
<td>NO₂-tyr</td>
<td>-</td>
<td>5.1b</td>
<td>0.2b</td>
<td>-</td>
<td></td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Asp</td>
<td>55.2</td>
<td>52.8</td>
<td>49.1</td>
<td>39.5</td>
<td>-15</td>
<td>19.0</td>
</tr>
<tr>
<td>Thr</td>
<td>12.6</td>
<td>12.5</td>
<td>13.5</td>
<td>10.4</td>
<td>-2</td>
<td>15.7</td>
</tr>
<tr>
<td>Ser</td>
<td>21.3</td>
<td>21.4</td>
<td>22.4</td>
<td>19.5</td>
<td>-2</td>
<td>13.9</td>
</tr>
<tr>
<td>Glu</td>
<td>28.6</td>
<td>29.0</td>
<td>31.0</td>
<td>29.5</td>
<td>0</td>
<td>30.6</td>
</tr>
<tr>
<td>Pro</td>
<td>9.6</td>
<td>10.0</td>
<td>10.0</td>
<td>8.9</td>
<td>-1</td>
<td>11.8</td>
</tr>
<tr>
<td>Gly</td>
<td>19.5</td>
<td>19.3</td>
<td>20.3</td>
<td>18.9</td>
<td>0</td>
<td>38.8</td>
</tr>
<tr>
<td>Ala</td>
<td>26.7</td>
<td>22.1</td>
<td>26.5</td>
<td>14.8</td>
<td>-12</td>
<td>16.4</td>
</tr>
<tr>
<td>1/2 Cys</td>
<td>4.1</td>
<td>-</td>
<td>-</td>
<td>4.0</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>Val</td>
<td>20.7</td>
<td>21.1</td>
<td>19.9</td>
<td>19.1</td>
<td>-2</td>
<td>14.4</td>
</tr>
<tr>
<td>Met</td>
<td>10.0c</td>
<td>-</td>
<td>8.9</td>
<td>5.5c</td>
<td>-4</td>
<td>5.8</td>
</tr>
<tr>
<td>Ile</td>
<td>10.0</td>
<td>10.0</td>
<td>10.7</td>
<td>11.1</td>
<td>(+1)</td>
<td>6.2</td>
</tr>
<tr>
<td>Leu</td>
<td>21.0</td>
<td>21.0</td>
<td>18.6</td>
<td>15.7</td>
<td>-5</td>
<td>10.6</td>
</tr>
<tr>
<td>Phe</td>
<td>19.1</td>
<td>16.8</td>
<td>19.2</td>
<td>16.8</td>
<td>-2</td>
<td>4.6</td>
</tr>
<tr>
<td>His</td>
<td>2.8</td>
<td>3.7</td>
<td>2.9</td>
<td>3.4</td>
<td>0</td>
<td>3.1</td>
</tr>
<tr>
<td>Lys</td>
<td>35.2</td>
<td>33.0</td>
<td>35.2</td>
<td>30.6</td>
<td>-4</td>
<td>7.2</td>
</tr>
<tr>
<td>Arg</td>
<td>4.1</td>
<td>3.7</td>
<td>4.4</td>
<td>3.6</td>
<td>0</td>
<td>9.5</td>
</tr>
<tr>
<td>Trp</td>
<td>5.0</td>
<td>-</td>
<td>-</td>
<td>(5)</td>
<td>(0)</td>
<td>3.1</td>
</tr>
</tbody>
</table>

**TOTAL** 314 266 -48 207

*Determined using MCD.

*Determines both spectrally and by amino acid analysis.

*Amino terminal residue is methionine in both the native and the trypsin-treated protein.
Figure 17. Circular dichroism of DNA binding proteins. A) E. coli DNA unwinding protein ($1.6 \times 10^{-5} \text{ M}$) in $0.01 \text{ M}$ Tris-HCl, $1 \times 10^{-4} \text{ M}$ dithiothreitol, $0.015 \text{ M}$ NaClO$_4$, 10% glycerol, pH 8, 25°.

B) Gene 32 protein in $0.01 \text{ M}$ Tris-HCl, pH 8, 25°. (-----) Native protein ($2.8 \times 10^{-6} \text{ M}$); (-----) nitrated monomers ($2.0 \times 10^{-6} \text{ M}$); (---) human carbonic anhydrase C.
208 and 240 nm with contributions from a combination of α, β, and random polypeptide structures gives the most adequate fit for the spectra of both proteins with the following approximate percentages of the three conformations considered: 20% α, 20% β and 60% random coil.

Optical activity of the aromatic chromophores makes a large contribution to the near ultraviolet circular dichroism of both proteins (Figure 17). The magnitude of these Cotton effects is particularly striking in the _E. coli_ unwinding protein with a negative band ([θ] ~ -5 x 10⁴ deg cm²/dmole) centered at 280 nm.

Using a molecular weight of 22,000 daltons for the _E. coli_ protein, determined by SDS-acrylamide gel electrophoresis using gels (Figure 1) calibrated with a series of known molecular weight standards including trypsin (MW 23,300) (Weber and Osborn, 1969) and in agreement with the determination of Sigal et al. (1972), it can be estimated that the protein contains approximately 207 residues (Table III). The amino acid composition for the _E. coli_ protein reported here differs little from that published by Weiner et al. (1975) but in this laboratory the protein never migrated on gels at the low 19,000 MW value reported by Weiner et al. (1975). Using the same techniques (see Figure 24 below) the gene 32 protein contains about 314 residues (Table III). The gene 32 protein contains 35 lysyl residues as well as a relative abundance of aromatic residues. The tryptophan content as determined by magnetic circular dichroism is 3 residues per molecule for the _E. coli_ protein and 5 residues per molecule for the gene 32 protein (Table III).

**CD Assay of the Binding of the Gene 32 and _E. coli_ Proteins to DNA**

As demonstrated for gene 5 protein, the change in CD of fd DNA or poly [d(A-T)] above 250 nm induced by DNA-protein complex formation can
be used to assay the formation of complexes of both gene 32 and the E. coli protein with the nucleic acids. Representative CD titrations of fd DNA with gene 32 protein and the E. coli protein are shown in Figures 18 and 19 respectively. From these titration curves the protein to base ratio at saturation of the DNA can be determined for both the E. coli and gene 32 proteins by plotting the change in ellipticity, $\Delta[\theta]$, versus the protein to nucleotide base ratio, $1/R$. The titration with gene 32 protein is complete at a $1/R$ value of 0.09 indicating that each protein covers $\approx 11$ bases (Figure 18, insert). This is close to the stoichiometry of 10 bases per protein reported from optical absorption measurements (Alberts and Frey, 1970). With the E. coli protein saturation is reached at $1/R = 0.07$ (Figure 19, insert) indicating that each protein covers approximately 14 bases, a number significantly different than the previously reported value of 8 determined from optical spectroscopy (Sigal, 1972). However, the higher number approaches the stoichiometry observed using filter binding techniques (Weiner et al., 1975).

There is no change in the far ultraviolet CD of either protein upon binding to DNA, suggesting that no gross changes in protein secondary structure are induced. The concentration of proteins and nucleic acids used and the sharp endpoint of both titrations (Figures 18 and 19) show that the dissociation constants for the complexes of both proteins with fd DNA must be less than $3 \times 10^{-8}$ M at 10 mM Na$^+$ in solution.

Gene 32 protein and the E. coli unwinding protein both bind tightly to the synthetic ribonucleotide polymer, poly A, as indicated by the large reversible changes induced in the CD spectrum of the polymer by
Figure 18. Circular dichroism spectra of fd DNA showing changes induced by the addition of gene 32 protein. fd DNA (2.5 x 10^{-5} \text{ M}) was initially present in 2.4 ml under conditions of Figure 17B. Changes in CD on addition of 10 \mu l and 20 \mu l aliquots of gene 32 protein (8.4 x 10^{-5} \text{ M}) in the same buffer to yield final protein concentration of 4.1 x 10^{-6} \text{ M} are indicated by curves 1-9. For curve 9, the protein/DNA base molar ratio = 1:5.8. Insert: Change in DNA ellipticity as a function of molar ratio of gene 32 protein/DNA bases; 1/R; (●) fd DNA (270 nm); (○) poly [d(A·T)] (260 nm).
Figure 19. Circular dichroism spectra of fd DNA showing changes induced by the addition of *E. coli* DNA-unwinding protein. fd DNA (2.5 x 10^{-5} M) was initially present in 2.4 ml under conditions of Fig. 17A. Changes in CD on addition of 50 µl aliquots of *E. coli* unwinding protein (1.6 x 10^{-5} M) in the same buffer to yield a final protein concentration of 1.8 x 10^{-6} M are indicated by curves 1-6. For curve 6, protein/DNA base molar ratio = 1:12.7. Insert: Change in DNA ellipticity (corrected for slight amount of ellipticity due to protein) as a function of molar ratio of *E. coli* unwinding protein/DNA bases, 1/R; (●) fd DNA (270 nm).
addition of the proteins as illustrated in Figure 20 for the gene 32 protein.

Effect of Cations on Nucleic Acid Complexes with E. coli DNA-Unwinding Protein and Gene 32 Protein

The effect of increasing Mg\(^{2+}\) concentrations on the CD spectra of the gene 32 protein-poly [d(A\*T)] complex (Figure 21) is a gradual return of the region above 250 nm to a spectra characteristic of the nucleic acid alone as complex dissociates. Data for other combinations of protein, nucleic acid, and cation are summarized in Figure 22.

In general the reversal of the CD of the complexes by increasing salt concentrations follows the concentration dependence predicted from the elution profiles seen during DNA-cellulose chromatography (Alberts and Frey, 1970; Molineaux et al., 1974; Weiner et al., 1975). As salt concentration increases, the E. coli protein complex begins to dissociate earlier than the gene 32 protein complex, but dissociation is spread over a much wider concentration range (Figure 22A). The apparent wider salt concentration range required to bring about the complete dissociation of the E. coli protein-DNA complex compared to the gene 32 protein-DNA complex does not seem to be explained by a very slow dissociation rate for the complex. The observed CD spectrum of the mixture at intermediate salt concentrations does not change upon extended incubation at room temperature. As with the gene 5 protein results (Figure 4) these dissociation profiles indicate the greater degree of dissociation caused by low concentrations of divalent metal ions as opposed to monovalent metal ions, especially with poly [d(A\*T)]. Again this reflects the additional stabilization of the double-stranded conformation of DNA by Me\(^{2+}\) rather than some direct effect on the proteins.
Figure 20. Circular dichroism changes upon addition of gene 32 protein to poly A: (-----) 59 nmole of poly A in 2.5 ml 0.01 M Tris-HCl, pH 8, 0.001 M EDTA; (-----) 1-4, poly A with addition of 25 µl and 50 µl aliquots of stock gene 32 protein (8.4 x 10^{-6} M) to yield protein to nucleotide base molar ratios of (1) 1:14, (2) 1:7, (3) 1:6, (4) 1:4; (---) gene 32 protein-poly A complex dissociated with addition of 375 µl 4 M NaCl to give Na^+ concentration of 0.52 M. Difference between poly A curve and that of dissociated complex can be accounted for by a summation of the intrinsic ellipticity of the protein, direct Na^+ effects on poly A ellipticity, and dilutional effects.
Figure 21. Circular dichroism changes upon reversal of gene 32 protein-poly [d(A·T)] complex with $\text{Mg}^{2+}$: 110 $\mu$l of stock gene 32 protein solution ($8.3 \times 10^{-5}$ M) was added to 57 n mole of poly [d(A·T)] in 2.4 mls of 0.01 M Tris-HCl, pH 8, yielding a solution of complex with slight excess of protein. 5 and 10 $\mu$l aliquots of 4 M MgCl$_2$ were added to dissociate the complex. Final concentration of Mg$^{++}$ (curve 6) is 0.08 M. (-----) poly [d(A·T)] alone; (•••) curve 0, poly [d(A·T)] plus excess gene 32 protein without Mg$^{2+}$; (——) curves 1-6, addition of Mg$^{2+}$ to concentrations of 0.01, 0.02, 0.04, 0.05, 0.06, and 0.08 M respectively. The difference between the curve representing the dissociated complex (curve 6) and that for poly [d(A·T)] alone represents the summation of intrinsic protein ellipticity, dilution effects and Mg$^{2+}$ induced changes on poly [d(A·T)] ellipticity.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 22. Dissociation of DNA-protein and RNA-protein complexes by cations as followed by circular dichroism. Saturating amounts of protein were added to solutions of nucleic acid (2.5 x 10^{-6} M) under salt and temperature conditions as in Figure 17, mixed, and a CD spectrum taken. Aliquots of concentrated (4 M) solutions of the chloride salts of the cations of interest were added, thoroughly mixed for ten minutes, and a CD spectrum taken. This process was repeated until no further change in the CD spectra could be noted. At the end point, the CD spectrum coincided with the spectrum of free nucleic acid at similar salt concentrations in each case. The % maximum ellipticity change (at 260 nm for poly [d(A-T)], at 270 nm for fd DNA, and at 265 nm for Poly A complexes) is plotted as a function of the molarity of added cation. 

A) E. coli unwinding protein: (●) Mg^{2+} - poly [d(A-T)]; (▲) Mg^{2+} - fd DNA; (△) Na^{+} - fd DNA. 
B) Gene 32 protein: (□) Mg^{2+} - poly [d(A-T)]; (■) Mg^{2+} - fd DNA; (◇) Na^{+} - poly A; (△) K^{+} - poly [d(A-T)]; (○) Na^{+} - poly [d(A-T)]; (●) K^{+} - fd DNA; (▲) Na - fd DNA.
The interaction of the E. coli unwinding protein with poly [d(A*T)] is completely abolished at less than 0.01 M Mg\(^{2+}\) (Figure 22A), while greater than 0.05 M is required to completely dissociate the gene 32 protein-poly [d(A*T)] complex (Figure 22B). The gene 5 protein-poly [d(A*T)] complex dissociates at an intermediate value of about 0.02 M Mg\(^{2+}\). These differences must reflect structural differences in the respective complexes; electron microscopy suggests a different form of gross supercoiling for the complexes with the three proteins as will be discussed later. The gene 32 protein-poly A interaction is less stable than that with fd DNA, but it still requires approximately 0.4 M Na\(^{+}\) to completely dissociate the complex (Figures 20 and 22B).

Nitration of Gene 32 Protein and the E. coli DNA-Unwinding Protein

Gene 32 protein was nitrated under the same conditions used for the gene 5 protein modification: 100-fold molar excess of C(NO\(_2\))\(_4\) in 0.05 M Tris-HCl, pH 8, reacted for 30 minutes at 25°. This results in the nitration of 5 of the 9 tyrosyl residues in the protein determined spectroscopically (Figure 23) and by amino acid analysis (Table III). The relative magnitudes and contours of the MCD bands corresponding to the tryptophanyl absorption bands at 293 and 286 nm are unchanged in the nitrated protein suggesting that no modification of tryptophanyl residues has occurred. Hydrolysis of the modified protein yields no additional cysteic acid suggesting that extensive oxidation of cysteine is not induced by the C(NO\(_2\))\(_4\) reaction. Thus it is concluded that only the 5 tyrosyls are modified in the reaction. Their nitration radically alters the near ultraviolet CD of the aromatic residues (Figure 17) and destroys binding of the protein to DNA as assayed by the CD binding assay. Nitration does not significantly alter the secondary structure.
Figure 23. Absorption spectrum of native and nitrated gene 32 protein in $1 \times 10^{-2}$ M Tris-HCl, pH 8: (-----) native gene 32 protein, $1.7 \times 10^{-5}$ M; nitrated gene 32 protein, $1 \times 10^{-5}$ M.
of the polypeptide, since the far ultraviolet CD is unchanged after the reaction (Figure 17). The alteration of the CD of the aromatic chromophores induced by nitration may represent some change in their conformation, but it may reflect a contribution from the intrinsic optical activity of the nitrotyrosyl chromophore which has additional strong absorption bands in the 270-280 nm region (Sokolovsky et al., 1966). Prior complexation of gene 32 protein with fd DNA prevents the nitration (Table III) and prevents the change in the CD of the aromatic chromophores. Thus the native conformation and electronic structure of 5 of the 9 tyrosyl residues of gene 32 protein are required for DNA binding.

In contrast to both gene 32 protein and gene 5 protein, none of the tyrosyl residues in the E. coli DNA unwinding protein are available for nitration (Table III).

Nitration of gene 32 protein using tetranitromethane results in crosslinking of the protein into dimers, trimers and larger oligomers (Figure 24A), a reaction which also occurs on nitration of the gene 5 protein. All experiments on DNA binding and physico-chemical properties of the nitroprotein described here were performed on nitrated monomers separated from the other oligomers by Sephadex G-75 chromatography. Reaction of the protein with tetranitromethane in the presence of fd DNA results in minimal dimer formation (Figure 24B) as expected from the shielding of the reactive tyrosines (Table III).

**Enzymatic Cleavage of Gene 32 Protein**

Immediately after the final purification step, gene 32 protein migrates on polyacrylamide gels as a single band of M.W. 35,000. However, throughout the earlier purification stages a smaller band consistently appears at a position indicating a MW of ~30,000 daltons. This band
Figure 24. SDS-polyacrylamide gel (10%) showing: A) nitratred gene 32 protein; B) Gene 32 protein bound to fd DNA before reaction with tetranitromethane; C) Purified gene 32 protein stored at 4°C in 10 mM Tris-HCl (pH 8) for 1 week; D) Gene 32 protein after treatment with trypsin as in text; E) Gene 32 protein bound to fd DNA before trypsin treatment and recovered in void volume with DNA after chromatography on Sephadex G-100; F) Protein standards as in Methods; G) Purified gene 32 protein stored as in (C) for 6 weeks; H) Total mix after trypsin digest of gene 32 protein; I) Material from trypsin digest applied to DNA-cellulose column and recovered in initial wash (non-binding); J) Material from trypsin digest of gene 32 protein applied to DNA-cellulose column and recovered after 0.6 M salt elution as expected for the native protein.
reappears and increases with storage of the protein, particularly when stored prior to the final purification steps, and derives from the gene 32 protein itself (Figure 24C). Hosoda et al. (1974) have independently found a similar phenomenon when they purified gene 32 protein from an overproducer strain of T4 similar to the one employed in these studies.

Treatment of the intact protein with trypsin (1:100 w/w for 2 hours at 25°C) converts almost all of the protein to the 30,000 MW form with a few lower molecular weight fragments detectable (Figure 24D). This large fragment can be purified by binding to a column of DNA-cellulose; it can be eluted from the column with 0.6 M NaCl (Figure 24J). Trypsin treatment of gene 32 protein in the presence of fd DNA gives the same results as with the protein alone. Figure 24E shows a sample from the void volume fraction eluted from a Sephadex G-100 column to which a trypsin digest of the complex was applied. The 30,000 MW peptide moves with the large DNA molecule.

The CD of the 30,000 MW peptide and its binding to fd DNA (by CD assay) are similar to these parameters as measured for native gene 32 protein, implying that the major structural properties of the protein are preserved after cleavage (Figure 25). There is, however, a significant increase in ellipticity of the positive CD band at 250 nm and a significant decrease in the magnitude of the negative band at 208 nm, as well as a red shift to 210 nm. These changes imply at least moderate conformational changes in the protein upon removal of the small peptide. Trypsin treatment of the gene 32 protein-fd DNA complex also results in the conversion of all the protein to the 30,000 MW form. The modified protein-DNA complex migrates into the void volume of a Sephadex G-100 column (10 mM Tris-HCl, pH 8), again indicating that the binding function

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 25. Circular dichroism of gene 32 protein after digestion with trypsin and purification of active, cleaved 30,000 MW with DNA-cellulose chromatography: (---) cleaved protein at $3.6 \times 10^{-6}$ M in $1 \times 10^{-2}$ M Tris-HCl, pH 8; (-----) intact gene 32 protein spectrum for comparison (from Figure 17).
of the cleaved protein is intact.

Amino acid analysis of the 30,000 MW peptide reveals that approximately 50 residues are lost from the native gene 32 protein. Fifteen of the cleaved residues are aspartic acid. Depending on the unknown amount of amidation present, this could represent a loss of considerable negative charge. There is also a loss of 5 positively charged lysyl residues, however. A majority of the cleaved residues are accounted for by two species, alanyl and aspartyl residues. Thus the cleaved region has an unusual amino acid composition and could consist of some type of repeated structure. The amino terminal residue of both the native gene 32 protein and the trypsin-cleaved protein is methionine suggesting, but not proving, that cleavage occurs at the carboxy end of gene 32 protein.
CHAPTER IV

DISCUSSION

The CD assay used for protein-DNA binding in this study appears to be a reliable measure of the interactions of "DNA-unwinding" proteins with single-stranded DNA. The reversal of the CD spectra of the DNA in the complexes to that of free DNA on dissociation of the binding protein provides a precise method for determining the effect of various cations on the stability of the protein-DNA complexes as well as demonstrating that the induced CD changes are a result of a specific protein-DNA interaction. The relative sensitivity of the protein-fd DNA complexes to dissociation by various species of cations as measured by the CD changes (Figures 4 and 22) correlates with the known affinities of cations for nucleotides (Ross and Scruggs, 1964; Gordon, 1965; Sander and Tso, 1971). The reversal of the CD of the complexes as a function of the concentration of cation is also roughly consistent with the observed elution behavior of unwinding proteins from DNA cellulose columns (Alberts et al., 1972; Alberts and Frey, 1970; Molineux et al., 1974). Other techniques of measuring complex formation show similar salt sensitivities; for example, Sigal et al. (1972), using absorption spectroscopy, report that the E. coli protein-T4 DNA complex is disrupted by 30 mM Mg$^{2+}$; Oey and Knippers (1972) state that formation of the gene 5 protein-fd DNA complex is prevented by 0.4 M K$^+$ using a nitro-cellulose filter binding assay. The gene 32 protein and the gene 5 protein stoichiometries in their complexes as indicated by the CD assay (Figures -87-.
2 and 18) are in agreement with the stoichiometries determined by other methods (Alberts et al., 1972; Pratt et al., 1974; Alberts and Frey, 1970). Electron micrographs of complexes formed from our preparations of gene 5 protein and fd DNA appear identical to those previously published by Alberts et al. (1972).

The different value found for the stoichiometry of the _E. coli_ unwinding protein-DNA complex found in this work as opposed to that determined by Weiner et al. (1975), Molineux et al. (1974), and Sigal et al. (1972) might be explained by the different techniques used to determine protein concentration. The investigators quoted used the technique of Lowry et al. (1951) in their concentration determinations (standard(s) used were not stated) while this study used amino acid analyses. Lowry et al. (1951) note that up to a 3-fold difference in color yield can be obtained with different peptides. Weiner et al. (1975) note that the staining efficiency of the unwinding protein with Coomassie brilliant blue was much different than that of a bovine serum albumin standard. While the Coomassie brilliant blue staining mechanism is unknown but probably different than that of the Folin-Lowry reagents, the question of anomalous readings with this concentration determination technique does arise. Also the results with the Coomassie staining might be explained if less actual protein was being put on the gels than indicated by the Lowry determination.

The extinction coefficient at 280 nm for the _E. coli_ DNA unwinding protein calculated from the amino acid composition (Table III) by the method of Wetlaufer (1962), 23,350, is only 50% of that calculated from the absorption spectra and concentration (determined on the amino acid analyzer) of 43,750. It does not seem likely that a systematic error.
was made in calculating concentrations because the concentration and protein composition of tryptophan determined independently using magnetic circular dichroism (McFarland and Coleman, 1972) are internally consistent. The amino acid composition data for aromatic and other residues obtained here agrees quite closely with that of Weiner et al. (1975) except for the presence of a sulfhydryl residue in our analysis, if the different molecular weights used (22,000 and 19,500 respectively) are taken into account. The anomalous extinction coefficient remains unexplained; the possible presence of some prosthetic group or carbohydrate moiety has not been ruled out.

A significant feature of these CD titrations is that the contribution of the DNA to the CD of the complex (i.e. the portion of the final spectra above ~250 nm) is approximately the same (when corrected for a small contribution from the protein) whether the E. coli, gene 32, or gene 5 proteins are used to form the complex. The same holds for the complexes with poly[d(A·T)]. Thus the final CD of the DNA does not appear particularly sensitive to subtle differences in the conformation of the DNA, differences which clearly must exist between the three complexes judging from the length measurements and the gross features of electron micrograph images of the complexes (see below). While definitive statements about the structure of the DNA in the complex with gene 5 protein cannot be made from the CD, the dramatic reversal of the 269-270 nm ellipticity band is similar to the CD changes seen when DNA is complexed with polylysine (Chang et al., 1973). Similar changes, however, are observed when DNA is dissolved in concentrated alcohol (Green and Mahler, 1971; Girod et al., 1973) or in very concentrated salt solutions (Studdert et al., 1972). The basis of these CD changes
could include dehydration in the vicinity of the DNA molecule and charge neutralization of the phosphate groups of the DNA (Chang et al., 1973; Wolf and Hanlon, 1975), aggregation or condensation as occurs in alcohol (Girod et al., 1973), as well as changes in the secondary structure of the DNA. The changes in the CD spectra of the nucleic acids does resemble that of the shift from a predominantly "B" structure to one similar to the "C" form as seen by Hanlon et al. (1975) when increasing concentrations of salt was added to calf thymus DNA and by Chang et al. (1973) upon adding polylysine to DNA. The "C" type CD spectra is also seen with the non-specific aggregated states formed by DNA in concentrated alcohol or polymer solutions however (Gorod et al., 1973; Jordan et al., 1972; Green and Mahler, 1971).

It is thought that the formation of salt linkages between protein lysyl e-amino groups and the DNA backbone phosphates provide a major part of the energy for many stable DNA-peptide interactions, particularly those involving histones (Malchy and Kaplan, 1974; Simpson, 1971). A number of studies have dealt with polylysine interactions with nucleic acids as a model for protein-nucleic acid systems. In most cases the mixture of polylysine and DNA form micellar or liquid crystal phases (Haynes et al., 1970; Shapiro et al., 1969). The interaction under appropriate conditions is stoichiometric with one lysine per nucleotide and shows some preference for A + T-rich DNA (Shapiro et al., 1969). The melting temperature of polylysine complexed DNA is increased (Pinkaton and Li, 1974) as one would expect upon adding effective counter-ions to the backbone phosphates. It is interesting that Hanlon et al. (1974) have found using CD and salt dissociation techniques that histone binding does give chromosomal DNA a CD spectrum characteristic
of DNA in the "C" conformation.

Binding of polyarginine to nucleic acids shows characteristics similar to polylysine binding as monitored by increases in melting temperature and induced CD changes (Epstein et al., 1974; Yu et al., 1974). There exists some disagreement as to the exact form arginine-phosphate bonds exist in solution; they are apparently at least as strong as lysine-phosphate bonds and more structurally specific (Saenger and Wagner, 1972; Cotton et al., 1973).

The dissociation of the gene 5 protein-fd DNA complex induced by salt and the prevention of complex formation by N-acetylation of the lysyl ε-amino groups suggest that the lysyl residues of gene 5 protein play a major role in formation of the DNA complex. As might be expected, all the ε-amino groups of the lysyl residues appear to be on the surface of the molecule and are susceptible to acetylation and maleylation. The introduction of acetyl groups at the ε-amino groups does not appear to produce changes in the secondary or tertiary structure of the protein, since no change occurs in the CD of the acetylated protein in which the O-acetyl groups have been removed (Figure 14). The rapid disruption of preformed gene 5 protein-DNA complexes by mild acetylation of the complex suggests that the lysyl ε-amino-phosphate linkages may occur on the exterior of the complex.

While there is no direct evidence, the salt sensitivity of the nucleic acid complexes of the gene 32 protein and E. coli DNA-unwinding protein are consistent with an important role for basic residues in the electrostatic interactions with nucleic acid backbone phosphates in complex formation. That over 10% of the amino acid residues of gene 32 protein are lysines is consistent with this concept (Table III).
Tyrosyl residues have been implicated in the active sites of two enzymes involved in the processing of DNA: *Staphylococcus aureus* extracellular nuclease, by X-ray diffraction (Arnone et al., 1971), fluorescence studies (Cuatrecases et al., 1967), and chemical modification (Cuatrecases et al., 1968), and pancreatic deoxyribonuclease I, by chemical modification (Hugli and Stein, 1971). Fluorescence studies have indicated an interaction between the tyrosyl residues of histones and DNA (Matsuyama and Nagata, 1970). The tyrosyl-rich amino-terminal portion of the lac repressor apparently contains the amino acid sequence which binds to the lac operator (Adler et al., 1972; Beyreuther et al., 1973; Platt et al., 1973). Iodination of the lac repressor, under conditions such that most of the label appears on tyrosine in the amino-terminal tryptic peptide blocks binding to the lac operator (Fanning, 1975).

Many model systems have also been exploited to look at tyrosyl-nucleotide interactions. Friedman and Ts'o (1971) using sucrose gradient electrophoresis and polymers of various amino acids found that poly-L-tyrosine was the only "uncharged" polypeptide that could bind to nucleic acids. Novak and Dohnal (1973) and Santella and Li (1974), using alternating copolymers of tyrosine with glycine and lysine respectively, showed that poly-tyrosyl binding to nucleic acids occurred and resulted in an increased melting temperature when bound to double-stranded DNA. Helene et al. (1971) have demonstrated that tyrosine and tyramine in solution mixed with DNA and RNA lead to fluorescence quenching and NMR up-field shifts consistent with a "stacking" interaction between the phenol rings and nucleotide bases. Similar studies employing short oligopeptides with various combinations of aromatic
(tyrosyl, phenylalanyl, tryptophanyl) and basic residues have yielded fluorescence and NMR evidence that at least two types of complexes form with nucleic acids (Dimicoli and Helene, 1974a,b; Gabbay et al., 1972,1973; Brun et al., 1975; Toulmé et al., 1974). Durand et al. (1975) have also used CD to look at these mechanisms. The first mechanism, intercalation, involves partial overlapping of the aromatic ring of the amino acid with the heterocyclic ring of the nucleotide bases and induces changes in the nucleic acid conformation (Helene et al., 1971; Gabbay et al., 1973; Sellini et al., 1973; Dimicoli and Helene, 1974). A preference of tyrosine for intercalation with A-T over G-C sites has been observed (Gallay et al., 1973). This mode of binding is favored with single stranded polynucleotides (Brun et al., 1975; Toulmé et al., 1974). A second mechanism of binding, particularly with tyrosyl-containing oligopeptides, is hydrogen bonding. Infrared and proton magnetic resonance studies in organic solvents have demonstrated the ability of analogues of tyrosine to hydrogen bond with the DNA bases via the phenolic hydroxyl (Sellini et al., 1973). (Carboxylic acids also hydrogen bond to DNA bases.) Santella and Li (1974) have pointed out that much of the data using model systems to demonstrate specific modes of interaction are far from conclusive; much of the NMR shifts and broadening could be well explained by tyrosine stacking along the nucleic acid phosphate backbone and fluorescence quenching can also arise from deprotonation of the tyrosyl hydroxyls in complex formation.

Another type of model system thought to be appropriate to protein-DNA interactions is the binding of large planar aromatic ring dyes and antibiotics to nucleic acids. Sobell (1973) has proposed a model based
on x-ray diffraction results which summarizes the features of the much studied actinomycin D interaction with double-stranded DNA. In his model the actinomycin D phenoxazone ring intercalates between G-C base pairs while the cyclic peptide moieties fit into the minor groove of the helix and form specific hydrogen bonds with nucleotide base groups as well as hydrophobic bonds with the DNA sugars (Sobell et al., 1971; Jain and Sobell, 1972; Sobell and Jain, 1972; Sobell, 1973; Sobell, 1974). Since the presentation of this model proton and phosphorus NMR studies (Krugh and Neely, 1973; Patel, 1974a,b) as well as kinetic studies (Bittman and Blau, 1975) have been consistent with its major features. The binding mechanism of actinomycin D to DNA is intrinsically interesting because of its ability to block in vivo transcription by DNA-dependent RNA polymerase at the elongation step (Hyman and Davison, 1970).

A similar intercalation model for the binding of the drug ethidium bromide to dinucleotides has also been proposed based on x-ray diffraction results (Tsai et al., 1975). This planar ring molecule binds to both RNA and DNA (Waring, 1965) and also acts in vivo by blocking DNA-directed RNA polymerase activity but unlike actinomycin D its major effect is apparently on the initiation step (Richardson, 1973). Crystal structures of the aminoglycosyl antibiotic puromycin and of the mutagen 9-aminoacridine with a dinucleotide further confirm the possibility of intercalation as an important mechanism in complex formation with DNA (Sundaralingam and Arora, 1972; Seeman et al., 1975). To balance the picture, Festy and Daune (1973) report that fluorescence measurements indicate that the trypanocidal drug hydroxystilbamide, composed of two substituted benzyl rings connected by an ethylene linkage, binds
to DNA using a hydrogen bonding mechanism. A hydroxyl group on one ring increases the molecule's specificity. *In vivo* it apparently disrupts DNA replication.

Three of the five tyrosyls of gene 5 protein are on the surface of the molecule exposed to solvent as indicated by their reactivity toward acetylation and nitration as well as their chemical shift seen with $^{19}$F NMR. Myers and Glazer (1971) make the argument that susceptibility to modification is not a sufficient criteria of residue exposure, however Figure 9 indicates that nitration changes the pK's of the 3 tyrosyls which initially ionized at a pH characteristic of free tyrosine in solution as opposed to the two tyrosyls which did not ionize until a quite high pH was obtained. In addition, a number of the tyrosyls in this protein, though there is no data to particularly implicate specific residues, are involved in a structure, possibly some form of stacking, that markedly enhances the optical activity of their chromophores (Beychok and Fasman, 1964; Simmons and Glazer, 1967; Chen and Woody, 1971).

The protection by DNA of the three accessible tyrosyls of gene 5 protein from nitration with TNM indicates that these tyrosyls, initially on the surface of the isolated protein, become buried in the gene 5 protein-DNA complex. The tightness of the interaction is brought out by the observation that binding to a single tetranucleotide can shield the tyrosines from $\text{C(NO}_2\text{)}_4$ (Table I). This is consistent with the decrease in ellipticity of the protein at 228 nm accompanying formation of the DNA complex which has been interpreted as reflecting a shift of the tyrosyl residues to a more hydrophobic environment (Day, 1973). These three exterior tyrosyl residues could interact either with the
DNA by hydrogen bonding or intercalation between bases or with the adjacent gene 5 protein molecules.

Nitro substitution on the phenoxazone ring of actinomycin does not interfere with its ability to intercalate with DNA as reflected in the similar dissociation constants of the DNA complex with the substituted and unsubstituted antibiotic (Müller and Crothers, 1968). However, the situation with the nitrated tyrosines of gene 5 protein and single-stranded DNA may not be entirely analogous to the models for intercalation with double-stranded DNA (Sobell, 1972). The spectral characteristics of model compounds suggest that the nitro group ortho to the ionized phenolic oxygen is forced to deviate from a coplanar alignment with the phenolate ring. This could increase the effective ring thickness from 3.4 Å to greater than 5 Å. Thus the loss of binding affinity of nitrated gene 5 protein for DNA might involve interference of the nitrotyrosine ring with intercalation. The increased bulk of the nitrophenol ring might also sterically interfere with complex formation by coming into contact with the nucleic acid backbone. Reduction of the nitrotyrosines to aminotyrosines does not alter the decreased binding to DNA (Figure 8); again the substituent group, though smaller than the nitro is not coplanar with the phenol ring.

Alternatively, the fall in the pKₐ of the phenolic hydroxyl group of the 3-nitrotyrosyl residue to ~7 (Figure 9) might introduce negative charges which interfere with complex formation or which prevent hydrogen bonding with the DNA bases. These latter alternatives are less likely since the gene 5 protein containing fluorotyrosyl residues exhibits no detectable loss of binding affinity despite the fact that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
the pK\textsubscript{a} of the tyrosyl hydroxyl groups is lowered by 2 pH units and there is no increase in binding affinity when the nitro-protein is reduced to the aminotyrosine form with a concurrent return of the phenolic pK to near normal values (Sokolovsky et al., 1967).

The intercalation mechanism might also explain the apparent changes in DNA conformation suggested by the CD spectra (Figure 2) and the electron micrographs. Up-field chemical shifts of the $^{19}\text{F}$ resonances of the surface tyrosyls observed on the binding of tetranucleotides to fluorotyrosyl gene 5 protein are compatible with the intercalation mechanism (Figure 12).

Five of the nine tyrosyls of gene 32 protein can be implicated in the DNA binding reaction using the same criteria of protection. There is of yet little further data on the tyrosyl residues of gene 32 protein. However the aromatic Cotton effects in the gene 32 protein CD spectrum between 240 and 310 nm are similar to those observed previously in carbonic anhydrase C (Figure 1) (Coleman, 1968). From the extensive solution data and the crystal structure of the latter enzyme these Cotton effects appear to arise from an aromatic cluster near the N-terminus of the molecule containing two tryptophanyl and one tyrosyl residue stacked with several prolinyl and phenylalanyl residues (Lindskog et al., 1971). This suggests that some type of aromatic residue interaction may be present in the gene 32 protein similar to that seen between tyrosyl residues in gene 5 protein.

Electron micrographs of the complexes of all three proteins with fd DNA have revealed major differences in the nature of the complexes of the three different DNA binding proteins with the same DNA. Gene 5 protein collapses circular fd DNA such that two oppositely directed
protein covered strands are pulled into a helical rod-like structure (Alberts et al., 1972; Pratt et al., 1974). It is not known whether gene 5 protein interacts with 4 bases on 1 strand or with two bases on each of two strands. Length measurements show a linear length of complex corresponding to ~3.8 Å per base using revised figures for the size and number of nucleotides in fd DNA as per Berkowitz and Day (1974). The detailed electron micrographs suggest considerable supercoiling of the DNA in the complex. Thus the actual distance between bases along the DNA strand itself must be considerably longer, perhaps as much as 6 Å between bases. The gene 32 protein-fd DNA complex leaves the circle open, but length measurements indicate a linear length of the complex of 5.3 Å per base, considerably greater than fd DNA alone (Delius et al., 1972). Again supercoiling of the DNA in the complex suggests a greater actual distance between bases along the extended DNA chain.

On the other hand the E. coli protein contracts the circumference of the circular complex with fd DNA to 2.1 Å per base (Sigal et al., 1972). While supercoiling also appears to be present in this complex, the distance between bases along the DNA stand itself is probably not as great in this complex as in the other two. Intercalation of aromatic residues between bases in single-stranded DNA could account for some of the expansion of the DNA length observed in complexes with gene 5 and gene 32 proteins, both of which seem to have tyrosyl residues which participate in DNA binding. The E. coli protein apparently does not have accessible surface tyrosyl residues and would appear not to use tyrosyl intercalation as part of the binding mechanism. These structural differences may determine how the DNA interacts with other proteins and enzymes in the cell.
The CD changes on complex formation with the synthetic ribonucleotide polymer, poly (A), are similar with all three of the proteins studied (Figures 3 and 20). The potential of these proteins to bind to cellular RNA as well as DNA may be of importance in considerations of their physiological functions in the cell.

The *in vivo* significance of the tryptic-like cleavage of gene 32 protein is not certain, but a reasonable postulate can be made. Loss of a large, highly charged fragment from the protein as well as the accompanying small conformational changes in the molecule might alter protein-protein interactions and the configuration of the DNA-protein complex recognized by enzymes such as the T4 DNA polymerase, which function in the replication and recombination of T4 DNA.

The fact that the modified gene 32 protein melts T4 DNA at physiological temperatures (Hosoda *et al.*, 1974), while the parent molecule of 35,000 M.W. does not (Alberts and Frey, 1970), suggests a physiological role for the proteolytic cleavage. A molecule participating in replication and recombination of T4 might reasonably be expected to melt the phage DNA at physiological temperatures. The failure of gene 32 protein to do so has been puzzling.

Carroll *et al.* (1975) have reported that the aggregation of their preparations of gene 32 protein decreases with storage at 4°C. Although they see no change in the gel electrophoretic properties of the protein accompanying the changes in self-association, such a change could relate to limited proteolysis of the molecule. The similarity between their published amino acid composition for the gene 32 protein and ours confirm that the same polypeptide is being studied (Carroll *et al.*, 1974) (Table I). Purification of gene 32 protein from overproducer mutants
of T4 may result in residual traces of trypsin-like proteolytic contamina-
tion not detectable by polyacrylamide gel electrophoresis in our
preparations.

A new trypsin-like activity is induced in *E. coli* at about 15
minutes after infection with T4 phage (Poglazov and Levshenko, 1974).
There is evidence that limited proteolysis of several of the structural
proteins of the mature phage plays an important part in the maturation
and assembly of T4 (Eiserling and Dickson, 1972).

The amino acid sequence of gene 5 protein (see Figure 26) reveals
an uneven distribution of charged residues along the polypeptide chain,
with the basic residues clustered near the amino and carboxyl termini
and the acidic residues located mainly in the central part of the
chain. At physiological pH gene 5 protein, like the gene 32 protein,
has a net positive charge (Alberts et al., 1972). A crude calculation
of the average degree of hydrophobicity of the gene 5 protein using the
methods of Bigelow (1967) gives a number significantly higher than that
for all the common globular proteins, with the most markedly hydrophobic
residues distributed uniformly over the sequence except for a relative
dearth over the amino terminal one-fourth of the protein.

The single sulfhydryl group of gene 5 protein seems to be centrally
involved in the maintenance of tertiary structure as well as centrally
located in the primary structure. The lack of reactivity of the -SH
group suggests that it is probably not sufficiently exposed to interact
directly with the DNA. Yet addition of two positive charges at this
site by Hg(II) complexation completely destroys the folding of the
protein. Protection of the -SH from reaction with Hg(II) by the DNA
implies that the DNA prevents access from the solvent to this residue.
It is remarkable that gene 5 protein is stable in 6 M urea and in 3 M guanidine HCl, but its native structure appears to be very sensitive to alterations in charge distribution over the molecule. The introduction of negative charges on the $\alpha$ and $\varepsilon$-amino groups by maleylation denatures the protein (Figure 14); a similar phenomenon is seen with E. coli acyl carrier protein where neutralization of the amino groups results in an electrostatically mediated disruption of the native structure (Schulz, 1975). This result, along with the sensitivity of the structure to the manipulation of the charge on the single sulfhydryl group suggests that electrostatic interactions may be as important, if not more so, than hydrogen bonding in maintaining the tertiary structure of the gene 5 protein.

Three-Dimensional Model of the Gene 5 Protein-Tetranucleotide Complex

Chou-Fasman Secondary Structure Prediction. Chou and Fasman (1974a,b) have developed a simple technique for the prediction of regions of secondary structure - $\alpha$, $\beta$, or random coil - in proteins by analyzing the primary amino acid sequence of the protein utilizing empirically derived probability rules. These authors tabulated the frequency of occurrence of each of the 20 common amino acids in the regions of $\alpha$ and $\beta$ structure in fifteen proteins whose structures have been determined to high resolution by X-ray diffraction. They used their frequency data to compute indices for the probability of occurrence of each type of amino acid in $\alpha$-helix, $\beta$-sheet, $\beta$-turn, or random coil regions.

The application of the Chou-Fasman analysis to the primary sequence of gene 5 protein (Figure 26) as determined by Nakashima et al.
Figure 26. The amino acid sequence of the gene 5 protein from bacteriophage fd as elucidated by Nakashima et al. (1974a,b). The amino acids are characterized as per the Chou and Fasman (1974a,b) schema as H (strong former), h (medium former), I (weak former), i(indifferent), b (breaker), and B (strong breaker) of helices in the first row and ß structure in the second row. Predicted secondary structural regions are enclosed in boxes. Residues 7-10, 24-27, 38-41, and 71-74 are predicted to be in ß-turns; these sequence regions are underlined. Tyrosyl residues modified by reaction with tetranitromethane are marked with asterisks ( * ).
Figure 26

\[ \text{NH}_3\text{-Met-Ile-Lys-Val-Glu-Ile-Lys-Pro-Ser-Gln-Ala-Gln-Phe-Thr-Thr-Arg-Ser-Gly-Val-Ser-Arg-Gln-} \]

\[ \begin{align*}
\alpha & \quad h & I & I & h & H & I & I & B & i & h & H & h & h & i & i & i & i & B & h & i & i & h \\
\beta & \quad H & H & b & H & B & H & b & b & b & h & I & h & h & h & h & i & b & i & h & b & i & h \\
\end{align*} \]

Gly-Lys-Pro-Tyr-Ser-Leu-Asn-Glu-Gln-Leu-Cys-Tyr-Val-Asp-Leu-Gly-Asn-Glu-Tyr-Pro-Val-Leu-Val-

\[ \begin{align*}
\alpha & \quad B & I & b & B & i & H & b & h & H & h & H & i & h & i & h & h & h & B & h & h & h & h \\
\beta & \quad i & b & b & h & b & h & b & B & h & h & h & h & H & i & h & i & b & B & h & b & H & h & h \\
\end{align*} \]

Lys-Ile-Thr-Leu-Asp-Glu-Gly-Gln-Pro-Ala-Tyr-Ala-Pro-Gly-Leu-Tyr-Thr-Val-His-Leu-Ser-Ser-Phe-

\[ \begin{align*}
\alpha & \quad I & I & i & H & i & h & B & h & B & h & B & B & B & B & h & b & i & h & h & h & i & i & h \\
\beta & \quad b & H & h & h & i & B & i & h & b & I & h & I & b & i & h & h & H & b & b & h & b & i & h \\
\end{align*} \]


\[ \begin{align*}
\alpha & \quad I & h & B & h & h & B & i & H & h & I & i & i & i & I & h & H & h & h & \quad B & H & I \\
\beta & \quad b & H & h & h & i & b & h & H & H & i & i & i & i & i & h & H & b & I & b \\
\end{align*} \]
(1974a,b) leads to striking results. Two long stretches of $\beta$-sheet structure are predicted with high probability in the regions of residues 31-50 and 61-85; a shorter region of $\alpha$-helix is indicated near the amino terminus. Three $\beta$-turns are indicated at residues 24-27, 39-42, and 71-74; the latter two fall almost exactly in the middle of the predicted $\beta$-sheet segments. The secondary structure derived from this analysis is illustrated in Figure 27. Both regions of $\beta$-structure have been folded at the predicted turns to yield two anti-parallel pleated sheets. This leaves 2 of the 3 tyrosines which are accessible to modification in the native protein (tyr-26, tyr-56) in relatively unstructured regions while the third (tyr-41) is at the end of a $\beta$-sheet. The reactive lysyl residues are shown in their positions near the ends of the sequence.

Several other nucleic acid binding proteins also appear to contain a significant proportion of $\beta$-structure. *Staphylococcus aureus* nuclease and ribonuclease S contain 25% and 44% $\beta$-structure respectively as shown by X-ray diffraction (Chou and Fasman, 1974b). It has also been hypothesized on the basis of various conformational parameters that large regions of $\beta$-structure exist in the amino-terminal operator binding region of the lac repressor protein (Chou et al., 1974; Patel, 1975) and in pancreatic deoxyribonuclease I (Cheng, 1966). A model for the interaction of polypeptides in the $\beta$-conformation with nucleic acid polymers has recently been proposed (Carter and Kraut, 1974).

**Tertiary Structure.** On the basis of the chemical modification data and physico-chemical studies reported here, the model of the secondary structure (Figure 27) has been folded into a tertiary configuration consistent with an interaction with a 5'-phosphate
Figure 27. Secondary structure of gene 5 protein as predicted from the conformational parameters for amino acids in \( \alpha \)-helical, \( \beta \)-sheet, and random coil regions as determined by the method of Chou and Fasman (1974).
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
tetra-deoxyribonucleotide. In the model pictured in Figure 28 the \( \beta \)-
structure containing residues 31-50 is central; it approximately aligns
with the axis of the tetranucleotide as accentuated by the glass rod.
Three of the tyrosyl residues of the protein are also aligned on this
axis; this "stacking" configuration is suggested by the prominent
ellipticity at 228 nm as discussed above. There is no absolute evidence
to suggest how many tyrosyls are involved in the "stack", but the
involvement of at least some of the nitratable tyrosyls which are
protected from reaction by complexation with DNA is indicated by the
change induced in the 228 nm peak upon DNA binding. In this model the
3 nitratable tyrosyls, tyr-26, tyr-41, and tyr-56, are stacked. How-
ever, it is impossible to bring tyr-26 and tyr-56, in the "coil regions",
closer than about 20 Å to tyr-41 which is in the midst of a sheet
region without disrupting the long regions of \( \beta \)-pleated sheet. It may
be that tyr-41 is involved in a protein-protein contact if the other
two accessible tyrosyl residues are involved with the DNA in the
complex. The relatively small magnitude, as well as the direction of
the up-field shifts seen in the \( { }^{19} \text{F NMR} \) spectrum of fluoro-tyrosyl
gene 5 protein upon tetranucleotide binding are indicative of the ring
current "shielding" which the fluorine nuclei on the tyrosyl rings
would experience if they were intercalated between nucleotide bases
(Millet and Raftery, 1972). Thus the three interacting tyrosyls are
shown as intercalated between nucleotide bases.

If the lysine-rich, relatively non-hydrophobic amino-terminal
region of the peptide chain is wound around to the tyrosyl region
(Figure 28, upper left), three lysyl residues (lys-3, lys-7, and lys-46)
can be manipulated to interact with three of the four phosphate groups
Figure 28. Model of tertiary structure of gene 5 protein complexed with 4 nucleotide bases constructed by folding the predicted secondary structure under constraints of spectroscopic and chemical results described in the text. The vertical glass rod goes through the "stack" of tyrosyl residues and nucleotide bases. The nucleotide shown, d(pTpA)<sub>4</sub>, has been extended in both directions to illustrate the twist that may be induced in the nucleotide backbone by the protein-protein interactions between adjacent monomers of gene 5 protein. Proceeding from the 5' end, residues 4 through 7 are bound to the monomer shown. Residues 1 through 3 would bind to the gene 5 monomer below the one shown, the axis of the nucleotide backbone (oblique glass rod) being rotated relative to the complex above. Residue 8 is again rotated and would be bound to the gene 5 monomer above the one shown.
STEREO VIEW OF GENE 5-nucleotide COMPLEX
of the 5'-phosphate tetranucleotide used. If the second β-sheet region (residues 61-84) is now brought up "behind" and perpendicular to the first β-sheet region (residues 31-50) a fourth lysyl residue (lys-69) can form a "salt-linkage" to the remaining phosphate. This folding leaves the relatively inaccessible cysteine (cys-33) "buried" behind the DNA-interacting tyr-26 in the center of the protein. The introduction of 2 positive charges on a large mercury atom amidst the concentration of basic residues already in the "DNA binding cavity" of the protein as depicted in this model could well disrupt the structure as observed. On the other hand the reversal of the lysyl charges by their reaction with maleic anhydride would concentrate an overabundance of negative charges in this central area which the model predicts already has some acidic residues (glu-5, glu-40). This folding also leaves the two solvent inaccessible tyrosyls (tyr-34 and tyr-61) in an internal position. Incidentally this leaves a concentration of basic (arginyl) residues surrounding a cavity at the "top" of the model and protrusion of acidic residues at the bottom.

The above model predicts that the DNA chain in the complex will be relatively extended since the separation of bases in the tetranucleotide is required to be over 6 Å. While direct measurements of the length of the in vivo complex from the electron micrographs show a linear length of complex corresponding to ~3.8 Å per base, the structure is clearly supercoiled. Hence the actual distance between bases along the axis of the DNA must be considerably longer, possibly as much as the 6 Å suggested by the model. The origin of the twisted structure of the overall complex is of course not indicated by the model and may derive as much from protein-protein interactions between adjacent gene 5
monomers as from the interaction of individual monomers with four bases.

Obviously innumerable alternatives to the above folding are possible. A number have been tried and found to be less satisfactory than the version presented here. For example, aligning the two \( \beta \)-pleated sheets results in a very awkwardly elongated structure with little possibility for "burying" residues found experimentally to be unreactive in the native protein. The structure of the protein and its possible mechanisms of binding to DNA as deduced from the above simple assumptions is quite different from the overlapping "fish scale" structure suggested by Marvin et al. (1974) for the binding of the largely helical coat protein to the DNA in intact filamentous phage. The speculative nature of this structure is emphasized, however, by a recent report by Schultz et al. (1974) demonstrating the tendency of the Chou and Fasman system to overpredict \( \beta \) structure.

It will be interesting to test predictions concerning reactive residues and speculations concerning possible nucleotide induced conformational changes and their relation to protein-protein interactions (Rasched and Pohl, 1974) using the (tenuous) framework generated by this model building. Preliminary results of tests of the model using NMR are presented in an appendix to this work.

Determination of the structure of these "unwinding" proteins and how they control the conformation of single-stranded DNA should reveal information about the basic mechanisms of DNA-protein interactions which are applicable in general. The role of aromatic and basic residues in the nucleic acid binding interaction as related in this work to the structure of the three protein-DNA complexes studied may be
a general feature of protein-nucleic acid interactions. The spectroscopic and chemical techniques used in the present work may be employed in the investigation of general mechanisms of the binding of the DNA unwinding proteins.

Knowledge of the molecular details of this type of interaction should be useful in attempting to understand the basis for the affinity exhibited by other, more specific DNA binding proteins such as the bacteriophage λ and lac operon repressor proteins, and the RNA polymerases.
REFERENCES


APPENDIX

A model for the binding of the gene 5 protein from bacteriophage fd to a tetradecoxyribonucleotide presented in the body of this work involves major predictions concerning the mode of basic residue and of tyrosyl interaction with components of the nucleic acid. In collaboration with Dr. I. Armitage, these predictions are being tested using NMR spectroscopy. Preliminary results of 31P and 1H NMR studies on the gene 5 protein interaction with a tetranucleotide are presented here.

Materials and Methods

1H NMR spectra were recorded on a FT-Bruker spectrometer operating at 270 MHz. D2O present in the sample was used as a field-frequency lock. Measurements were made at 25° ± 2° on 0.25 ml samples contained in 5 mm sample tubes. Chemical shift values are reported as parts per million (δ) downfield from 2,2-dimethyl-2-silapentane-5-sulphonate.

31P NMR spectra were recorded on a FT-Bruker HFX-90 spectrometer operating at 36.4 MHz. Deuterium oxide (D2O) present in the sample or in a 3 mm coaxial capillary insert was used as a field-frequency lock. All spectra were obtained under conditions of proton noise decoupling unless otherwise indicated using the Fourier transform method. A spectral width of 5000 Hz was used with an acquisition time of 0.2 s. This sweep width was chosen to maximize the S/N improvement from a 5000 Hz bandwidth crystal filter. For all spectra shown, an interpolation expansion routine was employed providing a resolution of 1.22 Hz/point. Measurements were made at 25 ± 2° using 5 mm sample tubes.
containing 0.5 ml of solution. To obtain satisfactory spectra, up to 250,000 transients were collected on individual samples. $^{31}$P chemical shifts were determined relative to external 85% H$_3$PO$_4$ and were found to be identical for both the internal and external $^2$H lock, eliminating corrections for possible changes in magnetic susceptibility.

Samples of gene 5 protein, 1-2 mM, were prepared by 5 dialyses against pure D$_2$O containing 20 mM sodium phosphate, pD 8.0. The tetranucleotide sample was prepared by repeated lyophilization from D$_2$O solution, and finally redissolved in the above buffer.

Results and Discussion

Proton magnetic resonance spectra of the gene 5 protein, the tetranucleotide, and of the complex are pictured in Figure A1. The NMR spectra of the protein (Figure A1a) is much as expected from previous $^1$H NMR spectra of small proteins (Campbell et al., 1973; Bradbury et al., 1973). The resonance from the -CH of the single histidine is farthest down-field followed by multiple aromatic (tyrosyl and phenylalanyl-CH groups) resonances. In the up-field region the resonances of the -CH$_2$ groups of lysyl and argininyl residues can be identified and are assigned in Figure A1a from the data on lysine- and arginine-rich histones (Bradbury et al., 1973). Aliphatic proton resonances dominate the farthest up-field end of this spectrum. Again tentative assignments of these are based on the histone spectra. Resonances in the tetranucleotide spectrum (Figure A2e) overlap with those of the protein but are distinguishable. The addition of tetranucleotides to the protein (Figure A2b-d) results in a marked shifting and broadening of the resonances of the aromatic protons, probably the tyrosyl protons, consistent with the $^{19}$F NMR results. While not analyzed in detail as
yet they appear to be within 1 ppm, compatible with ring current shifts. The histidine resonance (His-64) is not altered as predicted by the model (Figures 27 and 28). There is significant broadening, loss of amplitude and down-field shift of the resonances in the aliphatic region. This latter observation raises the possibility of a large change in environment of the more non-polar residues of the protein. This may reflect contact of aliphatic residues with the DNA in the binding crevice. DNA-binding may also induce conformational change in the molecule. Little change is seen in the lysyl and arginyl resonances; however, this does not rule out involvement with phosphate groups of the DNA backbone because, as shown in the model, these residues are quite exposed on the surface of the protein. They also acetylate readily in the complex. In Figure A1b-d the gradual rise of apparently free nucleotide resonances can be followed as an increasing excess of nucleotide is added.

Based on chemical modification data and sensitivities of DNA complex formation for the gene 5 protein (as well as for other DNA-binding proteins tested) it would appear that a number of lysyl residues of the protein are involved in electrostatic binding to DNA backbone phosphates. A $^{31}$P NMR spectra of a random sequence, 5'-phosphate tetranucleotide in 10 mM Tris-HCl, pH 8, is shown in Figure A2a. While the resonances in this spectrum are quite broadened, the ratio of the area of the up-field phosphodiester resonance to that of the phosphate monoester (5') is \(\approx 3:1\) confirming the previous assignment of tetranucleotide based on the column elution profile (see Materials and Methods). The broadening seen here is probably a result of both the effects of the Tris-HCl buffer (via alteration in proton exchange rates) as well
as paramagnetic metal ions bound tightly to the tetranucleotide phosphates.

Changes in the $T_1$ and $T_2$ values for the phosphorus nuclei of the nucleotides under varying conditions of buffer and metal ion contamination require caution in interpreting changes in the phosphorus resonances upon binding to the protein; however, preliminary studies show that significant changes do occur in the phosphorus resonances of the nucleotide on binding to gene 5 protein.

The phosphorus resonances of the tetranucleotide in $D_2O$ and 20 mM phosphate buffer after extraction of metals with a $CCl_4$ solution of dithizone is shown in Figure A2b. Under these conditions the resonance of the 5' phosphate is not well separated from the resonance of free phosphate and the lines are still somewhat broadened. Addition of EDTA will narrow them, but use of these samples for the 270 MHz proton spectrum made it undesirable to add EDTA. The $^{31}P$ study is best carried out separately with EDTA present and is the object of continuing investigation. Nevertheless, preliminary interpretations can be made of the changes occurring on the addition of gene 5 protein (Figure A2c).

On formation of the gene 5 protein-tetranucleotide complex (Figure A2c) there is a significant down-field shift in the resonance of the 5' phosphate of the nucleotide, from -2.12 ppm in the free nucleotide to -2.71 ppm in the complex. There is also a small up-field shift of the resonance of the 3'-5' phosphates, from +1.33 ppm in the free nucleotide to +1.63 ppm in the complex. The tetranucleotide was added to the protein as a solid after lyophilization from phosphate buffer which accounts for the increase in the resonance from free inorganic phosphate (Figure A2c).
The most striking finding is that the $T_1$ for the 3'-5' phosphates (determined by the inversion recovery method) has decreased from $\sim$3.0 sec in the free nucleotide to $\sim$1.0 sec on the bound form. Thus the chemical environment of both the 5' and the 3'-5' phosphates are altered on binding to the protein and the 3'-5' phosphates appear to be held in contact with the protein structure as suggested by the considerable enhancement of their longitudinal relaxation. This is consistent with a binding mechanism involving electrostatic interactions between protein residues (lysyl amino groups) and the tetranucleotide phosphates.

Further work must be done with both the $^1$H and $^{31}$P NMR investigation. Attempts to produce difference spectra are in progress. Spectra of nitrotyrosyl gene 5 protein should make individual assignments of tyrosyl resonances possible, since large shifts occur in these resonances on nitration (Snyder et al., 1975). Extension of both the $^1$H and $^{31}$P NMR studies to complexes with an octanucleotide should allow a determination of what effects the protein-protein interactions involved in the cooperative binding have on the structure of the protein and the nucleotide.

Bibliography
Figure A1: Fourier transform proton magnetic resonance spectra of gene 5 protein and its complex with a tetranucleotide: a) native gene 5 protein at 1-2 mM in 20 mM sodium phosphate in D$_2$O, pD 8; b-d) successive additions of 2 equivalents of deuterium exchanged tetranucleotides to give a final tetranucleotide concentration of 8 mM; e) tetranucleotides in same buffer as in a) at 8 mM. 3000-6000 scans. Tetranucleotide concentration given in terms of 4 base unit, that is [P]/4.
Figure A2: Fourier transform $^{31}$P NMR spectra a tetranucleotide and changes induced by complex formation with gene 5 protein: a) tetranucleotide, $1.3 \text{ mM}$ in $10 \text{ mM}$ Tris-HCl, pH 8; b) tetranucleotide, $8 \text{ mM}$ in $20 \text{ mM}$ sodium phosphate, pH 8; c) tetranucleotide, $1.5 \text{ mM}$, plus gene 5 protein, 1-2 mM in the phosphate buffer.