21

Single-Stranded DNA Binding Proteins

STEPHEN C. KOWALCZYKOWSKI • DAVID G. BEAR • PETER H. VON HIPPEL

THE ENZYMES, Vol. XIV

ISBN 0-12-122714-6

I. Introduction and Overview

In its broadest context, the title of this chapter could easily encompass the subject matter of many of the chapters of this and the companion volume. As we define it the topic is more circumscribed, and our discussion is restricted to proteins that (1) bind preferentially and relatively nonspecifically to single-stranded DNA, and (2) have no other (enzymatic) activity.

It is becoming apparent that proteins that fit the above definition are essential to many physiological functions, including replication, recombination, and repair, in a host of organisms ranging from bacteriophage to higher eukaryotes. Furthermore, despite the apparent simplicity of their central function, it turns out that many of these proteins have subtle and sophisticated features that not only permit them to *participate* in these processes, but also to play important roles in *controlling* and *directing* them. In this sense current single-stranded DNA binding proteins may represent systems that have evolved substantially beyond primitive precursors, which may only have been capable of direct and uncontrolled nucleic acid binding.

Therefore in this chapter we focus not only on the DNA binding properties of the proteins, but also (to the extent information permits) on molecular aspects of their involvement in entire systems of DNA replication, recombination, and repair. To this end we attempt to bring out the structural, thermodynamic, and functional principles that unify the single-stranded DNA binding proteins and serve to define them as a class.

After a brief summary of the nature and measurement of DNA-protein interactions (Section II), and a presentation of general purification strategies for single-stranded DNA binding proteins (Section III), we discuss in some detail the properties of representative members of this class (Section IV). These "case histories" deal with proteins isolated from both prokaryotic and eukaryotic cells, and although significant differences in some properties exist, we hope the reader will be more impressed by the many apparent underlying similarities of the proteins in terms of their functions, and of the structures that have evolved in support of the functions.

In Section V we review briefly certain ways in which the single-stranded DNA binding proteins have been exploited as tools in molecular biological research, particularly in the electron microscopy of biological macromolecules, and in certain biochemical assays.

In conclusion (Section VI) we attempt to draw these threads together to present a series of generalizations to help involved workers develop an overview of the field, and also to set up certain experimental criteria that

might guide and facilitate the characterization and functional interpretation of the properties of proteins that will be examined in the future. We also hope that a better understanding of the structure and properties of the single-stranded DNA binding proteins will help to provide further insight into the mechanistic details of the physiological systems of which these proteins form an integral part.

The nomenclature of the single-stranded DNA binding proteins is still in a relatively unsatisfactory state. This reflects, in part, the fact that the role(s) of these proteins in the various integrated physiological systems in which they are involved is still far from completely understood. Most names used in the past reflect the (at least potential) ability of these proteins to shift nucleic acid helix-coil conformational equilibria by binding preferentially to the single-stranded "coil" form of DNA or RNA. For this reason these ligands have been called melting proteins, unwinding proteins, helix-destabilizing proteins (HDPs), and single-strand binding (SSB) proteins. An effort initiated by Bruce Alberts (*1*) to arrive at a consensus on the use of the term "helix-destabilizing protein" as a generic name for this class has been only partially successful. In the absence of a final consensus, and to avoid further confusion, we generally use (or, at least, indicate) the names employed in the original articles in presenting detailed descriptions of individual protein systems in Section IV. In general discussions we use the terms single-stranded (DNA) binding protein (SSBP) and helix-destabilizing protein (HDP) interchangeably.

II. Theoretical and Experimental Considerations (2)

A. The DNA Substrate

Single-stranded nucleic acid sequences comprise the primary binding substrates for the proteins discussed here, and thus we consider briefly the main structural features of these target lattices. First, however, we must recall that almost all of the nucleic acid components of the cell are *double-stranded* in nature, either as double-helical DNA or as base-paired

1. B. M. Alberts and R. Sternglanz, *Nature (London)* **269,** 655 (1977).

2. We note that such binding interactions (with either single- or double-stranded nucleic acids) comprise a central component of the general interaction of virtually all genome regulatory proteins (including repressors, polymerases, nucleases, gyrases, helicases, etc.) with nucleic acid lattices. Thus it is important to note that the theoretical and experimental considerations outlined in this section in terms of HDPs are general, and form, in exactly the same terms, a part of the description of all nucleic acid binding proteins.

regions of secondary structure in the various cellular RNAs (*3*). These double-stranded structures enter our considerations in two ways: (1) They can compete directly (depending on relative binding affinities) for DNA binding proteins; and (2) they can serve (via local fluctuation-driven double-helix $\rightleftarrows$ coil transconformation reactions) as additional sources of single-stranded binding sites. The first problem is generally obviated by the fact that these proteins bind sufficiently weakly to base-paired sequences to prevent double-stranded nucleic acids from serving as effective binding competitors at physiological levels of protein and nucleic acid, and at physiological salt concentrations. The second problem is central, and has both equilibrium and kinetic aspects.

All duplex nucleic acid sequences are, of course, in potential equilibrium with their single-stranded constituents. Thus, from an equilibrium perspective, these latter forms will bind protein if the (favorable) binding free energy made available on forming such protein–single-stranded nucleic acid complexes exceeds the conformational free energy lost on disrupting the otherwise stable duplex sequences. The point at which complex formation is favored for each sequence depends on the length of the particular double-stranded segment, its structure (i.e., does it contain single-stranded loops or mispaired bases?), its base composition, the solvent environment (increasing salt concentrations generally stabilize the nucleic acid duplex and destabilize protein–nucleic acid complexes), the binding constant, and the concentration of free protein. (For a further discussion of these aspects, see Refs. *4, 5*.)

In addition to considering the equilibrium situation, one must also ask whether conformational equilibrium is actually reached in a finite time under various experimental conditions. For example, results with gene 32 protein (*4, 6, 7*) suggest that this protein is kinetically blocked from invading double-stranded DNA (and probably RNA) sequences of significant length, even under conditions where single-stranded nucleic acid binding is favored at equilibrium. Thus, for at least some DNA binding proteins, "opening" fluctuations of the DNA duplex may not be of sufficient size or

3. RNA lattices are included in this discussion because they can serve as binding sites for autogenous regulatory interactions (at least for phage T4-coded gene 32 protein) and because they comprise an appreciable fraction of the nucleic acid composition of the cell, and thus may compete with the primary single-stranded DNA target sequences for free single-stranded binding protein.

4. D. E. Jensen, R. C. Kelly, and P. H. von Hippel, *JBC* **251**, 7215 (1976).

5. J. W. Newport, N. Lonberg, S. C. Kowalczykowski, and P. H. von Hippel, *JMB* **145**, 105 (1981).

6. B. M. Alberts and L. Frey, *Nature (London)* **227**, 1313 (1970).

7. J. W. Newport, Ph.D. Thesis, University of Oregon, Eugene, Oregon, 1980.

frequency to nucleate the transition to the single-stranded–protein-complexed form. As discussed in Section IV,A, such kinetic blocks may serve as important physiological control elements. Other single-stranded DNA binding proteins can take advantage of thermally driven conformational fluctuations of the DNA double-helix to melt such structures to equilibrium. However we note that in at least some replication complexes (e.g., T4, *E. coli*) other protein components are involved in opening the replication fork and making binding sites available for the subsequent (and passive?) binding of the SSBP.

The actual interaction of a DNA binding protein with a single-stranded nucleic acid lattice clearly involves the participation of many functional groups. A substantial component of the binding free energy is generally electrostatic in nature, involving interaction of DNA phosphates with appropriately placed, positively charged (Arg, Lys, His) amino acid residues of the protein. In addition, hydrogen bonding, dipolar, and hydrophobic (solvent-driven) interactions between functional groups of the protein and those located on the various components of the nucleic acid may provide some additional stability to the binding complex, as well as varying degrees of binding specificity. (For a recent discussion of these aspects see Ref. *8*.]

B. Binding Parameters and Interactions

1. *Thermodynamics of Binding*

A description of the nonsequence-specific binding of proteins to nucleic acid lattices involves thermodynamic considerations beyond those needed to characterize the simple binding of ligands to independent binding sites. On binding to a nucleic acid lattice, a DNA binding protein generally covers (i.e., makes unavailable to another incoming protein) more than one nucleotide residue. If binding is nonspecific each nucleotide residue of the lattice can then be considered, in principle, to comprise the beginning of a potential protein-binding site n residues in length, where n (in units of nucleotide residues) represents the protein site size. When binding to an otherwise "naked" lattice the protein not only occludes the lattice site to which it actually binds, but it also partially covers $2n$ minus two other potential binding sites. As more protein is bound, however, the number of potential binding sites occluded per binding event decreases, and the number of binding sites remaining unoccupied at any particular protein

8. P. H. von Hippel, *in* "Biological Regulation and Development" (R. F. Goldberger, ed.), Vol. I. p. 279. Plenum, New York, 1979.

binding density is not a linear function of the number of ligands bound. As a consequence of this "overlap" type of binding, curved Scatchard plots are obtained even for the noncooperative binding of homogeneous protein ligands (*9, 10*).

General approaches for extracting intrinsic ligand-lattice binding constants (K), as well as site size parameters (n), from binding data on such systems have been developed elsewhere (*9–11*). We note that because of this overlap effect, complete equilibrium saturation of a nucleic acid lattice by a noncooperative binding protein cannot be attained. As the binding density of protein on the lattice increases, most of the remaining vacant binding loci are less than n residues in length, and thus cannot be used. To create more usable binding sites some of the smaller vacancies must be rearranged into fewer larger ones. The unfavorable entropy involved in such rearrangements eventually exceeds the potential free energy to be gained by binding another protein molecule, and the binding stops short of lattice saturation.

This problem can be overcome and binding saturation attained if binding is cooperative, i.e., if binding of a ligand adjacent to one previously bound is more favorable than isolated binding. Many physiologically active DNA binding proteins show such binding cooperativity, which is characterized thermodynamically by the unitless parameter, ω. This parameter corresponds to the equilibrium constant for moving a protein already bound to the lattice from an isolated to a contiguous binding position. As an indication of the magnitudes involved, $\omega \cong 2 \times 10^3$ for the cooperative binding of T4 gene 32 protein to a single-stranded DNA or RNA lattice. This value of ω represents a favorable increment in the net binding free energy (for contiguous over isolated protein binding) of about -4 kcal per mole of protein monomer.

These three parameters (n, K and ω) suffice for a complete *thermodynamic* description of protein-nucleic acid binding interactions of this type. The definitions of these parameters are summarized and illustrated in Fig. 1.

2. *Molecular Characterization of the Binding Interaction*

In addition to the "bare-bones" thermodynamic description of a protein–nucleic acid interaction provided by n, K, and ω, it is often possible to learn more about the protein binding domain, and about the com-

9. J. D. McGhee and P. H. von Hippel, *JMB* **86,** 469 (1974).

10. J. A. Schellman, *Isr. J. Chem.* **12,** 219 (1974).

11. S. C. Kowalczykowski, L. S. Paul, J. W. Newport, N. Lonberg, and P. H. von Hippel, in preparation.

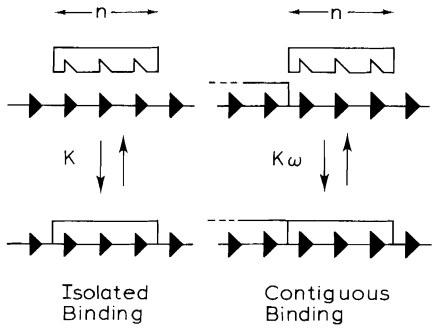


FIG. 1. Definitions of the thermodynamic parameters describing the interaction of a binding protein with a nucleic acid lattice. Each arrowhead represents a lattice site (i.e., a nucleotide residue) and the illustrated protein covers three such sites ($n = 3$). K (in M^{-1}) is the intrinsic association constant for protein binding to the lattice at an isolated site, and ω (dimensionless) represents the cooperativity of binding (ω is defined as the equilibrium constant for moving a protein from an isolated to a contiguous binding site). Thus $K\omega$ is the net binding constant per contiguously-bound protein molecule. If contiguous binding is favored, $\omega > 1$; if contiguous binding is disfavored, $\omega < 1$; and if the binding is noncooperative, $\omega = 1$.

plementary nucleic acid surface, by a variety of other approaches. Thus the number of nucleotide residues actually interacting with the protein (m), as opposed to merely being covered by it (n), can often be determined by measuring the apparent binding constants of a series of oligonucleotides (of length l) to the binding site; a particularly simple system of this type is portrayed in Fig. 2. When the length of the test oligonucleotide (l) exceeds the number of interacting residues (m), further increase in lattice length should merely increase the apparent oligonucleotide binding constant by a statistical factor, $l - m + 1$ (12). In addition, at $l = m$, the

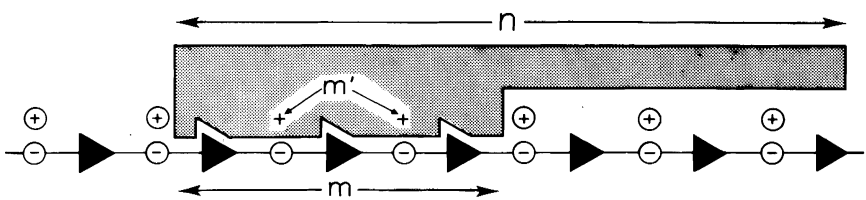


FIG. 2. Definitions of molecular binding (interaction) parameters. Here the arrowheads represent the nucleoside (sugar-base) residue, the negative charges in the backbone represent backbone phosphates, the positive charges in the protein represent basic amino acid residues, and the positive changes in the solution represent monovalent counterions. The illustrated protein here *covers* (occludes) six nucleotide residues ($n = 6$), but interacts with only three sugar-base units ($m = 3$) and forms two charge–charge interactions ($m' = 2$) with the nucleotide backbone.

apparent protein-oligonucleotide binding constant should equal the intrinsic K obtained with long DNA or polynucleotide lattices. Study of the base and sugar dependence of oligonucleotide binding constants may also provide information about binding specificities (*12–14*).

This approach depends on several crucial premises, including the assumptions (1) that there is no change in protein conformation or binding site geometry in going from the oligonucleotide binding form of the protein to the (often cooperatively bound) polynucleotide binding form, and (2) that the oligonucleotide is free to bind statistically (i.e., to shuffle) in the protein binding site. We note that neither of these assumptions appear to be valid for the T4 gene 32 protein-nucleic acid binding interaction (*14*) (see Section IV,A).

Record *et al.* (*15*) have shown that monitoring the dependence of K (or $K\omega$ for cooperatively binding proteins) on salt concentration can be developed into a molecular probe of considerable power and generality. These workers point out that (in simple cases) one can treat a protein–nucleic acid interaction as a three component system involving the protein ligand, the nucleic acid lattice, *and* the counterions that are bound tightly to this lattice. On binding the (in this sense) polycationic protein ligand, some of these counterions are displaced from the lattice; this results in a dependence of the observed binding constant on salt concentration, which can be interpreted to determine m', the number of charge–charge interactions involved (per protein monomer) in the formation of the protein–nucleic acid complex (see Fig. 2). This viewpoint also shows clearly that the major source of binding free energy for such electrostatic interactions comes from the entropy of mixing (or dilution) of the ions displaced from the polynucleotide lattice. Such determinations of m' have been made for a number of systems (*14–17*). In some situations anion as well as cation displacement may be involved, and the situation can be more complicated (*14*).

3. *Kinetics of Binding and Binding Pathways*

All of the previous considerations are essentially equilibrium in nature. In "real life" binding, equilibrium is often not attained at every step,

12. R. C. Kelly, D. E. Jensen, and P. H. von Hippel, *JBC* **251,** 7240 (1976).
13. D. M. Draper, and P. H. von Hippel, *JMB* **122,** 321 (1978).
14. S. C. Kowalczykowski, N. Lonberg, J. W. Newport, and P. H. von Hippel, *JMB* **145,** 75 (1981).
15. M. T. Record, Jr., T. M. Lohman, and P. L. deHaseth, *JMB* **107,** 145 (1976).
16. P. L. deHaseth, T. M. Lohman, and M. T. Record, *Biochemistry* **16,** 4883 (1977).
17. A. Revzin and P. H. von Hippel, *Biochemistry* **16,** 4769 (1977).

especially in complex and interlocking multicomponent systems. Thus information about association and dissociation *rates* and *pathways* is often required as well. Complete information of this type is not at hand for any DNA binding protein, though approaches have been made and some kinetic questions relevant to physiology have been raised (see Section IV,A,2) (*18–22*).

C. Methods for Monitoring Binding

1. *Measurement of Binding Parameters*

The binding parameters previously described (n, K, ω, m, m' and the related kinetic constants) can be measured in a given system by a variety of approaches. Generally titrations are involved, in which the progress of the reaction is measured by monitoring either spectroscopic or hydrodynamic changes due to complex formation. Binding site size (n) can be (and usually is) determined independently by titrating protein with nucleic acid (or vice versa) to saturation under tight-binding (typically low salt) conditions (*4*). Binding constants (K) and cooperativity parameters (ω) can be determined by titrating under conditions where binding is less tight, and appreciable (measurable) concentrations of free protein ligands and free nucleic acid binding sites are present in equilibrium with the complexes (*14, 16, 17, 23–25*). In addition, all these parameters can also be obtained by analyzing the *shapes* of binding protein-perturbed nucleic acid melting profiles (*4, 26*).

a. Nucleic Acid Signals. The interaction of DNA binding proteins with single-stranded nucleic acid sequences often results in appreciable deformation of the sugar–phosphate backbone, with concomitant unstacking of

18. B. F. Peterman, and C. W. Wu, *Biochemistry* **17,** 3889 (1978).
19. I. R. Epstein, *Biopolymers* **18,** 2037 (1979).
20. S. C. Kowalczykowski, N. Lonberg, J. W. Newport, L. S. Paul, and P. H. von Hippel *BJ* **32,** 403 (1980).
21. T. M. Lohman, *BJ* **32,** 458 (1980).
22. B. M. Alberts, J. Barry, P. Bedinger, R. L. Burke, U. Hibner, C. C. Liu, and R. Sheridan, *in* "Mechanistic Studies of DNA Replication and Genetic Recombination" (B. Alberts and C. F. Fox, eds.), ICN-UCLA Symp. Mol. Cellular Biol., Vol. 19. Academic Press, New York, 1980.
23. D. E. Jensen and P. H. von Hippel, *JBC* **251,** 7198 (1976).
24. D. E. Draper and P. H. von Hippel, *Biochemistry* **18,** 753 (1979).
25. T. L. Lohman, C. G. Wensley, J. Cina, R. R. Burgess, and M. T. Record, Jr. *Biochemistry* **19,** 3516 (1980).
26. J. M. McGhee, *Biopolymers* **15,** 1345 (1976).

adjacent nucleotide bases. As a result these processes can frequently be followed by monitoring changes in the spectroscopic properties of the nucleic acid; circular dichroism [backbone deformation; see, e.g., Refs. (*4, 27–29*)] or UV hyperchroism [base unstacking; see, e.g., Refs. (*4, 6*)] are generally used.

In addition (covalently) modified nucleic acids are often useful. For example, the chemical modification of polyriboadenylic acid to polyriboethenoadenylic acid makes this moiety fluorescent (*30*), and the fluorescence is greatly enhanced by the base-unstacking brought about, for example, by T4 gene 32 protein (*14, 31*). Nitroxide spin labels have also been attached to polynucleotides, and changes in the resulting ESR spectra on protein binding have been followed (*32*).

b. Protein Signals. Many single-stranded DNA binding proteins show appreciable quenching of intrinsic protein fluorescence on interacting with nucleic acid lattices. These changes (generally in tryptophan, but sometimes in tyrosine fluorescence) can also be monitored to follow protein-nucleic acid binding reactions [see, e.g., Refs. (12, 33–36)].

c. Other Approaches. Changes in sedimentation, electrophoretic, and gel exclusion chromatographic behavior are also used to follow protein-nucleic acid interactions [see, e.g., Refs. (*4, 23–25, 37*)]. DNA cellulose chromatography (*38*), utilizing the nucleic acid lattice as the stationary phase and the protein ligand as the mobile phase, has also been applied to the measurement of DNA-protein binding constants (*14, 16, 39*). Quantitative photoaffinity cross-linking studies, in which the competition for a DNA binding protein between a photoaffinity-labeled nucleic acid component and a nonlabeled polynucleotide is monitored, also can be made to yield measurements of binding parameters (*40*).

27. R. A. Anderson and J. E. Coleman, *Biochemistry* **14,** 5485 (1975).
28. A. P. Butler, A. Revzin, and P. H. von Hippel, *Biochemistry* **16,** 4757, (1977).
29. L. A. Day, *Biochemistry* **12,** 5329 (1973).
30. J. A. Secrist, R. J. Bario, N. J. Leonard, and G. Weber, *Biochemistry* **11,** 3499 (1972).
31. J. J. Toulmé and C. Hélène, *BBA* **606,** 95 (1980).
32. A. M. Bobst and Y.-C. Pan, *BBRC* **67,** 562 (1975).
33. R. C. Kelly and P. H. von Hippel, *JBC* **251,** 7229 (1976).
34. C. Hélène, F. Toulmé, M. Charlier, and M. Yaniv, *BBRC* **71,** 91 (1976).
35. H. T. Pretorius, M. Klein, and L. A. Day, *JBC* **250,** 9262 (1975).
36. I. J. Molineux, A. Pauli, and M. L. Gefter, *Nucleic Acids Res.* **2,** 1821 (1975).
37. K. R. Yamamoto and B. M. Alberts, *JBC* **249,** 7076 (1974).
38. B. M. Alberts, F. J. Amodio, M. Jenkins, E. D. Gutmann, and R. L. Ferris, *CSHSQB* **33,** 289 (1968).
39. P. L. deHaseth, T. M. Lohman, R. R. Burgess, M. T. Record, Jr. *Biochemistry* **17,** 1612 (1978).
40. S. C. Kowalczykowski and L. S. Paul, unpublished results.

III. Protein Isolation and Purification: Procedures and Strategies

Techniques used to isolate and purify single-stranded DNA binding proteins generally share many common features. In particular, one or more DNA-cellulose columns, first developed by Alberts and co-workers (*38*) as a form of affinity chromatography for DNA binding proteins, occupy a central position in nearly every purification scheme. Here we outline some general procedures and strategies that have been used; these procedures are summarized (in outline) in Table I where we present a sample scheme (generally the most recent or most widely used variant if several procedures are in common use) for the isolation of four of the best-characterized prokaryote single-stranded DNA binding proteins. Our purpose in this section is to pinpoint certain general approaches that have been widely employed to purify SSBPs and should probably be considered in developing procedures for the isolation of new members of this protein class.

TABLE I

PURIFICATION OF PROKARYOTIC SINGLE-STRANDED DNA BINDING PROTEINS

Protein	Procedure	Yield[a] (mg)	References[b]
T4 Gene 32 protein	Lysis (sonication and blending), DNase, LSC[c], D, ssDNA-cellulose, norleucine-Sepharose, phosphocellulose or phenyl-Sepharose	~16 ~100–250[d]	(*56*)
fd Gene 5 protein	Lysis (sonication), DNAse, LSC[c], HSC[c], ssDNA-cellulose, DEAE-cellulose	~60	(*41*)
E. coli SSB protein	Lysis (sonication), LSC,[c] PEG[c] ppt. of DNA, LSC[c], D[c], ssDNA-cellulose, DEAE-Sephadex	~3 ~30[d]	(*42*)
T7 DNA binding protein	Lysis (Sonication), LSC[c], PEG[c] ppt. of DNA, LSC[c], D[c], ssDNA-cellulose, Mg^{2+} ppt., DEAE-cellulose	~3	(*161*)

[a] Yield from 100 gm (net weight) of *E. coli* or phage-infected *E. coli*.

[b] Only the most recent, or most detailed, reference is listed.

[c] Abbreviations: LSC, low-speed centrifugation to remove cellular debris; HSC, high-speed centrifugation to remove ribosomes; PEG, polyethyleneglycol; D, dialysis into low ionic strength buffer + EDTA.

[d] Yield from overproducing strain of *E. coli* or phage.

A. Production of Cellular Extracts

Because most of the proteins discussed in detail in this chapter are derived from uninfected or phage-infected bacterial cells, we confine the majority of our remarks in this section to the production of cellular extracts from prokaryotic cells. Specific references should be consulted for procedures involved in the growth and lysis of eukaryotic cells.

1. *Strains*

The choice of a strain for the isolation of bacterial and phage-coded, single-stranded DNA binding proteins depends on the nature of the experiments to be undertaken. If the amount of protein required is small, isolation from wild-type strains is often adequate. For example, ~8 mg of T4 gene 32 protein can be isolated from 50 g of *E. coli* cells infected with wild-type T4 phage (*6*), and ~30 mg of fd gene 5 protein can be obtained from 50 g of Ff-infected *E. coli* (*41*). However, in some cases the yield of protein from wild-type strains is rather poor, or at least insufficient for large-scale physical studies; for example, only ~0.5 to 1.5 mg of SSB protein are obtained per 50 g of wild-type *E. coli* (*42*).

Thus it is often useful to isolate strains of bacteria or phage that will overproduce the protein desired. Sometimes this can be done in a fairly straightforward fashion; thus strains of *E. coli* that overproduce SSB protein were obtained by simply inserting plasmids or λ phage carrying the *ssb* gene into *E. coli* (*42*). Overproducers of *E. coli lac* repressor were found by screening for "up" promoter mutants in the *lac* i-gene (repressor) promoter (*43*).

Sometimes, however, straightforward approaches may not work, either because the overproduced protein is lethal to the cell or because the wild-type free protein level is controlled by autoregulatory feedback mechanisms. In the latter case an appropriate overproducer may be designed by taking advantage of the operation of the regulatory system. For example, the free concentration of T4 gene 32 protein is autoregulated at the translational level (*44, 45*). The system works on the basis of a cascade of binding specificities, as follows (see Section IV,A). First, sufficient protein is produced to saturate all the single-stranded DNA sequences in the cell. Then, after the free concentration of gene 32 protein has risen to a

41. B. M. Alberts, L. Frey, and H. Delius, *JMB* **68**, 139–152 (1972).
42. J. W. Chase, R. F. Whittier, J. Auerbach, A. Sancar, and W. D. Rupp, *Nucleic Acids Res.* **8**, 3215–3227 (1980).
43. B. Muller-Hill, L. Crapo, and W. Gilbert, *PNAS* **59**, 1259 (1968).
44. H. M. Krish, A. Bolle, and R. H. Epstein, *JMB* **88**, 89 (1974).
45. L. Gold, P. Z. O'Farrell, and M. Russel, *JBC* **251**, 7251 (1976).

critical level, the protein binds specifically and reversibly to a critical control sequence on gene 32 mRNA and prevents further synthesis (*5, 46*). The key to overproducing this protein, then, was to find a way to keep the *free* protein level below the shut-off level while the *total* protein concentration was increased. This was achieved for gene 32 protein by infecting with a T4 phage that had mutations in several nucleic acid processing enzymes. The net effect of these mutations is to increase greatly the steady-state concentration of single-stranded sequences in the newly synthesized phage DNA. These regions then bind greatly increased quantities of gene 32 protein, and thus permit overproduction by delaying shut-off synthesis (*45*).

In general, before time and effort is put into cloning a gene for the purpose of overproducing a specific protein, it is important to establish whether the protein may be toxic in excess or whether its synthesis is autoregulated, in order that an appropriate overproducing strategy can be devised.

2. *Cell Lysis and Processing of the Protein Extract*

Procedures commonly used for bacterial cell lysis, such as grinding with glass beads or alumina, pressure disruption, sonication, and lysozyme-detergent treatment, have been used in isolating single-stranded DNA binding proteins. Pressure disruption or lysozyme-detergent treatment are generally the methods of choice for large-scale preparations. A problem that is sometimes encountered with bacteria (particularly *B. subtilis*), and is very common with cells of higher organisms, is that of intracellular proteases. Thus, after the cell is broken, normally compartmentalized or membrane-bound proteases may start to attack the desired protein product. This not only reduces yields considerably, but can also generate proteolytic fragments that are difficult to separate from the intact protein. Both general covalently binding protease inhibitors, such as phenylmethylsulfonyl fluoride (PMSF) and diisopropyl fluorophosphate (DFP), and specific complex-forming moieties, such as soybean trypsin or chymotrypsin inhibitors, have been employed. An excellent review of this subject is available (*47*).

Other types of covalent modification of protein during isolation are a potential problem. For example, in several eukaryotic proteins the binding activity depends on the level of protein phosphorylation (see Section IV,E), and these proteins may be inactivated by phosphatases acting in the cell extract during purification.

46. G. Lemaire, L. Gold, and M. Yarus, *JMB* **126,** 73 (1978).
47. J. R. Pringle, *Methods Cell Biol.* **12,** 149–184 (1975).

After the cells have been lysed, free DNA may make the lysate enormously viscous. This viscosity must be reduced to permit effective removal of cell debris by centrifugation prior to fractionation of the extract. Sonication and treatment with DNase I are the two techniques most frequently used to degrade the free DNA. Preparations of DNase I should be treated with PMSF prior to use to remove contaminating proteases (*48*).

The concentration of salt present in lysing and extraction buffers during protein purification is also often crucial. High salt concentrations are generally used in early steps to liberate the desired protein from single- or double-stranded DNA fragments. The (protein-free) DNA is then removed by procedures such as PEG-dextran two-phase extraction, or precipitation with streptomycin or polyethyleneimine.

After initial centrifugation to remove cell debris, various approaches can be employed to achieve a gross fractionation of the cell extract. These techniques may include further centrifugation (to remove ribosomes), heat treatment, ammonium sulfate precipitation, and dextran sulfate or PEG extraction. The final extract is generally dialyzed against low salt buffer containing EDTA (to inhibit nucleases) in preparation for DNA-affinity chromatography.

B. DNA-Affinity Chromatography

At the heart of almost every scheme for purifying a DNA binding protein lies one or more DNA-cellulose or DNA-agarose column chromatography steps. The columns carry either single- or double-stranded DNA, and the protein extract to be resolved is generally loaded on the columns at low salt and then eluted with a continuous or a step salt gradient. This approach offers a powerful means to isolate and separate the DNA binding proteins of the cell.

On the other hand, the simple criterion that a particular protein binds to a DNA affinity column at low salt concentrations has often been used to identify a protein isolated from an otherwise uncharacterized extract as a binding protein. This may be a mistake; at low salt concentrations, in particular, a DNA column can function as a nonspecific cation-exchanger and many proteins stick only because of nonspecific charge-charge interactions. In functional terms proteins isolated this way may have absolutely nothing to do with DNA metabolism or genome function. It is therefore essential to prove by genetic complementation, or by some other biochemical assay, that a particular DNA binding protein does, in fact, play a role in genome manipulation *in vivo*. In contrast it is also possible for proteins that do have an actual DNA-binding function *in vivo* to be

48. P. A. Price, T. Y. Liu, W. H. Stein, and S. Moore, *JBC* **244**, 917 (1969).

lost, damaged, or modified during cell fractionation procedures, and thus fail to bind to DNA-affinity columns.

A general procedure for the identification and selective purification of biologically significant DNA binding proteins has been proposed (*49*). In this procedure the cell extract is first passed through a native DNA-cellulose column under defined conditions, and is then loaded onto a column containing denatured DNA. After a washing step, a solution of dextran sulfate is applied to the column. This step should remove proteins that stick to the column only because of weak nonspecific ionic interactions. The remaining proteins are then eluted with high salt. While this protocol may not fractionate all single-stranded DNA binding proteins, any proteins that survive the double-stranded DNA cellulose column and dextran sulfate cuts are reasonable candidates for consideration as biologically relevant DNA binding proteins.

The support matrix, and the method used to couple DNA to it, are extremely important considerations in DNA affinity chromatography. Initially, noncovalently linked DNA-cellulose columns (*50*) and DNA-agarose columns (*51*) were prepared by mixing DNA with such support materials and then drying to attach the DNA. Under certain circumstances such preparations are quite suitable for quantitative protein fractionation. However, such columns often shed DNA, and thus contaminate protein fractions with large quantities of nucleic acid that may be difficult to remove. In addition, the useful lifetime of affinity columns prepared in this manner is often short. These problems have been circumvented by covalently coupling the DNA to the cellulose matrix by ultraviolet photocrosslinking (*52*) and to agarose by cyanogen bromide activation (*53*). The relative merits of agarose and cellulose must be decided in each case. Nonspecific adsorption of the protein to the matrix can effect the salt concentration required to elute a DNA binding protein from the column; thus elution patterns for DNA-agarose and DNA-cellulose are not always the same. The use of chelating agents to inhibit divalent-cation-dependent nucleases increases the lifetime of the DNA affinity columns significantly.

C. Additional Purification Procedures

Although DNA affinity chromatography plays the most significant role in the purification of DNA binding proteins, most proteins are still some-

49. G. Herrick and B. M. Alberts, *JBC* **251,** 2124 (1976).
50. B. M. Alberts and G. Herrick, "Methods in Enzymology," Vol. 21D, p. 198, 1971.
51. H. Schaller, C. Nüsslein, F. J. Bönhoeffer, C. Kurz, and I. Nietzschmann, *EJB* **26,**
52. R. M. Litman, *JBC* **243,** 6222–6233 (1968).
53. D. J. Arndt-Jovin, T. M. Jovin, W. Bähr, A. Fischauf, and M. Marquardt, *EJB* **54,** 411 (1975).

what contaminated after this step. Further purification by ion exchange, gel filtration, and hydrophobic or affinity chromatography (see Table I for examples) is usually required. Affinity chromatography on columns containing conjugated dyes (*54*) or nucleotides (*55*) has become increasingly popular. Hydrophobic chromatography has also been employed to remove nuclease contaminants from T4 gene 32 protein (*56*).

D. Assays and Criteria of Purity

Since DNA binding proteins are not enzymatically active, they are difficult to assay in crude extracts. Occasionally an empirical criterion, such as the retention of *E. coli* SSB protein on nitrocellulose filters at very high salt concentrations, can be used (*57*). And sometimes activation of DNA polymerases or nucleases by specific DNA binding proteins can serve as the basis for a biochemical (complementation) assay. However such approaches tend to suffer from considerable variability in crude cell extracts.

Generally the purity of DNA binding proteins is established by demonstration of a homogeneous band (or set of bands) on an SDS-polyacrylamide gel, and by the absence of contaminating nucleic acids and enzymatic activities. Many of the tryptophan-containing proteins described in Section IV show 280 : 260 nm absorbance ratios as high as 1.5 to 2.0; lower 280 : 260 ratios generally reflect nucleic acid contamination.

In addition to testing for chemical purity as above, it is also important to demonstrate that the purified protein retains biological activity, if such a property has been established. Thus the operation of the pure protein as a specific activator of an *in vitro* DNA replication, recombination, or repair system can sometimes be monitored. Such assays may reveal the need for a specific cofactor, a special state of aggregation of the protein, or a covalent modification of the protein that is required for function.

IV. Structure, Properties, and Nucleic Acid Binding Interactions of Several Single-Stranded DNA Binding Proteins

In this section we describe several single-stranded DNA binding proteins and their interactions with various types of nucleic acid "substrates," as well as with other proteins of the relevant DNA replication

54. R. R. Meyer, J. Glassberg, J. Y. Scott, and A. Kornberg, *JBC* **255**, 2897–2901 (1980).
55. E. Calva and R. R. Burgess, *JBC* **255**, 11017 (1980).
56. M. Bittner, R. L. Burke, and B. M. Alberts, *JBC* **254**, 9565 (1979).
57. R. F. Whittier and J. W. Chase, *Anal. Biochem.* **106**, 99 (1980).

complexes. (Very little information is available at the molecular level about the participation of any of these proteins in integrated recombination or repair systems.) Our knowledge of each of these systems is incomplete, but to a certain extent the available information is complementary, and the reader is urged to take an integrated view. To aid in this we present an extensive summary of the known properties and interactions of several of the major single-stranded DNA binding proteins isolated from prokaryotes in Table II; a similar (though much more limited) compilation is presented for certain eukaryotic SSBPs in Table III.

The following considerations suggest how one might profitably consider the results presented here in a broader context. T4 gene 32 protein is by far the most carefully studied protein from a physiochemical point of view. Many features of other binding proteins are, at least qualitatively and semiquantitatively, similar to those of gene 32 protein. Thus the approach to (and results from) the measurement of thermodynamic and molecular parameters (e.g., n, K, ω, m, m') for this system certainly provides methodology (and probably also reflects orders of magnitude) that may well apply to similar systems. The only SSBP for which a detailed molecular structure is in hand is fd gene 5 protein. Careful analysis of this structure certainly suggests ways that other SSBPs might be built, and ways they could interact with their nucleic acid substrates. The interactions of the *E. coli* SSB protein with other components of reconstituted replication and recombination systems have been extensively studied; consideration of the enormous "skein" of interactions that involve this protein in so many facets of genome regulation and expression in *E. coli* certainly provide insight into what one might expect from similar proteins in less well-studied systems.

At the same time these are all individual proteins, and will clearly be found to have many features that are less general. Thus, fd gene 5 protein serves a special role in the life cycle of this unusual phage, and its molecular structure and interactions with DNA may well reflect some elements of this. Similarly, while gene 32 protein binds to DNA lattices as a monomer, gene 5 protein probably binds as a dimer and *E. coli* SSB protein binds as a tetramer. These binding differences require somewhat different interpretations of thermodynamic parameters such as site size (n) and cooperativity (ω).

A. Bacteriophage T4-Coded Gene 32 Protein

The product of bacteriophage T4-coded gene 32 was the first DNA binding protein to be isolated utilizing the DNA-cellulose chromatography method (*38*). Identification of the protein was achieved by infecting with

TABLE II

PROPERTIES OF PROKARYOTIC SINGLE-STRANDED DNA BINDING PROTEINS

Property	T4 Gene 32 protein	*E. coli* SSB protein	fd Gene 5 protein	T7 DNA binding protein
Physical				
Molecular weight (monomer) (kdal)	34	19.5	9.7	31 (25)
Oligomeric state	Monomer[b] ⇄ dimer ⇄ infinite	Tetramer	Monomer ⇄ dimer[b]	Monomer
Sedimentation coefficient	N.D.	4.7 S	1.35 ⇄ 1.95	N.D.
Number of amino acid residues	301	~190	87	N.D.
Extinction coefficient	37,000 M^{-1} cm^{-1} (280 nm)	30,000 M^{-1} cm^{-1} (280 nm)	7100 M^{-1} cm^{-1} (276 nm)	N.D.
Isoelectric point	5.5	6.0	8.0	Acidic
Binding				
Polynucleotide site size (n)	7 nuc/monomer	8 nuc/monomer	4 nuc/monomer	N.D.
Relative affinity	DNA > RNA	DNA > RNA	DNA > RNA	N.D.
Cooperativity	Yes	Yes	Yes	N.D.
Ability to melt duplex DNA	No (GP32*I, yes)	Yes	Yes	N.D.
Ability to reanneal denatured DNA	Yes	Yes[c]	No	N.D.
Ability to melt poly(dA-dT)	Yes	Yes	Yes	Yes
[NaCl] for elution from ssDNA-cellulose[a]	>0.6 M, <2.0 M	1.0–2.0 M	0.6 M	0.4–0.6 M
Biological				
Copies per cell	10,000 (monomer)	300–800 (tetramer)	75,000 (dimer)	N.D.
Stimulation of DNA polymerase	T4 pol	pol II, pol III, T7 pol	pol II[d]	T7 pol
Stimulation of DNA repair	Yes	Yes	N.D.[e]	N.D.
Stimulation of DNA recombination	Yes	Yes	N.D.	N.D.

[a] Because of cooperativity this is a function of protein concentration.
[b] Probable nucleic acid binding form.
[c] Requires presence of polyamines.
[d] Stimulates amount of DNA synthesis, not rate, only at low protein concentration.
[e] N.D., not determined.

TABLE III

PROPERTIES OF EUKARYOTIC SINGLE-STRANDED DNA BINDING PROTEINS

Property	Calf thymus HDP-I	Mouse myeloma	Ustilago maydis	Lily	Adenovirus
Physical					
Molecular weight (monomer) (kdal)	24	27	20	35	72
Isoelectric point	7.8	6.6	N.D.	N.D.	7.5[a]
Oligomeric state	Monomer	Monomer	Monomer	Monomer	Monomer
Sedimentation coefficient	N.D.[c]	N.D.	2.6 S	3.3 S	3.3 S
Binding					
Polynucleotide site size (n)	7 nuc	5–7 nuc	3–7 nuc	N.D.	7 nuc
Relative affinity (DNA vs. RNA)	DNA > RNA	DNA > RNA	N.D.	DNA > RNA	DNA > RNA
Cooperativity	No	No	[b]	[b]	[b]
Ability to melt duplex DNA	Yes	N.D.	N.D.	Yes	N.D.
Ability to melt poly(dA-dT)	Yes	Yes	Yes	N.D.	No
[NaCl] for elution from DNA-cellulose	0.4 M	0.4 M	1.6 M	N.D.	1.0 M
Biological					
Copies per cell	800,000	1,000,000	300,000	N.D.	70,000,000
Stimulation of DNA polymerase	Yes	Yes	Yes	N.D.	Yes

[a] Nonphosphorylated species.
[b] Possibly cooperative, but not fully established.
[c] N.D. = not determined.

phage that contained *amber* mutations in gene 32 and by demonstrating the loss of the corresponding polypeptide from the resulting polyacrylamide gel patterns; controls using phage defective in other genes (*30, 41–45*) had no effect on the band identified as the product of gene 32.

Both because gene 32 protein had been shown by genetic analysis to be essential to phage DNA replication (*57a*) and recombination (*58, 59*), and because it is produced in large (stoichiometric) quantities (*60, 61*) [~10,000 copies/cell (*62*)], Alberts *et al.* (*6, 38*) undertook biochemical studies of this protein. These early studies showed that (1) the protein binds to single-stranded DNA in preference to double-stranded DNA; (2) binding appeared to be cooperative in protein concentration; (3) gene 32 protein can denature duplex poly(dA-dT), but not native duplex DNA; (4) the protein accelerates the renaturation of denatured DNA; and (5) the protein appears to stimulate specifically the activity of T4 DNA polymerase. Each of these properties of gene 32 protein has since been studied in more detail, and the results are summarized below.

1. *Physical Properties of the Protein*

Gene 32 protein has a monomeric molecular weight of approximately 35,000, as determined by SDS gel electrophoresis and by a combination of sedimentation and gel filtration studies; it behaves hydrodynamically as if it were a prolate ellipsoid with an axial ratio of 4 : 1 (≅120 Å in length) (*6, 38*). A determination of the complete amino acid sequence of the protein yielded a more exact molecular weight of 33,466 (*63*).

a. Aggregation State. Although gene 32 protein exists predominantly as a monomer in dilute solution (<0.025 mg/ml), the protein has been shown to undergo extensive self-association at higher concentrations (*64, 65*). The extent of this aggregation is dependent on many variables, including ionic strength, pH, temperature, and glycerol concentration. The aggregation process seems to consist of at least two different types of

57a. R. H. Epstein, A. Bolle, C. M. Steinberg, E. Kellenberger, E. Boy de la Tour, R. Chevalley, R. S. Edgar, M. Susman, G. H. Denhardt, and A. Lielausis, *CSHSQB* **28,** 375 (1963).

58. J. Tomizawa, N. Anraku, and Y. Iwama, *JMB* **21,** 127 (1966).

59. A. W. Kozinski and Z. Z. Felgenhauer, *J. Virol.* **1,** 1193 (1967).

60. D. P. Snustad, *Virology* **35,** 550 (1968).

61. N. K. Sinha and D. P. Snustad *JMB* **62,** 167 (1971).

62. B. M. Alberts, *FP* **29,** 1154 (1970).

63. K. R. Williams, M. B. LoPresti, M. Setoguchi, and W. H. Konigsberg, *PNAS* **77,** 4614 (1980).

64. R. B. Carroll, K. E. Neet, and D. A. Goldthwait, *PNAS* **69,** 2741 (1972).

65. R. B. Carroll, K. E. Neet, and D. A. Goldthwait, *JMB* **91,** 275 (1975).

association. One leads to a unique dimeric species; the other, via an indefinite association process, leads to large aggregates. The indefinite aggregation is particularly susceptible to various agents, being virtually completely inhibited by high concentrations of salt (>1.0 *M* KCl), high temperature, high pH (>pH 10), and the addition of saturating concentrations of oligonucleotides [e.g., $d(pT)_{10}$]. Under each of these conditions the limit aggregate is a dimeric species, even at protein concentrations as high as 1 mg/ml.

Although these protein-protein association reactions might be considered to reflect the interactions responsible for the cooperative binding of gene 32 protein to nucleic acid lattices, quantitative study of the nucleic acid binding reaction shows that this interpretation is not consistent with the facts. For example, the unitary free energy of the aggregation process is about −9 kcal/mol, whereas the free energy of the monomer-monomer interaction involved in cooperative binding is only about −4 kcal/mol under the same conditions (*12*). In addition, while the indefinite self-association shows a very large dependence on salt concentration, the cooperativity parameter for nucleic acid binding is not dependent on this variable (*5*). Thus the nonelectrostatic part of the protein-protein self-association free energy may contribute to the cooperativity of nucleic acid binding, but it is clear that the interactions involved in self-association are not, *in toto,* the same as those involved in cooperative binding of this protein on a nucleic acid lattice.

b. Protein Domain Structure. The observation that gene 32 protein can bind to (polyanionic) DNA-cellulose, despite the fact that it carries a net negative charge at pH 8 and is eluted from DNA-cellulose at high concentrations of salt, suggested very early that gene 32 protein must contain a positively charged binding site (*6, 38*). The idea that gene 32 protein might be composed of several functionally distinct domains was given a firmer basis when it was discovered that the native protein can be converted into three discrete products by limited tryptic digestion. These products have been called G32P*I, G32P*II and G32P*III (*27, 66, 67*). The G32P*I fragment results from the removal of ~50 amino acid residues from the carboxy terminus of the native protein, the G32P*II product is obtained by removing ~20 amino acid residues from the amino terminus, and the G32P*III fragment is produced by the removal of both the amino and carboxy termini. The G32P*III protein core is quite resistant to further proteolytic degradation.

66. J. Hosoda, B. Takas, and C. Brock, *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **47,** 338 (1974).

67. H. Moise and J. Hosoda, *Nature (London)* **259,** 455 (1976).

These proteolytic products differ in their elution properties from DNA-cellulose columns. Thus the native protein and G32P*I (at fairly high protein concentrations approaching DNA site saturation within the protein band) require ~2.0 *M* NaCl for elution from single-stranded DNA-cellulose, whereas G32P*II and G32P*III can be eluted at ~0.4 *M* NaCl concentrations. This latter salt concentration is also sufficient to elute noncooperatively bound native gene 32 protein (*38*), in keeping with the conclusion (see Section IV,A,2,c) that G32P*II and G32P*III retain most of their intrinsic affinity for single-stranded DNA, but no longer can bind cooperatively. In contrast G32P*I (and, to some extent, G32P*III) bind more tightly to double-stranded DNA-cellulose than the native protein or G32P*II (*67*), presumably reflecting the fact that the *I (and the *III) species can denature the duplex DNA and bind to the resulting single-stranded sequences (*66, 68*) (see also Section IV,A,2,c).

Additional evidence for the differential involvement of the amino and the carboxy-terminal domains of gene 32 protein in nucleic acid binding was provided when it was shown that the rate at which these regions are removed by proteolytic cleavage depends on whether or not the protein is bound to polynucleotides (*69, 70*). The cooperative binding of gene 32 protein to polynucleotides was found to enhance the rate of cleavage at the carboxy terminus, but to protect the amino terminus against proteolytic attack. These effects were not observed in the presence of short oligonucleotides [$<d(pT)_8$]. These studies reinforced the suggestion that the amino terminus is a part of a distinct functional domain involved in the cooperative binding of the protein to DNA, while the carboxy terminus is exposed on binding to DNA.

To more closely define the regions within the carboxy and amino terminal domains responsible for the control of these activities, proteolytic products were also prepared using *Staphylococcus aureus* protease. Cleavage with this enzyme yielded products lacking only 9 amino acid residues at the amino-terminal end, and only ~25 residues at the carboxy-terminal end (*71–73*). These products were shown to behave essentially identically to the tryptically modified protein with respect to elution from both single and double-stranded DNA-celluloses; in addition

68. J. Greve, M. F. Maestre, H. Moise, and J. Hosoda, *Biochemistry* **17**, 893 (1978).
69. K. R. Williams and W. H. Konigsberg, *JBC* **253**, 2463 (1978).
70. J. Hosoda and H. Moise, *JBC* **253**, 7547 (1978).
71. J. Hosoda and H. Moise, personal communication.
72. J. Hosoda, R. L. Burke, H. Moise, I. Kubota, and A. Tsugita, *in* "Mechanistic Studies of DNA Replication and Genetic Recombination" (B. M. Alberts and C. F. Fox, eds.), ICN-UCLA Symp. Mol. Cellular Biol., Vol. 19 Academic Press, New York, 1980.
73. K. R. Williams and W. H. Konigsberg, *in* "Gene Amplification and Analysis," (J. C. Chirikjian, ed.) Vol. II, Elsevier/North Holland, Amsterdam, 1981.

the $^*II_{staph}$ and $^*II_{staph}$ proteins do not self-associate appreciably. Thus it appears that the residues controlling both cooperative DNA binding (*74*) and protein self-association must be located within 9 amino acids of the N terminus, and that the part of the polypeptide chain responsible for control of DNA melting ability (and interaction with T4 DNA polymerase; see below) is included within the first 25 C-terminal residues.

The amino acid sequence data for gene 32 protein can provide additional insight into the domain structure of this protein. Using an empirical secondary structure prediction scheme (*75*), the secondary structure for this molecule has been determined (*76*). The results suggest that the molecule can be divided into three regions on this basis: The amino terminal (residues 1–35) and the carboxy terminal (residues 187–301) sequences appear to be primarily α-helical, and the middle (core) sequence is predicted to be primarily β-sheet. Whether these predictions reflect reality will eventually be established crystallographically; meanwhile, we find that these numbers are in reasonable accord with circular dichroic results for the whole protein. Thus CD data (*27, 77*) suggest that gene 32 protein contains ~20% α-helix, 20–25% β-sheet, and 55–60% disordered regions, whereas the secondary structure prediction approach gives totals of 36, 18, and 46% in each of these categories.

c. Calorimetry. Calorimetric studies on the denaturation of gene 32 protein and its proteolytic products have shown that removal of either the N- or the C-terminal peptide decreases the thermal stability of the protein (*73, 78*). Under comparable conditions these proteins denature at about 56° (gene 32 protein), 54° (G32P*II), 51° (G32P*I), and 46° (G32P*III). Each protein is further stabilized by single-stranded polynucleotide binding. The thermal denaturation of protein–nucleic acid complexes of gene 32 protein and G32P*I takes place over a much narrower temperature range than does the denaturation of either the free proteins or complexes of either G32P*II or G32P*III with nucleic acid lattices. This suggests that native gene 32 protein and G32P*I bind cooperatively, whereas G32P*II and G32P*III do not.

2. *Protein–Nucleic Acid Interactions*

As previously pointed out, several physicochemical features of the binding of gene 32 protein to nucleic acid substrates were recognized in

74. N. Lonberg, S. C. Kowalczykowski, L. S. Paul, and P. H. von Hippel, *JMB* **145**, 123 (1981).
75. P. Chou and G. Fasman, *Advan. Enzymol.* **47**, 45 (1978).
76. K. R. Williams, M. LoPresti, and M. Setoguchi, *JBC* **256**, 1754 (1981).
77. J. Greve, M. F. Maestre, H. Moise, and J. Hosoda, *Biochemistry* **17**, 887 (1978).
78. K. R. Williams, L. Sillerud, D. Schafer, and W. H. Konigsberg, *JBC* **254**, 6426 (1979).

the initial studies of protein retention on DNA-cellulose columns (*38*). The fact that the protein elutes from double-stranded DNA-cellulose at lower salt concentrations than from single-stranded DNA-cellulose also suggested that the net affinity of the protein for single-stranded DNA should exceed that for double-stranded DNA, and that binding involves ionic interactions. Also, the observation that the amount of gene 32 protein retained on the single-stranded DNA cellulose column depends on the initial concentration of the protein suggested that the binding must be cooperative in nature; no effect of protein concentration on binding affinity was observed with double-standed DNA.

a. Interactions with Single-Stranded DNA. These early observations formed the basis for subsequent, more quantitative, investigations. Alberts and Frey (*6*) showed, using sedimentation techniques, that saturation of single-stranded (fd phage) DNA occurred at a ratio of ~10 nucleotide residues per protein monomer. In addition, these workers estimated that the affinity of gene 32 protein for a site adjacent to a bound protein molecule should be at least 80-fold greater than its affinity for free DNA.

Electron microscopic visualization of gene 32 protein–nucleic acid complexes supported the idea that the protein binds cooperatively to single-stranded nucleic acids. With subsaturating amounts of gene 32 protein, binding was seen to occur in clusters; some fd DNA molecules were coated entirely while others showed no bound protein. The protein coats single-stranded DNA to form a flexible, rodlike complex with a diameter of 60 Å and an internucleotide spacing of 4.6 Å (*79*). In addition, these studies also showed that although gene 32 protein does not melt native double-stranded DNA, it can invade A-T rich regions of duplex DNA *under irreversible binding conditions* (e.g., here with glutaraldehyde as a covalent binding agent).

i. Polynucleotides. Quantitative studies of the interaction of gene 32 protein with nucleic acids have utilized changes in the optical properties of either the protein or the nucleic acid that result from complex formation (Section II,B). Because binding of gene 32 protein results in an unstacking of the bases within the nucleic acid, changes in circular dichroism (*4, 27, 77*), UV absorbance (*4, 5, 14*), and fluorescence of nucleic acid analogues (*5, 14*) can be used to monitor binding. In general, the binding of gene 32 protein to polynucleotides results in changes similar to those observed upon thermal denaturation of polynucleotides; i.e., a decrease in the CD signal and an increase in the UV absorbances of the solution. In addition the binding of protein is monitored by following the quenching of the

79. H. Delius, N. J. Mantell, and B. M. Alberts, *JMB* **67**, 341 (1972).

intrinsic fluorescence of the protein. Upon binding to nucleic acids this fluorescence is quenched as much as 60%; the exact value depends on the length and base composition of the nucleic acid lattice used (*5, 12, 14, 33, 34*).

Under conditions of tight binding, the binding site size (n) can be determined (see Section II,B,2). Values of n ranging from 5 to 11 nucleotide residues/protein monomer (*12, 27*) have been obtained, with most values clustering at $n = 7 \pm 1$ (*4, 5, 14, 34*).

Under such stoichiometric binding conditions [in general (NaCl) ≤ 0.1 M], it is only possible to determine a lower limit for the net binding constant; i.e., $K\omega \geq 10^8\ M^{-1}$ (*4, 12, 27*). However, by performing titrations in increasing concentrations of NaCl, it is possible to weaken the affinity of the protein so that nonstoichiometric titrations are observed (*5, 14*). As shown in Fig. 3, at 0.01 M NaCl the binding is very tight; however, at 0.35

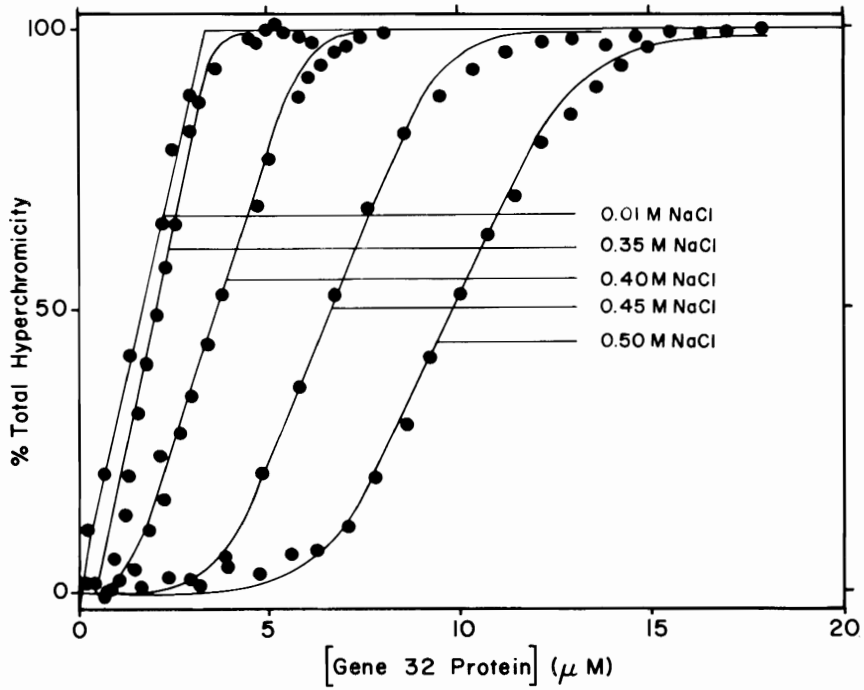


FIG. 3. Titration curves for the binding of T4 gene 32 protein to a poly(rA) lattice at 25° as a function of NaCl concentrations. The solid curves (except for that labeled "0.01 M NaCl," for which binding was stoichiometric) are calculated using the following best-fit parameters: $n = 7$ nucleotide residues, $\omega = 2 \times 10^3$ and K determined from the measured value of $K\omega$. The points represent the experimental data [taken from Ref. (*14*)].

M NaCl, there is a slight (sigmoid) lag before binding and this lag becomes very pronounced at still higher NaCl concentrations. These data display classical cooperative (sigmoid) binding profiles, and quantitative analysis shows that each of these curves can be fit with a cooperativity parameter (ω) of $\sim 2 \times 10^3$. Values of the net affinity constant, $K\omega$, can be determined from each set of data by utilizing the fact that at the midpoint of each titration, $K\omega = 1/[\text{protein}]_{\text{free}}$ (*9*). The values of $K\omega$ for the four nonstoichiometric binding curves in Fig. 3 are $1.7 \times 10^6\ M^{-1}$, $8.2 \times 10^5\ M^{-1}$, $2.5 \times 10^5\ M^{-1}$, and $1.6 \times 10^5\ M^{-1}$, reading from left to right. This represents a significant dependence of the net binding affinity of gene 32 protein for polynucleotides on salt concentration, with $K\omega$ decreasing ~ 7 orders of magnitude per 10-fold increase in salt concentration (*14*).

These direct titration procedures, as well as a salt-induced dissociation procedure, were used to determine values of $K\omega$ for a variety of polynucleotides as a function of salt concentration (*5, 14*). The resultant plots of log $K\omega$ versus log [NaCl] demonstrate the steep dependence of the net protein affinity on [NaCl] ($\partial \log K\omega / \partial \log [\text{NaCl}] \cong -7$), as well as the fact that there is a distinct specificity of binding in terms of polynucleotide base and sugar composition. Qualitative experiments involving the competition of two nucleic acid lattices for limiting amounts of gene 32 protein have also established the existence of such a nucleic acid binding specificity by using either changes in electron spin resonance to monitor competition with nitroxide-labeled polynucleotides (*32*), or by using differences in hyperchromism in polynucleotides upon binding (*5*). The hierarchy of affinities that has been established by direct measurements of $K\omega$ is (in order of increasing affinity at 0.2 M NaCl): poly(rC) < poly(rU) < poly(rA) < poly(dA) < poly(dC) < poly(dU) < poly(rI) < poly(dI) < poly(dT) [see Table II of Ref. (5) for a quantitative comparison of $K\omega$ values]. In addition, quantitative analysis of gene 32 protein binding isotherms has yielded separate values of K and ω for a number of polynucleotides at different salt concentrations (*5, 14*). The results of these analyses show that the cooperativity parameter is independent of salt concentration and essentially independent of polynucleotide type (ω values range from 10^3 to 10^4), and that virtually all of the salt dependence of $K\omega$ resides in the intrinsic binding constant, K.

The results of these equilibrium studies on the interaction of gene 32 protein with single-stranded polynucleotides can be summarized as follows: (1) Binding is highly salt-dependent ($\partial \log K\omega / \partial \log [\text{NaCl}] \cong -7$); (2) binding is highly cooperative ($\omega \cong 10^3 - 10^4$); (3) ω is independent of salt concentration and relatively independent of polynucleotide type; (4) K is highly salt-dependent ($\partial \log K\omega / \partial \log [\text{NaCl}] \cong -7$); (5) there is distinct specificity of binding with respect to polynucleotide types, and this is

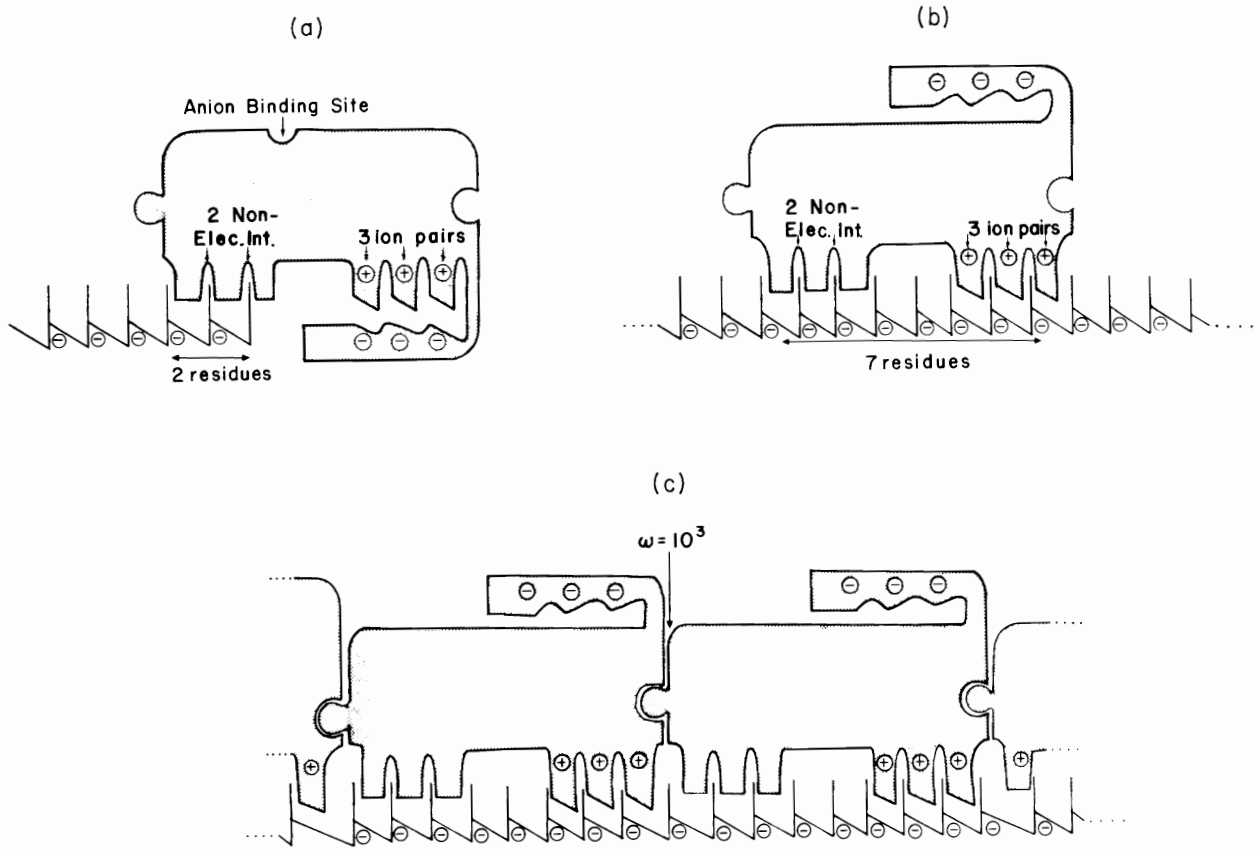
mainly reflected in differences in K; (6) the affinity for a *deoxy*polynucleotide is always greater than that for the homologous *ribo*polynucleotide (differences in $K\omega$ range between 10^1 to 10^4); and (7) values of $K\omega$ for randomly copolymerized polynucleotides and natural DNA can be calculated from compositionally weighted averages of $K\omega$ for the homopolynucleotides representing each of the constituent bases (*5*).

The approach of Record *et al.* (*15*) (see Section II,B,3) was used to analyze the salt (both cation and anion) concentration dependence of the binding of gene 32 protein to polynucleotide lattices (*14*). The results indicate that over one-half of the dependence of $K\omega$ on NaCl concentration results from the release of ~5 chloride ions from the protein on polynucleotide binding, whereas the remainder of the salt dependence is due to the formation of ~3 charge-charge interactions between basic amino acid residues on the protein and negatively charged phosphates of the polynucleotide backbone. For the other anions studied, the classical Hofmeister series of ion binding affinity was followed (*80*); e.g., only ~2 acetate ions, and ~0–1 fluoride ions were released per protein monomer bound.

ii. Oligonucleotides. The interactions of short ($l \leq 8$ nucleotide residues) oligonucleotides with gene 32 protein differ quantitatively and qualitatively from those observed with polynucleotides (*12, 14, 23*). The following generalizations emerge from these studies: (i) The oligonucleotide binding interaction shows very little salt dependence ($\partial \log K_{\text{oligo}} / \partial \log [\text{NaCl}] \cong -0.3$); (ii) this interaction shows little dependence on oligonucleotide base composition (factors of less than 2 to 3 in K_{oligo}); (iii) the dependence of the interaction on sugar type is small (K_{oligo} ~ twofold greater for oligodeoxyribonucleotides than for the homologous oligoribonucleotides); and (iv) there is little dependence of the observed value of K_{oligo} on lattice length (i.e., the expected statistical factor is not observed; see Section II,B,3).

These results may be interpreted as follows: The low salt dependence suggests that, at most, one (and possibly no) ionic interactions are involved in the formation of the protein–oligonucleotide complex. The absence of the statistical factor in K_{oligo} suggests that the oligonucleotide is not able to move freely within the binding site, and thus that the protein may bind to an end of the oligonucleotide lattice. This also suggests that a dinucleotide monophosphate constitutes the minimum binding unit for oligonucleotide binding (*12*). Finally, the dinucleotides pApA and ApAp bind to gene 32 protein with identical affinities but cause very different levels of fluorescence quenching (2 versus 17%). This shows that binding

80. P. H. von Hippel and T. Schleich, *Acc. Chem. Res.* **2**, 259 (1969).


(a)
Anion Binding Site
2 Non-
Elec. Int.
3 ion pairs
2 residues
(b)
2 Non-
Elec. Int.
3 ion pairs
7 residues
(c)
ω = 10³

of oligonucleotides to the protein is polar; i.e., that an oligonucleotide can interact with the binding site in only one orientation.

iii. Binding models. The facts discussed in the previous sections concerning the interaction of gene 32 protein with polynucleotides and oligonucleotides, as well as some of the results of binding studies with the proteolytic products G32P*I and G32P*III, can be incorporated into a schematic model of gene 32 protein binding to nucleic acids, which is presented in Fig. 4. Two different types of binding conformations are proposed; these are termed the oligonucleotide (a), and the polynucleotide (b and c) binding modes.

In Fig. 4a (the oligonucleotide binding mode) the protein is shown to interact nonelectrostatically with two residues at the end of an oligonucleotide. This is consistent with the fact that a dinucleotide binds better than a mononucleotide and that further increases in length of the oligonucleotide has no effect on K_{oligo}. The lack of statistical effect on binding for larger oligonucleotides is accounted for in the model by the presence of a steric constraint (the "arm" or "flap" in the drawing). No electrostatic interactions are shown for oligonucleotide binding.

In Fig. 4 (the polynucleotide binding mode), an additional binding subsite becomes available due to a displacement of the flap. We assume this displacement occurs because polynucleotides provide very few lattice ends relative to the number of potential internal binding sites. Displacement of the flap is a free energy-requiring process; thus almost all of the values of K for polynucleotides are lower than the values of K_{oligo}. However the displacement of the flap uncovers ~3 positively charged amino acid residues on the protein that can form charge–charge interactions with the phosphate residues of the polynucleotide. Simultaneously, an anion binding site (or sites) is destroyed, resulting in the release of the anions upon formation of the complex. In addition, Fig. 4 shows that gene 32 protein covers ~7 nucleotide residues when bound, and indicates that

FIG. 4. Schematic models of three modes of gene 32 protein binding to nucleic acid lattices. (a) Binding in the oligonucleotide binding mode; note the presence of the block to statistical "shuffling" of the oligonucleotide in the binding site, as well as the presence of the anion binding site. (b) Isolated binding in the polynucleotide mode; note that the "shuffling block" has been moved away, exposing the positively changed binding subsite, and the anion binding site has been disrupted. Also the nucleic acid lattice between the two binding subsites is somewhat stretched, and the nonelectrostatic (XpX) binding subsite is somewhat altered, indicating that gene 32 protein binding in this mode (conformation) shows somewhat enhanced base compositional specificity. (c) Continguous binding in the polynucleotide mode; the gene 32 protein binding conformation is unchanged from that of (b), except for protein–protein interactions and cooperative extension of the nucleic acid lattice between and through contiguously bound protein monomers [taken from Ref. (*14*)].

binding results in increasing the internucleotide spacing of the DNA lattice to ~5 Å.

In Fig. 4c (the cooperative polynucleotide binding mode) contiguous proteins are shown to interact with one another; these interactions are nonelectrostatic since the magnitude of ω is salt-independent and reflects primarily protein–protein interactions [see Ref. (*74*)].

These models are also consistent with proteolytic digestion studies of gene 32 protein complexed with nucleic acids (*69, 70*). The enhancement of the rate of proteolysis of the carboxy terminus of DNA-bound gene 32 protein relative to that of the free protein, suggests that the flap in Fig. 4a may comprise the carboxy terminus of the polypeptide chain. In contrast, cooperative polynucleotide binding protects the amino terminus, just as the "bump" on the left of each protein monomer is shown to be protected in Fig. 4c. Finally, the binding of oligonucleotides has no effect on the proteolytic digestion patterns obtained (relative to those obtained with the free protein), which is consistent with the model shown in Fig. 4a.

This schematic model of gene 32 protein binding also predicts that G32P*I should display altered oligonucleotide binding properties, that G32P*II should not bind polynucleotides cooperatively, and that G32P*III should share both of the above properties. These predictions are consistent with experimental observations (see Section c below).

b. Interaction with Double-Stranded DNA. Most studies on the interaction of gene 32 protein with duplex DNA have focused on the effect of protein on the thermal denaturation profile of the nucleic acid (*4, 6, 77*). It has been shown that gene 32 protein will denature poly(dA-dT), but is incapable of denaturing native T4 or T7 DNA (*4, 6*). The binding of gene 32 protein to duplex DNA is quite weak, with values of K ranging from $4 \times 10^4\ M^{-1}$ at 0.02 M NaCl to $8.0 \times 10^3\ M^{-1}$ at 0.05 M NaCl (*4*). In binding to duplex DNA, gene 32 protein covers approximately 5 base pairs and forms 1 to 2 electrostatic interactions with phosphate groups. There is no evidence for cooperative binding to duplex DNA.

These binding parameters for gene 32 protein to duplex DNA can be utilized, together with theoretical approaches to ligand-perturbed double-helix $\rightleftarrows$ coil transitions, to determine values of K and ω for the gene 32 protein interaction with single-stranded DNA (*26*). The results of such calculations, based on the thermal denaturation measurements of poly(dA-dT) at 0.01 M NaCl, are that $n \cong 7.5$, $K \cong 10^7\ M^{-1}$, and $\omega \cong 10^3$ (*4*); these parameters are in excellent agreement with those extracted from direct measurements of the affinity of gene 32 protein for single-stranded nucleic acids.

Using these values, it is possible to calculate the expected thermal denaturation temperature (T_m) of natural duplex DNA in the presence of

gene 32 protein. Although it can be calculated that gene 32 protein should lower the T_m of the T7 DNA by ~60°, no destabilization was observed under any conditions. No melting of the duplex DNA by gene 32 protein was observed for either whole or sonicated T7 DNA, even after 8 hours at a temperature 20° below the unperturbed T_m. Similar results were obtained for other natural DNAs, including those extracted from calf thymus (58% A-T), *Clostridium perfringens* (69% A-T), and *Micrococcus lysodeikticus* (28% A-T). These results strongly suggest that the melting of native duplex DNA by gene 32 protein is kinetically blocked.

c. Interaction of the Proteolytic Digestion Products of Gene 32 Protein with Nucleic Acids. The digestion products of gene 32 protein formed by limited proteolysis have been shown to possess different DNA binding properties, as assessed by their affinities for DNA-cellulose (*67*) (also see Section b above]. To more fully understand the molecular aspects of the interactions of gene 32 protein with nucleic acids, and to investigate the functional role of the N- and C-terminal domains in these interactions, the binding of these products to nucleic acids has been studied (*67, 68, 74, 81*).

*i. G32P*I.* The oligo- and polynucleotide binding properties of G32P*I are very similar to those of the native protein (*74*). With respect to oligonucleotide binding, all interactions are essentially the same for the two proteins except that G32P*I shows a greater electrostatic component of the binding free energy and a greater salt dependence of binding for 6-mers and 8-mers, with the nonelectrostatic component remaining the same for both species. This result is most easily interpreted in terms of the model shown in Fig. 4a, by postulating that this proteolytic cleavage removes at least part of the flap on the lower right-hand side of the model, thus making the charge-containing binding protein subsite available to the longer oligonucleotides, at least in part.

The polynucleotide binding properties of the native protein and G32P*I are also virtually identical; the only significant difference is that the value of K is 2- to 3-fold greater for G32P*I for all the polynucleotides. This single (and small) thermodynamic difference between the two proteins is particularly noteworthy when we recall that the gene 32 protein cannot melt native double-stranded DNA, whereas G32P*I can (*66, 68*). It seems unlikely that this small difference in binding affinity can explain the melting differences, suggesting that melting by the two proteins must involve very different kinetic pathways (*4*).

This conclusion, and the thermodynamic differences have been confirmed by comparing the differences between the melting temperature depressions induced by gene 32 protein and G32P*I on poly(dA-dT). The

81. E. K. Spicer, K. R. Williams, and W. H. Konigsberg, *JBC* **254**, 6433 (1979).

observed change in T_m (~12°) (*70*) corresponds to about a twofold difference in net binding affinity. Based on these results (since G32P*I denatures native T4 DNA at ~70° below the unperturbed T_m in 0.01 *M* NaCl) (*37*), we would expect that the T_m of this DNA should be lowered ~60° in the presence of native gene 32 protein if this melting goes to equilibrium. A T_m depression of exactly this magnitude has, in fact, been calculated, but was not observed (*4*). This confirms that gene 32 protein is indeed kinetically blocked from melting native DNA, and that this block is effectively removed when G32P*I serves as the melting protein.

*ii. G32P*III.* This proteolytic product of gene 32 protein differs appreciably in its nucleic acid binding properties from both the native protein and G32P*I, though it retains some features of these precursors. The most striking change is that binding cooperativity is abolished; G32P*III binds to polynucleotides with a measured value of $\omega \cong 1$ (*74*). G32P*III has a smaller site size (n = 5–6) than either the native protein or GP32*I; in addition, the magnitude of the salt dependence of binding of this product to both polynucleotides and oligonucleotides is somewhat changed, though the overall relative specificity of polynucleotide binding displayed by the native protein is approximately retained.

Despite the fact that G32P*III binds to polynucleotides noncooperatively, binding induces the same changes observed with the cooperatively bound proteins in the optical properties of the polynucleotides (*74*). This suggests that lattice distortion [lattice-mediated cooperativity; see Ref. (*82*)] is not centrally involved in cooperative binding, and that protein–protein interactions are predominantly responsible for the binding cooperativity of gene 32 protein (and GP32*I). This cooperativity appears to be localized at the amino terminus of the protein, and the results of the *Staphalococcus aureus* protease digestion experiments suggest that essential residues for this property fall within 9 residues of the N terminus (*71*). The "ball" of the "ball-and-socket" interaction between adjacent protein molecules in Fig. 4c provides a simple functional representation of the residues that are cleaved off in forming G32P*III (and G32P*II).

*iii. G32P*II.* No direct quantitative studies of the interaction of G32P*II with nucleic acids have as yet been carried out. However, based on its DNA-cellulose elution behavior and on the studies with G32P*I and *III, it is possible to establish some of the properties of this proteolytic product by inference.

Since the N terminus is required for cooperative interactions, G32P*II should bind to polynucleotides noncooperatively. This conclusion is con-

82. P. H. von Hippel, D. E. Jensen, R. C. Kelly, and J. D. McGhee, *in* "Nucleic Acid-Protein Recognition" (H. J. Voegl, ed.), p. 65. Academic Press, New York, 1977.

sistent with both its single-stranded DNA-cellulose binding behavior (*67*) and its calorimetric properties (*73*) (see above). In addition, since the N terminus may also be responsible, at least in part, for protein self-association, the G32P*II product should not form indefinite aggregates; this is also confirmed by preliminary studies (*71, 73*). Furthermore, since G32P*II retains its C-terminal peptide, we may also expect that this product (like the native protein) will be kinetically blocked from melting native double-stranded DNA. This is supported by the fact that G32P*II (like native gene 32 protein but unlike G32P*I and *III), does not bind to double-stranded DNA-cellulose (*67*).

d. Kinetics of the Binding of Gene 32 Protein to Nucleic Acids. As discussed in the preceding section, the *kinetics* (as well as the thermodynamics) of the interaction of gene 32 protein with its various nucleic acid substrates must be elucidated in order to develop a complete understanding of the physicochemical and biological properties of this protein. Such studies are quite incomplete, and have focused mainly on double-stranded DNA, poly(dA-dT) denaturation and renaturation rates (*6*), and the kinetics of the binding of gene 32 protein to single-stranded DNA (*18, 20, 21, 83*).

i. Denaturation of poly(dA-dT). Since gene 32 protein is kinetically blocked from denaturing double-stranded DNA (*6, 12*), denaturation rate studies with the native protein have been possible only with poly(dA-dT). The melting of this model DNA duplex has been monitored spectrophotometrically, using the increase in OD_{260} to follow the reaction. The rate of melting is slow, and as expected depends strongly on salt concentration (*6*); e.g., a half-time for denaturation (at 25°) of ~20 min was observed in 10 m*M* $MgSO_4$. This $t_{1/2}$ increased to ~300 min in ~40 m*M* $MgSO_4$. More detailed kinetic studies on this system, and of the kinetics of the denaturation of native DNA by G32P*I, are in progress in this laboratory (*84*).

ii. Renaturation of double-stranded DNA. Since DNA denaturation is reversible the perturbation of the rate of *renaturation* of DNA by gene 32 has also been studied. The DNAs used in these studies were first denatured by alkali, then (after neutralization of the solution) gene 32 protein in various concentrations was added, and finally $MgSO_4$ was added to induce renaturation. The reactions were monitored by following decreases in OD_{260}. Studies of this sort have shown that renaturation rates can be accelerated over 1000-fold by gene 32 protein. As observed in the denaturation rate studies, these effects are salt-dependent, with the stimulatory

83. P. Suau, J. J. Toulmé, and C. Hélène, *Nucleic Acids Res.* **8**, 1357 (1980).

84. N. Lonberg and S. C. Kowalczykowski, unpublished observations.

effect of gene 32 protein on the rate decreasing as the salt concentration is increased.

Studies in which the concentration of DNA has been varied at less than saturating concentrations of gene 32 protein have shown that the process that is being accelerated is primarily the bimolecular (nucleation) step in the reassociation of the complementary DNA strands (*6*); the rate of the subsequent "zippering" process is not appreciably affected. It has been concluded that gene 32 protein accelerates renaturation by holding the single-stranded DNA moieties in a favorable, unfolded conformation, which increases the probability of pairing by complementary sequences during strand collisions. It has also been suggested, on the basis of these studies, that bound gene 32 protein might be displaced as renaturation proceeds (*6*).

iii. Kinetics of the association and dissociation reactions of gene 32 protein with single-stranded nucleic acids. The kinetics of the interaction of gene 32 protein with various single-stranded nucleic acids has been investigated by using stopped-flow techniques; the course of the reaction is followed by monitoring the changes in optical properties of either the protein or polynucleotide upon complex formation (*18, 20, 21, 85*).

Dissociation kinetics can be studied by subjecting the preformed protein-nucleic acid complex to an ionic strength jump, resulting in dissociation of the complex (*18*). The rate of dissociation was found to depend strongly on the extent of protein saturation of the nucleic acid lattice (*18, 20, 21*), as well as on the composition of the nucleic acid. In fact, the same order of polynucleotide affinities observed in the equilibrium experiments is reflected in the dissociation rates as well (*20, 21*). It has been suggested that the dissociation of gene 32 protein occurs primarily from the ends of cooperatively bound clusters of protein, and that this type of mechanism facilitates the renaturation of DNA strands during various biological processes (*18*). In addition, similar studies have suggested that the bound protein molecules do not behave independently during dissociation, but may be somewhat mobile on the DNA lattice (*20*).

Studies of the kinetics of the association of gene 32 protein have shown that this is a multistep process (*85*). The steps include (at least) the preequilibrium formation of noncooperatively bound protein, the growth of cooperatively bound clusters of gene 32 protein, and finally the redistribution of the clusters to form a final equilibrium state. The measured bimolecular association rate constant is $3 \times 10^6\ M^{-1}$ (nucleotide) sec^{-1} [or $2 \times 10^7\ M^{-1}$ (protein) sec^{-1}]. In addition, the data suggest that the

85. T. M. Lohman and S. C. Kowalczykowski, *JMB*, in press (1981).

cooperative growth step may occur by a process that involves translocation of bound proteins along the DNA lattice (*86*).

e. Molecular Details of the Binding Interaction. Several approaches have been made to attempt to elucidate additional molecular aspects of the nucleic acid binding reaction of gene 32 protein. Equilibrium binding studies (see Section a above) have suggested that two nucleotide residues and two or three nucleotide phosphates interact directly with the protein in the polynucleotide binding mode. We ask here which protein residues are involved in the interaction? This question can be quite definite since the amino acid sequence of the protein is available (*63*).

The involvement of aromatic amino acids in protein–nucleic acid interactions has been much investigated because of the potential of these moieties to stack on, or intercalate between, nucleotide bases [e.g., see (*92*)]. Furthermore tyrosine residues have been strongly implicated in the binding of phage fd gene 5 protein to DNA (see Section IV,B,2). Thus chemical modification studies of the nucleic acid binding site of gene 32 protein have also focused on aromatic residues. Tetranitromethane has been used as a chemical probe to determine the accessibility of tryosine residues in gene 32 protein. It was found that 4 or 5 out of 8 tyrosine residues on the native protein can be modified with this reagent. Nitration of these residues completely abolishes the DNA binding activity of the protein; furthermore, no tyrosine residues are modified when the protein is bound to single-stranded DNA (*27*). These results strongly implicate tyrosine residues in the nucleic acid binding interaction of gene 32 protein.

The position within the protein primary structure of the tyrosine residues nitrated in the above study is not known. However the amino acid sequence of gene 32 protein shows a very suggestive distribution of tyrosine residues. Five of the eight tyrosine residues are located within the proteolytically resistant protein core region, and are distributed at 7 to 9 residue intervals along the sequence (*63*). Whether these residues actually comprise part of the DNA binding site must await further characterization of the protein.

86. The processes by means of which *E. coli lac* repressor can translocate on DNA have been extensively studied, both theoretically and experimentally (*87–91*). These results may provide insight into analogous kinetic mechanisms which may be involved in single-stranded DNA binding protein-nucleic acid interactions.

87. P. H. Richter and M. Eigen, *Biophys. Chem.* **2,** 255 (1974).

88. O. G. Berg and C. Blomberg, *Biophys. Chem.* **4,** 367 (1976).

89. M. D. Barkley, P. A. Lewis, and G. E. Sullivan, *BJ* **32,** 452 (1980).

90. O. G. Berg, R. B. Winter, and P. H. von Hippel, *Biochemistry,* in press (1981).

91. R. B. Winter, O. G. Berg, and P. H. von Hippel, *Biochemistry,* in press (1981).

92. C. Hélène, *in* "Excited States in Organic Chemistry and Biochemistry" (B. Pullman and N. Goldblum, eds.), p. 65. D. Reidel, Holland, 1977.

The involvement of aromatic amino acids in the nucleic acid binding site of gene 32 protein is also currently being examined by ^{19}F-NMR techniques using fluorine-substituted analogues of these amino acid residues. Preliminary results show that all five tryptophan resonances can be resolved, and that none are shifted when the protein binds to various oligonucleotides (an upfield shift would be expected if extensive stacking interaction between tryptophan residues and nucleotide bases accompanied binding; see Section IV,B,2,e for a comparable study on fd gene 5 protein). Shifts of some of the fluorotyrosine resonances are observed on single-stranded DNA binding, suggesting that some tyrosine residues may be involved in polynucleotide binding (*93*).

3. *Biological Roles*

a. Replication. Numerous genetic studies have shown that gene 32 protein is absolutely essential to the DNA replication of the T4 phage (*38, 57a, 59, 94*). Gene 32 protein is required continuously throughout the replication process, and if a phage that contains a temperature-sensitive mutation in gene 32 protein is switched to a nonpermissive temperature, DNA replication ceases immediately (*38, 94*).

Some of the possible roles of gene 32 protein in DNA replication can also be demonstrated in *in vitro* experiments. Gene 32 protein has been shown to bring about a 5- to 10-fold increase in the rate of DNA polymerization by T4 DNA polymerase on a primed single-stranded homopolynucleotide template (*95*). Also sedimentation analyses have demonstrated that gene 32 protein binds weakly but specifically to T4 polymerase in solution (*95*). Neither of these effects is observed in heterologous systems, i.e., with the substitution of *E. coli* polymerase I for T4 polymerase, or the substitution of *E. coli* binding protein (SSB) for gene 32 proteins (*96–98*).

A recent kinetic investigation has confirmed and extended these results by demonstrating that the T4 DNA polymerase engaged in synthesizing a complementary strand to single-stranded fd phage DNA will pause at sites on the template that are capable of forming stable (about -15 kcal/mol)

93. S. C. Kowalczykowski, R. A. Anderson, V. Ochs, F. W. Dahlquist, and P. H. von Hippel, unpublished observations.

94. S. Riva, A. Cascino, and E. P. Geiduschek, *JMB* **54,** 85 (1970).

95. J. A. Huberman, A. Kornberg, and B. M. Alberts, *JMB* **62,** 39 (1971).

96. N. Segal, H. Delius, T. Kornberg, M. L. Gefter, and B. M. Alberts, *PNAS* **69,** 3537 (1972).

97. C. C. Liu, R. L. Burke, U. Hiber, J. Barry, and B. M. Alberts, *CSHSQB* **43,** 469 (1979).

98. R. L. Burke, B. M. Alberts, and J. Hosada, *JBC* **255,** 11484 (1980).

secondary structures, and that the addition of gene 32 protein causes a 7.5-fold stimulation of the rate of DNA polymerization in this system, presumably as a consequence of the destabilization of the hairpin structures (*99*).

If nicked double-stranded DNA is used as a template for DNA polymerase the situation becomes more complex, and the addition of gene 32 protein has only a limited effect (*100, 101*); less than 1% of the maximum *in vitro* rate is obtained (*101*). In addition, the product formed is largely A-T rich, and rapidly renaturable, indicating that A-T rich regions in the DNA are being copied preferentially (*100*). However, reasonable rates of DNA polymerization can be achieved if additional T4-coded proteins are added to the reaction mixture [for reviews see (*22, 97, 102–104*)]. The T4 accessory proteins (gene products 44, 45, and 62), together with gene 32 protein, are essential for efficient replication of double-stranded DNA on a nicked duplex template. The exact role of gene 32 protein in this process is not clear, but the free concentration of gene 32 protein in such *in vitro* reconstituted replication systems is very important. If gene 32 protein is omitted, no synthesis is observed; furthermore, the rate of replication fork movement increases almost linearly with increasing concentration of gene 32 protein, up to a rate of ~200 nucleotides/sec at 200 μg/ml (*22*). Clearly, gene 32 protein plays an important role in helix destabilization in this system. However, there must be additional sources of free energy for helix unwinding since the rate of this five-protein system is well below that observed *in vivo* [~750 nucleotides/sec (*105*)], and also below that observed when additional T4 proteins are included in the mixture to reconstitute the seven-protein system [~500 nucleotides/sec (*22, 101*)].

In addition to the stimulatory effects on DNA polymerization rates, which are due to the ability of gene 32 protein to destabilize weak hairpins in template DNA, this protein may also stimulate DNA synthesis by increasing the processivity of the T4 polymerase. An increase in pro-

99. C. C. Huang and J. E. Hearst, *Anal. Biochem.* **103,** 127 (1980).

100. N. G. Nossal, *JBC* **249,** 5668 (1974).

101. N. K. Sinha, C. F. Morris, and B. M. Alberts, *JBC* **255,** 4290 (1980).

102. B. M. Alberts, C. F. Morris, D. Mace, N. Sinka, M. Bittner, and L. Moran, *in* "DNA Synthesis and Its Regulation" (M. Gouliun and P. Hanawalt, eds.), Vol. III, p. 241. Benjamin, Menlo Park, California, 1975.

103. B. M. Alberts, J. Barry, M. Bittner, M. Davies, H. Hama-Inaba, C. C. Liu, D. Mace, L. Moran, C. F. Morris, J. Piperno, and N. K. Sinka, *in* "Nucleic Acid-Protein Regulation" (H. J. Vogel, ed.), p. 31. Academic Press, New York, 1977.

104. N. G. Nossal and B. M. Peterlin, *JBC* **254,** 6032 (1979).

105. D. McCarthy, C. Minner, H. Bernstein, and C. Bernstein, *JMB* **106,** 963 (1976).

cessivity allows the polymerase to remain attached to the DNA primer-template for a longer time per polymerase binding event, thereby increasing the macroscopic rate of DNA synthesis without a concomitant change in the microscopic polymerization rate (*22, 101–103*). Recent quantitative studies on the processivity of the T4 replication systems have demonstrated that gene 32 protein is required for processive synthesis *in vitro;* if this protein is omitted, replication becomes much less processive or completely dispersive (*7, 106*).

It has been shown, by investigating the effects of replacing gene 32 protein by G32P*I in *in vitro* replication systems, that the carboxy-terminal domain of this protein is essential in at least two aspects of the replication process. G32P*I does not interact properly with T4 polymerase and also inhibits RNA primer formation (*72, 98*). T4 DNA polymerase does not cosediment with G32P*I as it does with gene 32 protein, and G32P*I inhibits DNA synthesis by T4 polymerase when single-stranded templates are used. In addition, G32P*I does not interact with the priming protein (gene product 61), leading to the disruption of both primer synthesis and primer utilization (*72, 98*).

b. Recombination and Repair. Numerous investigations have demonstrated that gene 32 protein is essential for recombination in the T4 genome (*58, 59, 107–109*). Recombination in T4 does not occur (*59*), or is greatly reduced, if the phages are grown under semipermissive conditions (*107*). In addition the formation of branched DNA molecules (intermediates in recombination) is reduced ~10-fold in phages that carry a defective gene 32 protein gene (*108*).

It has been suggested that the helix-destabilizing capacity of gene 32 protein and, perhaps more importantly, the DNA renaturation activity of this protein, may play a role in genetic recombination (*6*). The notion that this protein might catalyze the formation of heteroduplex molecules of DNA has been supported by *in vitro* experiments showing that the formation of joint DNA molecules infective in a spheroplast assay is stimulated ~5-fold by added gene 32 protein (*109*).

Gene 32 protein has also been implicated in DNA repair. Phages that carry a temperature-sensitive gene 32 are defective in the repair of UV-induced lesions at nonpermissive temperatures; this may be due to the

106. J. W. Newport, S. C. Kowalczykowski, N. Lonberg, L. S. Paul, and P. H. von Hippel, *in* "Mechanistic Studies on DNA Replication and Genetic Recombination" (B. M. Alberts and C. F. Fox, eds.), ICN-UCLA Symp. Mol. Cellular Biol., Vol. 19. Academic Press, New York, 1980.

107. H. Berger, A. J. Warren, and K. E. Fry, *J. Virol.* **3,** 171 (1969).

108. T. R. Broker and I. R. Lehman, *JMB* **60,** 131 (1971).

109. W. Wackernagal and C. M. Radding, *PNAS* **71,** 431 (1974).

inability of the mutant gene 32 protein to bind to (and protect against nucleases) the single-stranded, gapped sections of DNA produced in the excision-repair process (*110*). This supposition has been strengthened by the observation that T4 phages that are temperature-sensitive in gene 32 suffer rapid nucleolytic degradation of their DNA when shifted to non-permissive temperatures (*111*).

c. Autogenous Regulation of Gene 32 Protein Synthesis. Gene 32 protein regulates its own expression at the translational level (*44–46, 112–114*). The observations that have contributed to this conclusion include (1) nonsense mutations in gene 32 protein overproduce the nonsense fragment peptides, (2) overproduction in mixed infections is recessive, (3) the amount of overproduced gene 32 protein is related to the amount of single-stranded DNA present (*44, 45*), and (4) gene 32 protein mRNA is very stable. These primarily genetic conclusions have been confirmed by *in vitro* experiments showing that purified gene 32 protein can specifically inhibit translation of its own mRNA, and this inhibition occurs only at concentrations of gene 32 protein sufficient to complex all the single-stranded DNA present [for a review, see Ref. (*113*)].

These data are consistent with a model in which gene 32 is synthesized as needed to saturate regions of single-stranded DNA that are produced in the course of replication, recombination, and repair. When sufficient protein has been produced the excess gene 32 protein binds specifically to gene 32 protein mRNA, and reversibly shuts down further translation of this message. The molecular basis of this recognition and specific shutdown have been speculated upon elsewhere (*5, 46, 106*). However we note, at a minimum, that the protein must (i) demonstrate a strong effective preference for single-stranded DNA over RNA sequences, and (ii) show a significant preferential affinity for its own mRNA relative to other T4 mRNAs. Physical chemical studies have shown that the first requirement is met as a consequence of the high degree of cooperativity of gene 32 protein binding (*5*). The exact nature of the binding site on gene 32 mRNA may soon be elucidated, since the relevant piece of DNA has been cloned and partially sequenced (*114*).

110. J. R. Wu and Y-C. Yeh, *J. Virol.* **12,** 758 (1973).

111. M. J. Curtis and B. M. Alberts, *JMB* **102,** 793 (1976).

112. M. Russel, L. Gold, H. Morrissett, and P. Z. O'Farrell, *JBC* **251,** 7263 (1976).

113. L. Gold, G. Lemaire, C. Martin, H. Morrissett, P. O'Conner, P. O'Farrell, M. Russel, and R. Shapiro, *in* "Nucleic Acid-Protein Interactions" (H. J. Vogel, ed.), p. 91. Academic Press, New York, 1977.

114. H. M. Krish, R. M. Duvoisin, B. Allet, and R. H. Epstein, *in* "Mechanistic Studies on DNA Replication and Genetic Recombination" (B. M. Alberts and C. F. Fox, ed.), ICN-UCLA Symp. Mol. Cellular Biol., Vol. 19. Academic Press, New York, 1980.

4. *Genetic Approaches*

Elegant genetic studies involving a large number of temperature-sensitive and *amber* mutations in gene 32 protein have demonstrated that this protein participates in a variety of phage functions, and researchers have mapped the regions of the protein that are involved in these interactions (*115, 116*). The results show that gene 32 protein interacts with T4 DNA polymerase, ligase, and gene 46–47 nuclease. In addition these studies have shown that the N-terminal domain of the protein is involved in DNA binding and in interactions with proteins that initiate DNA replication and recombination, whereas the C-terminal region may be required to modulate the activity of the nucleases that act during recombination, and to protect the DNA from excessive degradation. Several *E. coli* gene products (*dna*C, *dna*G) may be able to substitute for gene 32 protein in the first round of DNA replication, but the specific interaction of gene 32 protein and T4 DNA polymerase is essential to subsequent DNA replication and recombination. Thus gene 32 protein must play a central role in coordinating and controlling the activities of many of the enzymes involved in T4 regulation and expression.

B. Filamentous Phage Gene 5 Protein

The DNA binding protein of the filamentous phages M13, fd, ZJ-2 and f1 is the product of gene 5 of these phage genomes (*41, 117*). The protein is essential to the life cycle of the phages, and has been shown to be essential in controlling the switch to the production of single-stranded viral DNA from the double-stranded replicative form during replication (*118*).

As observed for T4 gene 32 protein, fd gene 5 protein is produced in large quantities (~75,000 copies/cell) and binds preferentially and cooperatively to single-stranded DNA, thereby lowering the T_m of both poly(dA-dT) and native DNA (*41, 117*). Gene 5 protein is found closely associated with single-stranded viral DNA in a molecular ratio of ~1600 gene 5 protein monomers to one viral DNA molecule. It is clear that gene 5 protein is not a structural protein of the fd viral coat (*119*). Although gene 5 protein has DNA binding properties similar to those of the other prokaryotic HDPs, its role in phage replication is unique, and is discussed in the following sections.

115. A. M. Breschkin and G. Mosig, *JMB* **112,** 279 (1977).
116. A. M. Breschkin and G. Mosig, *JMB* **112,** 295 (1977).
117. J. L. Oey and R. Knippers, *JMB* **68,** 125 (1972).
118. J. S. Salstrom and D. Pratt, *JMB* **61,** 489 (1971).
119. T. J. Henry and D. Pratt, *PNAS* **62,** 800 (1969).

1. *Physical Properties of the Protein*

The product of fd gene 5 is a small protein, with a molecular weight of ~10,000 (*41, 117, 119*). The complete amino acid sequences of both the M13 and the fd gene 5 protein have been determined and the exact value of the molecular weight (from the amino acid composition) is 9688 (*120, 121*). The sequences of the M13 and fd proteins are identical.

a. Aggregation State. Gene 5 protein exists in solution mainly in a monomer–dimer equilibrium state (*35, 117, 122*). Early sedimentation studies showed that gene 5 protein sediments either as a monomer (1.3 S) or as a dimer (1.9 S), depending on the ionic composition of the solution (*117*). Increasing concentration of salt induces dissociation of the dimer, so only monomer is present at salt (NaCl or $NaClO_4$) concentrations exceeding ~0.7 *M* (*35*). Estimates of the monomer-dimer association constant range from 10^6 M^{-1} in 0.15 *M* NaCl to 10^3 M^{-1} in 0.68 *M* NaCl. In contrast, a subsequent sedimentation equilibrium study showed that the dimeric species appears to be stable to dilution to concentrations as low as 0.075 mg/ml. Under the conditions used the dimer was also found to be stable to extremes of salt (0–0.5 *M* KCl), pH (5–11), and temperature (5° and 20°) (*122*). Only at concentrations greater than ~1 mg/ml did some formation of tetramer (7–8%) become apparent. The protein remains dimeric in saturating amounts of $d(pT)_4$, but in the presence of $d(pT)_8$ gene 5 becomes tetrameric. This may be due to the binding of two protein dimers of this oligonucleotide, and as such may reflect the cooperative bonding to DNA seen with longer single-stranded lattices.

b. Protein Structure. The successful crystallization and subsequent determination of the crystal structure of gene 5 protein to a resolution of 2.3 Å (*123–126*) makes possible an examination of molecular details at a level that is not accessible for any other DNA binding protein. The monomer of gene 5 protein has molecular dimensions of 45 Å × 25 Å × 30 Å. Its secondary structure consists entirely of antiparallel β-sheets, and contains no α-helical sequences whatsoever. This result is in agreement with sec-

120. Y. Nakashima, A. K. Dunker, D. A. Marvin, and W. Konigsberg, *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **40,** 290 (1974).

121. T. Cuypers, F. J. van der Ouderaa, and W. W. de Jong, *BBRC* **59,** 557 (1974).

122. S. J. Cavalier, K. E. Neet, and D. A. Goldthwait, *JMB* **102,** 697 (1976).

123. A. McPherson, I. J. Molineux, and A. Rich, *JMB* **106,** 1077 (1976).

124. A. McPherson, F. Jurnak, A. Wang, F. Kolpak, I. J. Molineux, and A. Rich, *CSHSQB* **43,** 21 (1979).

125. A. McPherson, F. A. Jurnak, A. H. J. Wang, I. Molineux, and A. Rich, *JMB* **134,** 379 (1979).

126. A. McPherson, F. Jurnak, A. Wang, F. Kolpak, A. Rich, I. J. Molineux, and P. Fitzgerald, *BJ* **32,** 155 (1980).

ondary structure estimates from circular dichroism (*29*), and with secondary structure predictions based on amino acid sequence (*127*).

The secondary structure consists basically of three elements: A three-stranded antiparallel β-sheet (residues 12–49), a two-stranded antiparallel β-ribbon (residues 50–70), and a second two-stranded antiparallel ribbon (residues 71–82). The first β-ribbon is involved in the protein–protein interactions that maintain the dimer in solution, whereas the second β-ribbon is believed to participate in the nearest-neighbor interactions responsible for cooperative binding (*126*).

Although the structure of the protein–nucleic acid complex has not been determined, studies on the unliganded gene 5 protein show that a 30 Å-long groove exists in the molecule. It is believed that this groove comprises the DNA binding site, based on its size and shape as well as on the presence of amino acid residues in this region that have been identified in solution studies as being involved in DNA binding (see Section IV,B,2,d). A three-dimensional representation of gene 5 protein is shown in Fig. 5.

2. *Protein–Nucleic Acid Interactions*

a. DNA Binding. The DNA binding properties of gene 5 protein were first analyzed using a sedimentation velocity technique (*41, 117*); these studies showed that gene 5 protein can saturate single-stranded DNA at a ratio of ~1 protein monomer per 4 DNA nucleotide residues. It was also observed that a 10-fold increase in protein concentration dramatically increased the affinity of the protein for DNA, suggesting that binding is cooperative. From these data it was concluded that the affinity of the protein for a contiguous binding site must be at least 60-fold greater than that for an isolated site (*41*). In addition, under the conditions employed, gene 5 protein exhibited no affinity for double-stranded T4 DNA, for double-stranded (replicating form) fd DNA, or for ribosomal RNA.

Filter binding has also been used to monitor the formation of the gene 5 protein-nucleic acid complex (*117, 122, 128*). A semiquantitative analysis of the filter binding data demonstrated that the binding of the protein to DNA is neither totally noncooperative nor infinitely cooperative (*128*). These results also suggested that the affinity of the protein for a contiguous binding site is ~1000-fold greater than for an isolated site; this estimate of the cooperativity parameter is of the same magnitude as that measured for the T4 gene 32 protein.

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

128. A. K. Dunker, *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **52,** 323 (1975).

The optical properties of both the protein and the nucleic acid are altered upon complex formation. The intensity of a CD band that has been attributed to the interaction of tyrosine residues is reduced on nucleic acid binding; this change saturates at ~4 nucleotide residues per protein monomer (*29*). The observed effects on the protein CD spectrum suggest that no major alteration in protein secondary structure accompanies DNA binding; however, as indicated above, the spectral changes that were seen strongly implicate alterations on binding in the environment of at least some tyrosine residues (*29*). Similar conclusions have been derived from an examination of the effect of DNA binding on the tyrosine fluorescence of the protein; a quenching of up to 70% is observed at a binding stoichiometry of 4 nucleotide residues per protein monomer (*35*).

The optical properties of the nucleic acid showed changes in the CD and UV absorbance spectra of protein binding that are consistent with base unstacking (*29, 127*); again these effects saturate at a nucleotide residue to protein monomer ratio of ~4 : 1. The association constant under these low salt conditions has been estimated to be greater than $\sim 10^8\ M^{-1}$ for gene 5 protein binding to single-stranded DNA or RNA (*127*). As seen with gene 32 protein, this affinity is greatly reduced by increasing the salt concentration; the gene 5 protein–fd DNA complex dissociates completely at 0.5 *M* NaCl or 0.1 *M* $MgCl_2$. Binding to short oligonucleotides shows both base specificity and oligonucleotide length dependence (*127, 129*), with binding affinity increasing markedly with oligonucleotide lattice length to at least the 8-mer level.

b. Melting of Double-Stranded DNA. Since gene 5 protein binds tightly to single-stranded DNA, but shows little or no affinity for double-stranded DNA under the same conditions, it should destabilize the duplex form of DNA at equilibrium. Melting profiles of DNA in the presence of excess gene 5 protein confirm this expectation; the T_m values of poly(dA-dT), *C. perfingens* DNA, and T4 DNA are all lowered by ~40° relative to those of the free DNA in 0.03 *M* KCl (*41*). Thus, unlike T4 gene 32 protein, fd gene 5 protein does not seem to be kinetically blocked with respect to melting duplex DNA. This DNA denaturation is relatively nonspecific with respect to base composition, but shows a slight preference for A-T rich DNA, which appears to be destabilized somewhat more than G-C rich DNA under the same conditions. The rate of DNA denaturation was found to be slow unless a small amount of Mg^{2+} is present (<10 m*M*), but higher concentrations of Mg^{2+} inhibit the reaction. Thus Mg^{2+} may play a special role in either the structure of the protein or of the protein–

129. J. E. Coleman, R. A. Anderson, R. Ratcliffe, and I. M. Armitage, *Biochemistry* **15**, 5419 (1976).

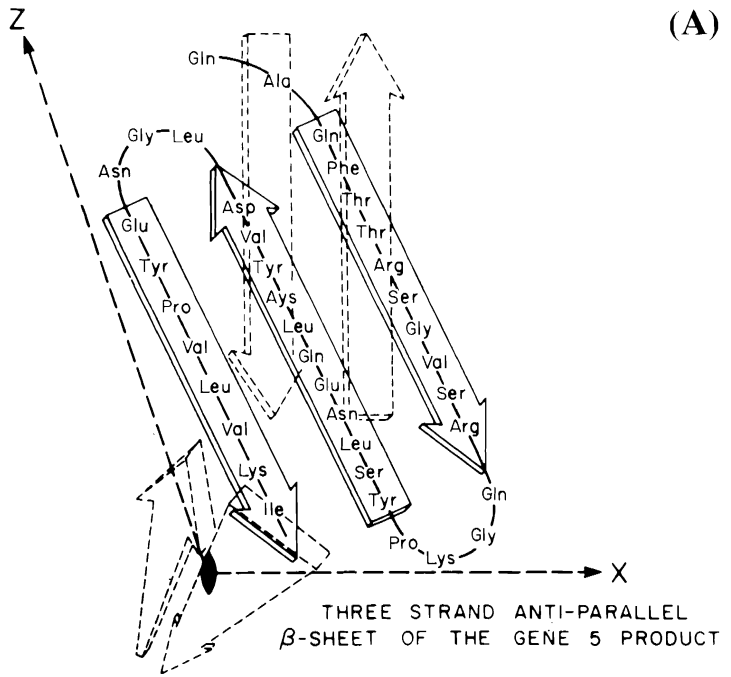


FIG. 5. (A) A schematic diagram that shows the three components of β-structure that comprise the major part of the gene 5 protein molecule. The amino acid residues forming the three-stranded sheet are indicated; these amino acids are primarily involved in the interaction with single-stranded DNA. (B) This schematic three-dimensional drawing, showing the course of the gene 5 polypeptide backbone, is taken directly from the electron density map. Domain (I) indicates the three-stranded β-sheet, which forms the major part of the DNA-binding interface. Region (II) shows the two strands of β-ribbon, which appear to be primarily responsible for maintaining the molecule as a dimer in solution by interaction with a symmetry-related β-ribbon. The second β-ribbon (III) is established diagonally across the β-sheet (I). This component may be the primary participant in the lateral interactions from which the cooperativity of the DNA binding arises. [After McPherson *et al.*, (*124*)].

DNA complex. Finally, in contrast to T4 gene 32 protein, no conditions could be found under which the gene 5 protein catalyzes the renaturation of denatured DNA (*41*); in this regard, gene 5 protein resembles the *E. coli* HDP.

c. Electron Microscopy. The gene 5 protein–fd DNA complex appears in the electron microscope as a rigid rod interrupted by occasional branched structures (*41*). These images contrast strikingly with those observed for complexes with DNA of T4 gene 32 protein, or of the *E. coli* or calf thymus HDPs, which all form expanded open circles with single-stranded fd DNA. The internucleotide spacing in the gene 5 protein–

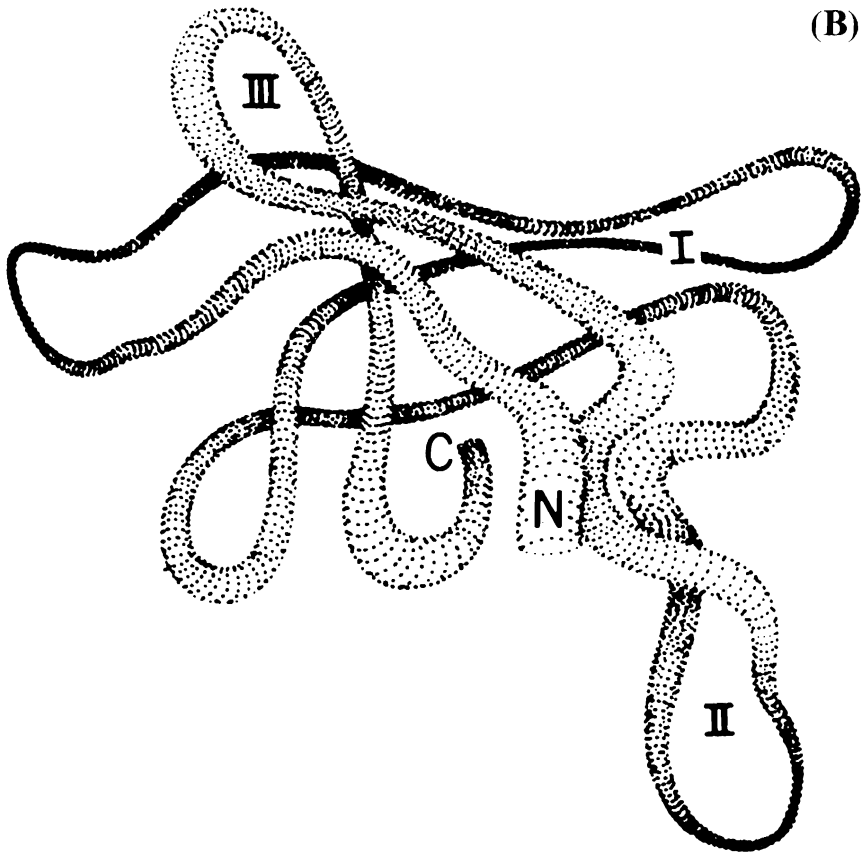


FIG. 5B.

fdDNA complex has been estimated to be at least 3.8 Å; in contrast the spacing for the T4 gene 32 protein–DNA complex is ~5.3 Å per nucleotide residue, and ~2.1 Å per residue for the *E. coli* SSB protein–DNA complex (*27*). The linear rodlike structure of the gene 5 protein–DNA complex seems to reflect the lateral association of two regions of protein-covered DNA, stabilized by the formation of back-to-back dimers. This complex, formed *in vitro* with fd DNA, is 100 Å wide and 7300 Å long, and is wound into an overall helix with a longitudinal repeat distance of 65–70 Å. Each turn of the helix contains at least six gene 5 protein dimers; we note that the asymmetric unit of gene 5 protein cocrystallized with oligonucleotides also contains 6 protein dimers (*126*).

Studies of complexes formed *in vivo,* however, show some differences from those formed *in vitro*. The *in vivo* complex consists of fibers ~40 Å in

width, which are supercoiled to an overall width of ~160 Å with a supercoil repeat of ~160 Å; the length of these complexes is ~1.1 μm (*130*). It has been observed that the binding stoichiometry of the complexes isolated *in vivo* differs from that of those formed *in vitro* (n = 4.7 and 4.0 nucleotide residues, respectively); this difference is not thought to be due to selective losses of protein in the isolation process (*35*). Also the *in vivo* complex appears to be more stable to salt dissociation. However, the addition of excess protein to the *in vivo* complex seems to result in the formation of the *in vitro* complex, as defined by sedimentation properties or stoichiometry. Whether these differences reflect isolation artifacts, are due to the unique ionic environment of the cell (e.g., critical concentrations of Mg^{2+}, spermidine, etc.), or represent different equilibrium or kinetic forms generated because of limited availability of protein during complex formation *in vivo,* remains to be seen.

d. Chemical Modification. The spectral results that appeared to implicate tyrosine residues in gene 5 protein–nucleic acid complex formation have been followed up by chemical modification studies (*127*). Reaction of gene 5 protein with tetranitromethane results in the nitration of three of the five tyrosine residues, and in a greatly reduced affinity of the modified protein for DNA (*127*). However all of the tyrosines are protected from the reagent if the protein is complexed with DNA prior to modification. The three tyrosine residues that are modified were identified as residues 26, 41, and 56. The presence of these residues on the surface of the protein has been confirmed by solvent perturbation (*35*) and by NMR studies (see next section).

Lysine residues also play a role in stabilizing the protein–nucleic acid complex, since reaction of the six lysine residues in gene 5 protein with *N*-acetylimidazole also abolishes the ability of the protein to bind DNA. Here, however, the residues are not protected against reaction by DNA binding (*127*).

Gene 5 protein also carries a single cysteine residue that is resistant to modification by DTNB unless the protein has been previously denatured (*35, 127*). However, this sulfhydryl is accessible to Hg^{2+}, and modification by this agent prevents binding of the protein to DNA. Conversely, complexation with DNA protects the sulfhydryl from modification (*127*).

Approximately 30% of the protein in an *in vivo* gene 5 protein–DNA complex can be covalently crosslinked to DNA by irradiation with UV light (*131, 132*). The covalent complex that results from irradiation of the

130. D. Pratt, P. Laws, and J. Griffith, *JMB* **82,** 425 (1974).

131. E. Anderson, Y. Nakashima, and W. H. Konigsberg, *Nucleic Acids Res.* **2,** 361 (1975).

132. L. Lica and D. S. Ray, *JMB* **115,** 45 (1977).

in vivo complex has been isolated, and the site of crosslinking has been shown to be between residues 70 and 77; it has been suggested that Ser-75, Gly-74, Phe-73, and Gly-71 might be the actual amino acid residues involved (*133*). In contrast, irradiation of the *in vitro* complex results in crosslinking at residue Cys-33, with an efficiency of crosslinking as high as 21% for poly(dT) (*131*). Whether this result represents another reflection of structural differences between *in vivo* and *in vitro* complexes is unclear.

e. NMR Studies. The role of aromatic residues in DNA binding has been extensively studied using NMR techniques, and on the basis of these results it has been suggested that intercalation of two tyrosines and one phenylalanine are involved in the recognition of the nucleic acid (*127, 129, 134–137*). ^{19}F-NMR has been used to probe the role of these protein residues in binding, ^{31}P-NMR has been employed to study the effects of complex formation on polynucleotide backbone structure, and ^{1}H-NMR has been utilized to investigate the structural changes in both the protein and the nucleic acid accompanying complex formation.

The fluorotyrosine derivative of gene 5 protein can be prepared by infection (with the wild-type bacteriophage) of *E. coli* tyrosine auxotrophs that have been grown in a medium containing *m*-fluorotyrosine (*127, 129*). This fluorotyrosine-substituted gene 5 protein has binding properties similar to those of native protein. The ^{19}F-NMR of this protein shows 5 resonances that correspond to the 5 tyrosine residues within the primary structure (*127, 129*). Three of the five resonances are located at spectral positions corresponding to free *m*-fluorotyrosine; the remaining two resonances are shifted downfield. The three upfield resonances have been assigned as *surface* residues (i.e., those accessible to nitration). This assignment is based on their spectral location, on their perturbation upon nucleic acid binding, and on the fact that they show a greater degree of rotational mobility than the other resonances as determined by the effect of proton decoupling on the intensity of the fluorine spectrum (nuclear Overhauser effect)(*135*). Based on similar reasoning, the two downfield resonances have been assigned as buried tyrosine residues.

Upon formation of a complex with either $d(pT)_4$ or $d(pA)_8$, the ^{19}F-resonances of fluorotyrosine gene 5 protein show two general characteris-

133. P. R. Paradiso, Y. Nakashima, and W. H. Konigsberg, *JBC* **254,** 4739 (1979).

134. G. J. Garssen, C. W. Hilbers, J. G. G. Schoenmakers, and J. van Boom, *EJB* **81,** 453 (1977).

135. J. E. Coleman and I. M. Armitage, *Biochemistry* **17,** 5038 (1978).

136. G. J. Garssen, R. Kaptein, J. G. G. Schoenmakers, and C. W. Hilbers, *PNAS* **75,** 5281 (1978).

137. G. J. Garssen, G. I. Tesser, J. G. G. Schoenmakers, and C. W. Hilbers, *BBA* **607,** 361 (1980).

tics: The downfield buried peaks shift slightly downfield, whereas the upfield surface resonances shift upfield (*135*). Two of the upfield peaks show chemical shifts upon complex formation, the magnitudes of the shifts depending on the base composition of the nucleic acid; a shift of ~0.4 ppm is observed with $d(pT)_4$ and a shift of ~0.7 ppm with $d(pA)_8$. This upfield shift is of the proper magnitude and direction to correspond to the nucleotide-base-induced ring current shifts that would be expected if intercalation of tyrosine residues between bases of the nucleic acid were taking place. It has therefore been suggested that two of the three tyrosine surface residues intercalate between nucleic acid bases upon complex formation (*129, 135*).

Studies involving proton NMR of gene 5 protein have confirmed the results based on ^{19}F-NMR, and have provided a more detailed picture of the intercalation mechanism (*129, 134–137*). These studies have shown that 30 to 40% of the aromatic protons of gene 5 protein show upfield shifts upon complex formation with oligonucleotides (*129*). The upfield shifts are on the order of 0.3 ppm for $d(pT)_8$, and ~0.8 ppm for $d(pA)_8$, which are similar to the values observed in the ^{19}F studies. Concommitantly with these changes, the C-6 proton resonances of the thymidine residues also shift upfield by 0.1 to 0.2 ppm. This observation is also consistent with intercalation, and could represent the expected upfield shift in the nucleotide base protons due to ring current effects from the aromatic residues of the protein.

Closer examination of the NMR proton spectra of gene 5 protein has shown that the 3,5-ring protons of the tyrosines are more affected by complex formation than the 2,6 protons (*134–137*), suggesting that only the leading edge of the tyrosine residues is inserted between the bases.

In addition to the effect on tyrosine protons that accompanies binding, it is clear that phenylalanine protons are also shifted upfield, and that these shifts are dependent on the base composition of the oligonucleotides of the complex. Thus, in addition to the two tyrosine residues, at least one phenylalanine is also involved in the intercalation process (*129, 137*).

Chemically induced dynamic nuclear polarization NMR has also been used to support the conclusions cited above on the basis of conventional NMR and chemical modification data (*136*). This technique has shown that tetranucleotide binding shields three tyrosine surface residues from reaction with a flavin dye. On the other hand, the dinucleotide $r(A)_2$ seems to protect only one residue when bound. Thus the decreased affinity of the shorter oligonucleotides (see above) may reflect the fact that the dinucleotides can interact with only one aromatic residue rather than with three.

Proton NMR has also been used to investigate the effects of nucleic acid binding on nonaromatic amino acid residues (*129*). The methyl protons of the ϵ-carbon of lysine, for example, show little chemical shift or line broadening upon complex formation, even though it is clear from chemical modification data that these groups are important in DNA binding. It has been concluded, since these groups show a large degree of rotational freedom on the NMR time scale, that they do not form rigid salt bridges with the phosphates of the nucleic acid, but are involved in a more delocalized binding interaction.

The resonances due to the arginine δ-CH_2 groups also show chemical shifts and/or broadening upon oligonucleotide binding (*129*). These changes may be due to a direct interaction of the amino acid residues with the phosphate backbone, but since the changes in the resonances differ if a tetranucleotide is used instead of an octanucleotide, these effects could also reflect a more general change in protein structure. This latter interpretation is consistent with the distinct changes that are observed in the aliphatic region of the spectrum and are dependent on the size of the oligonucleotide.

The structure of the nucleic acid backbone in the gene 5 protein complex has been studied using ^{31}P-NMR (*129, 134*). The diester phosphate resonances are shifted very little, indicating that gene 5 protein binding probably does not alter the conformation of the sugar-phosphate backbone appreciably. However, the fact that the resonances are broadened suggests that the phosphates are rigidly held in the complex. In addition, the apparent pK of the 5′-terminal phosphate is shifted in the complex, suggesting that the group is in close proximity to a positively charged amino acid residue of the protein (*134*).

f. Correlation of the Solution Results with the Protein Structure. The 30 Å cleft in the uncomplexed gene 5 protein structure, which has been proposed as the site of DNA binding, consists primarily of residues 12–49, 50–56, and 66–69 (*126*). The aromatic amino acids that are present in this region are the surface tyrosine residues 26, 41, and 56, as well as the buried Tyr-34, in addition, phenylalanine residues 13 and 68 are nearby. Of these residues only Tyr-56 and Phe-68 appear to point into the binding groove, but the other residues can be brought into the cleft by rotation about their respective β-carbon atoms.

All of these residues except Tyr-26 are clustered at one end of the binding cleft, and the side chains of Tyr-41, Tyr-34, and Phe-13 form a triple-stacked structure in which the residues are arranged in a fanned-out array. These tyrosines are probably located in the upfield region of the NMR spectrum, and may also be responsible for the protein circular

dichroism band at 228 nm, which has been attributed to Tyr–Tyr interactions. Binding to DNA results in a reduction of this signal by ~30%, perhaps reflecting an unstacking of the residues.

Cys-33 is also located on the inside of the DNA binding groove. This position is consistent with the fact that the bulky reagent, DTNB, is not able to penetrate to this residue, while Hg^{2+} can make contact. The fact that DNA binding prevents mercuration of Cys-33, and that this residue may be the site of UV-induced crosslinking to the nucleic acids (*133*), provides additional support for the identification of this region as the DNA binding site.

The distribution of the charged and uncharged residues within the binding groove is also noteworthy. Thus, while aromatic residues are located predominantly along the outside edge of the cleft, positively charged residues (particularly Arg-21, Arg-80 and Arg-82 and Lys-24 and Lys-46) are located within the groove. This suggests that the phosphate backbone of the nucleic acid is aligned and drawn into the binding region by these charged residues, and that subsequent stabilization of the complex may occur through interaction with the aromatic amino acids that rotate into position to react with the polynucleotide bases.

Final confirmation of this scheme must await direct high resolution X-ray data on crystals of gene 5 protein–nucleic acid complexes. The limited data presently available from such complexes indicate that the asymmetric unit consists of six gene 5 protein dimers. By slightly distorting the observed cylinderical arrangement into a helix, McPherson *et al.* (*126*) have proposed a structure for the gene 5 protein–DNA complex in which DNA strands bound to the gene 5 protein dimers are ~25 Å apart; the structure has a linear repeat of 80–90 Å and a diameter of ~100 Å. This proposal is in reasonable accord with some of the electron microscopic studies of gene 5 protein–DNA complexes (*41*).

3. *Biological Roles*

a. Replication. The main function of gene 5 protein in the life cycle of the filamentous phages is to control the switch from RF DNA replication to the replication of the single-stranded viral genome (*118, 138–140*). It has been clearly demonstrated, using phages that carry temperature-sensitive mutations in gene 5, that this protein is essential to the maintenance of the

138. B. J. Mazur and P. Model, *JMB* **78,** 285 (1973).
139. K. Geider, and A. Kornberg, *JBC* **249,** 3999 (1974).
140. A. Kornberg, "DNA Replication." Freeman, San Francisco, 1980.

production of single-stranded viral DNA (*118*); at nonpermissive temperatures with such mutants, single-stranded viral DNA is lost by continued synthesis of the double-stranded replicative form.

The free concentration of gene 5 protein in the cell may determine whether replicative form or single-stranded DNA is synthesized. This has been shown in phage in which DNA replication is blocked by a temperature-sensitive mutation in gene 2 (*138*). At nonpermissive temperatures gene 5 protein synthesis continues at a constant rate; on shifting to permissive temperatures only viral single-stranded DNA is accumulated, and no RF form can be detected.

In vitro experiments have also confirmed the *in vivo* observation that gene 5 protein inhibits the replication of viral single-stranded DNA (*139*). An *in vitro E. coli* replication system has been established that can convert viral DNA into RF DNA using purified DNA polymerase I and II, RNA polymerase, *E. coli* HDP, and DNA ligase. It was observed that purified M13 gene 5 protein could not substitute for the *E. coli* HDP; in fact the addition of gene 5 protein virtually abolished the reaction when present in excess of one protein per 4 nucleotide residues. Only at very low levels did gene 5 protein show a positive effect; under these conditions the protein appeared to stimulate the synthesis of RNA-primed M13 DNA.

It is clear that although the fd gene 5 product is a single-stranded DNA binding protein, its physiological role is very different from that of T4, T7, and *E. coli* (see the following sections) HDPs. The gene 5 protein plays a special role in the life cycle of the filamentous phages; unlike the HDPs previously listed, it acts as a specific switch in directing the DNA replication of this phage. In addition, it may be responsible for "prepackaging" the newly synthesized viral DNA and protecting it from nucleases until it reaches the cell membrane, at which point the gene 5 protein is displaced by the fd gene 8 coat protein.

C. *Escherichia coli* SINGLE-STRANDED DNA BINDING PROTEIN

In an attempt to determine whether a protein analogous to the T4-coded gene 32 protein exists in uninfected *E. coli* cells, Sigal *et al.* (*96*) chromotographed crude protein extracts of *E. coli* on denatured DNA cellulose columns. A single-stranded DNA binding protein with a monomeric molecular weight of about 20,000 was discovered and purified to homogeneity (see Section III). Subsequently this protein has been

shown to be necessary for DNA replication (*96, 141–143*), and has been implicated in recombination and repair (*144–148*). About 300–800 copies are present per bacterial cell (*96, 142*). This protein has unfortunately been called by several different names, including "DNA unwinding protein" [Chapter 18, this volume, and Ref. (*96*)], "DNA binding protein" (*149*), "*E. coli* helix destabilizing protein I" (*1*), and "*E. coli* single-strand binding protein (SSB)" (*148*). We call this protein the *E. coli* SSB protein in accord with the nomenclature that has been used by the authors of the more recent papers, and in consideration of the fact that the gene for this protein has now been formally named *ssb* (*150*).

1. *Physical Properties of the Protein*

a. Size and Structural Features. The *E. coli* SSB protein in its native form exists as a tetramer of four identical subunits (*151*). Each monomeric subunit has a molecular weight of about 20,000, with exact estimates (based on sodium dodecyl sulfate gel electrophoresis) ranging from 18,500 (*142*) to 22,000 (*27, 151*). From amino acid composition the molecular weight is estimated at 19,500 (*142*). The sedimentation coefficient of the native protein is ~4.8 S (*96, 142, 152*) corresponding to a tetramer of total molecular weight ~76,000.

The *C. coli* SSB protein (*27, 73, 142*) contains a rather large number of acidic residues. The isoelectric point was determined to be 6.0 ± 0.1 (*142*). Based on circular dichroism spectra, it has been estimated that *E. coli* SSB protein is about 20% α-helix and 20% β-sheet (*27*).

b. Spectroscopic Properties. The molar extinction coefficient (at 280 nm) for *E. coli* SSB, based on a molecular weight of 20,000 is 3.0×10^4 $M^{-1}\,cm^{-1}$ (*153*). The protein displays a fluorescence spectrum characteristic of many tryptophan-containing proteins; the wavelength of maximum excitation is 285 nm and the wavelength of maximum emission is 345 nm.

141. T. A. Kunkel, R. R. Meyer, and L. A. Loeb, *PNAS* **76,** 6331 (1979).
142. J. H. Weiner, L. L. Bertsch, and A. Kornberg, *JBC* **250,** 1972 (1975).
143. J. F. Scott, S. Eisenberg, L. L. Bertsch, and A. Kornberg, *PNAS* **73,** 1594 (1977).
144. J. Tomizawa, N. Anraku, and T. Iwama, *JMB* **21,** 247 (1966).
145. M. W. Baldy, *Virology* **40,** 272 (1970).
146. B. F. Johnson, *Mol. Gen. Genet.* **157,** 91 (1977).
147. J. Glassberg, R. R. Meyer, and A. Kornberg, *J. Bacteriol.* **140,** 14 (1979).
148. R. R. Meyer, J. Glassberg, and A. Kornberg, *PNAS* **76,** 1702 (1979).
149. I. J. Molineux and M. L. Gefter, *PNAS* **71,** 3858 (1974).
150. B. J. Bachmann and K. B. Low, *Microbiol. Rev.* **44,** 1 (1980).
151. I. J. Molineux, S. Friedman, and M. L. Gefter, *JBC* **249,** 6090 (1974).
152. W. T. Ruyechan and J. G. Wetmur, *Biochemistry* **15,** 5057 (1976).
153. W. T. Ruyechan and J. G. Wetmur, *Biochemistry* **14,** 5529–5534 (1975).

The binding of polynucleotides to *E. coli* results in up to 70% quenching of intrinsic protein fluorescence, unaccompanied by any shift in the wavelength of maximum emission (*36*).

c. Aggregation State and Heat Stability. The SSB protein exists as a tetramer within the protein concentration range of 75–750 μg/ml (*151*) at ionic strengths as high as 1.0 *M* (*152*). The *E. coli* SSB protein also appears to be remarkably heat stable. It has been reported that the protein loses none of its known activities as a consequence of heating to 100° for 2 minutes (*142*). This property has been exploited to purify the protein away from other proteins that precipitate upon heating. Even a temperature-sensitive mutant of the SSB protein is heat resistant and will function normally when cooled from 100° to the permissive temperature (*148*). However, preliminary differential scanning calorimetry studies have indicated that the wild-type protein undergoes an irreversible transition at about 70° when left at high temperature for an extended period of time (*73*). It is therefore important to exercise care in the use of heating in preparative protocols for the SSB protein.

2. *Protein–Nucleic Acid Interactions*

The binding of *E. coli* SSB protein shares many of the characteristics of the T4 gene 32 protein interaction with DNA; most importantly, like gene 32 protein, and SSB protein binds preferentially and cooperatively to single-stranded nucleic acids (*96*). In contrast to gene 32 protein, however, the SSB protein can, under some circumstances, denature double-stranded DNA. In addition, unlike gene 32 protein, which binds to the DNA as a monomer (and fd gene 5 protein, which seems to bind as a dimer), the SSB protein probably binds as a tetramer.

a. Binding to Single-Stranded Nucleic Acids. It has been demonstrated that SSB protein displays a large preference for single-stranded DNA, and shows no detectable binding to native duplex DNA or bacteriophage R17 RNA (*36, 96, 142*).

Chemical modification studies (*154*) have strongly implicated lysine residues in the binding of SSB protein to single-stranded DNA, and indeed, in common with the other proteins discussed in this review, a significant salt dependent of the interaction is seen. SSB protein elutes from denatured DNA-cellulose columns at between 1 and 2 *M* NaCl (*96*). A number of groups have reported that the affinity of the protein for both poly- and oligonucleotides decreases significantly as the ionic strength increases above 0.1–0.2 *M* in NaCl (*36, 142, 151–153*). Chemical modification studies

154. P. K. Bandyopadhyay and C. W. Wu, *Biochemistry* **17**, 4078 (1978).

with reagents specific for cysteine, arginine, and tyrosine seem to rule out the significant participation of these residues in the interaction of the protein with polynucleotides (*154*).

In addition to (monovalent salt concentration, both pH and Mg^{2+} concentrations also affect the SSB protein-nucleic acid interaction. Binding is strongest between pH 7.5 and 8.5, and the strength of the interaction appears to fall off rapidly below pH 6.5 and above pH 9.0 (*152*). Mg^{2+} concentrations above 10 m*M* decrease significantly the affinity of SSB protein for single-stranded oligo- and polynucleotides; thus most *in vitro* replication assays are carried out in the 1–10 m*M* Mg^{2+} concentration range. One should be careful to distinguish this effect of Mg^{2+} from the Mg^{2+}-induced stabilization of secondary structure in single-stranded DNA and RNA. In the latter case a decrease in the binding of the protein to nucleic acid occurs as a consequence of the stabilization of competing hairpin (duplex) structures, and this decreases the concentration of single-stranded nucleic acid available for binding.

The effect of *E. coli* SSB protein on base stacking interactions in single-stranded DNA has been examined by circular dichroism (*27*), sedimentation (*96*), and electron microscopic techniques (*96*). Circular dichroism spectra indicate that the protein unstacks the bases in a manner similar to that due to thermal denaturation, and density gradient centrifugation analyses suggest that the DNA is also held in a more open conformation, thus increasing the length of the polynucleotide backbone. However, electron microscopy studies of glutaraldehyde-fixed complexes are consistent with a 35% decrease in internucleotide spacing. This is to be contrasted to a 50% increase in length for DNA complexed to gene 32 protein (*79*); EM artifacts of fixation or spreading may account for this discrepancy.

b. Binding Specificity. The interaction of *E. coli* SSB protein with synthetic polynucleotides has been studied to look for base and sugar composition-based differences in binding affinity. Gel filtration studies (*143*) lead to the following order of SSB protein-polynucleotide affinities: poly(dT) > ssDNA (ϕX174) > poly(rU) > poly(dA) > dsDNA (T4) > poly(rA). By monitoring quenching of protein fluorescence, a somewhat different series was obtained: poly(dT), poly(dA) > ssDNA (fd) > poly(rU) $\cong$ poly(dC) > poly(rA) > $tRNA^{gln}$ > poly(rC) $\gg$ dsDNA (P22) (*36*). Obviously general preferences can be seen for single-stranded over double-stranded DNA, and for deoxyribose-containing over ribose-containing polynucleotides. Also poly(dT) seems to be bound more tightly than all other polynucleotides; these general features are also seen with gene 32 protein. Little, if any, nucleotide composition or sequence dependence is observed for the binding of hexanucleotides to SSB protein.

c. Dependence of Binding Affinity on Oligonucleotide Lattice Length. Equilibrium dialysis experiments (*152*) have suggested that the apparent binding affinity of SSB protein for various multimers of d(pCpT) depends on the length of the overall oligonucleotide lattice. The apparent binding constant to such oligomers increased ~10-fold in going from $d(pCpT)_2$ to $d(pCpT)_3$, and only ~2-fold further in going from $d(pCpT)_3$ to $d(pCpT)_{6-9}$. In partial contrast to these results, intrinsic protein fluorescence-quenching studies have shown that $d(pT)_{16}$ binds SSB protein ~200-fold more strongly than dT-containing oligomers 8 residues in length or less (*36*). The longer oligomers quench protein fluorescence much more effectively than the shorter lattices. These results are also in general accord with those observed with gene 32 protein.

d. Polynucleotide Binding Site Size and the Oligomeric State of SSB Protein. A variety of techniques have been used to determine the site size (n) of SSB protein for binding to polynucleotides; most approaches have yielded values of n of ~8 nucleotide residues per protein monomer (*36, 96, 143*). The molecular significance of this parameter has been hard to define because neither the state of aggregation of the protein in its binding form, nor the number of binding sites utilized per bound protein, have been established unequivocally. EM studies have suggested that the protein binds to single-stranded DNA as a tetramer, or alternatively, that it aggregates to tetramers on binding (*96*). It was also demonstrated by sedimentation experiments with oligo $d(pCpT)_{6-9}$ that no change in the state of aggregation occurs upon binding (*152*). This result has been confirmed by measurements of the rotational correlation time of the free protein and the complex by time-dependent emission anisotropy; no significant change in the protein structure upon binding to oligo$(dT)_{16}$ was revealed (*154*). In light of these results, a site size of about 32 nucleotides per tetramer may be a more physically significant way to view the binding of SSB protein to single-stranded DNA. More detailed work on the topography of this complex is obviously required.

e. Binding Cooperativity. The cooperative nature of the binding of *E. coli* SSB protein to single-stranded DNA and to synthetic polynucleotides has been demonstrated by several different techniques, including electron microscopy (*96, 153*), gel filtration (*142*), density gradient centrifugation (*96*), and fluorescence quenching (*36*). By comparing the binding constants for the interaction of SSB protein with pT_8 and pT_{16}, it was estimated that the value of the cooperativity parameter is at least 50 on the basis of fluorescence measurements (*36*), whereas on the basis of electron microscopic studies, ω was estimated to be $\sim 10^5$ (*153*). The estimate based on fluorescence is only a lower limit; however, the value obtained by electron microscopy may be high due to artifacts in the technique. Thus

the use of glutaraldehyde in the fixation of the complexes prior to spreading could have artificially increased the size of the protein clusters beyond equilibrium expectations. Such artifacts have been seen in EM estimates of cooperativity for gene 32 protein (*79*). An accurate measure of ω for SSB protein, obtained under physiological conditions, is not available. In addition, preliminary evidence (*153*) suggests that, unlike the situation with gene 32 protein, ω for SSB protein may be salt concentration dependent.

f. Denaturation and Renaturation of Duplex DNA. As noted above, *E. coli* SSB protein at low ionic strength can slowly denature duplex DNA (*96*). Surprisingly, this effect has not been heavily investigated, in contrast to the denaturation induced by the gene 32 protein *I fragment. Upon addition of *E. coli* SSB protein, it was found that denaturation of duplex DNA goes to completion in 12 to 30 min at 37° for duplex DNA; the DNA has an unperturbed T_m of 55° under the same solvent conditions (*96*). The effect is stoichiometric rather than catalytic; however, the quantitative aspects of the stoichiometry have not been characterized. In electron microscopic studies involving glutaraldehyde-fixed protein–DNA complexes it was found that the *E. coli* SSB protein attacks A-T rich regions, and tends to expand denatured loops, rather than to initiate new ones. The addition of Mg^{2+} completely reverses the denaturation effect.

Several groups have reported little or no affinity of *E. coli* SSB protein for double-stranded DNA. Also the addition of 5 m*M* spermidine to denatured DNA strongly decreases the affinity of the protein, presumably by stabilizing the secondary structure of the nucleic acid. It has also been demonstrated that *E. coli* SSB protein does not retain duplex DNA on nitrocellulose filters (*142*). The binding affinity of SSB protein to duplex DNA has been estimated to be at least 3 orders of magnitude weaker than that to denatured DNA (*36*).

These data suggest that, like the carboxy-terminal cleaved gene 32 protein (G32P*I), *E. coli* SSB protein binds to and traps sequences of DNA transiently opened by thermal fluctuations. Since A-T rich sequences are less stable with respect to thermal melting than those rich in G-C base pairs, the former are presumed to be prime candidates for such nucleating interactions. The protein, due to its significant cooperativity, would then tend to bind at contiguous sites and expand the denatured regions. The biological significance of such SSB protein-driven melting processes is questionable, due to the sensitivity of melting to the concentration of mono- and polyvalent ions. We return to this question in the following sections.

It has been reported that under certain conditions *E. coli* SSB protein

can catalyze the *renaturation* of denatured DNA (*155*). In the presence of NaCl (0.2 *M* or less), and in the presence or absence of $MgCl_2$, *E. coli* SSB protein does not significantly affect the rate of renaturation of λ DNA at pH 7.0. However, in the presence of 2 m*M* spermidine or spermine the renaturation rate is enhanced by a factor of about 5000. This rate-enhancement effect requires saturating amounts of SSB protein, and its efficiency increases with the length of the DNA. Because of the complex nature of the dependence on counterions and pH, the simple removal of intrastrand hairpin loops by SSB protein binding does not appear to fully explain this effect.

3. *Biological Roles*

a. Replication. For several years after the initial discovery of *E. coli* SSB protein, the effects of this protein on the *in vitro* replication of viral DNA by various viral and host-coded polymerases has been heavily investigated. However, it was not until a temperature-sensitive mutant was isolated (*148*) and found to be defective in DNA replication that one could be sure of the importance of *E. coli* SSB protein in phage and cellular metabolism *in vivo*. The mutation (designated *ssb*-1) has been localized at 90 to 91 minutes on the *E. coli* linkage map.

In vitro studies have demonstrated that the *E. coli* SSB protein specifically stimulates polymerase II-directed DNA synthesis on various phage templates (*96, 150*). This stimulation appears to be dependent on the ratio of binding protein to DNA, and independent of polymerase concentration, suggesting that the primary effect is at the DNA level (*151*).

The addition of *E. coli* SSB protein to *in vitro* replication systems has a multitude of effects, including (i) stimulation of the initiation of RNA primer-directed DNA synthesis by polymerase III holoenzyme on single-stranded phage DNA templates (*139, 142*), (ii) stimulation of the elongation rate of DNA synthesis by pol III (*142*), (iii) increase in the fidelity of DNA synthesis by pol III (*141*), and (iv) stimulation of the initiation and elongation of DNA synthesis by pol II on gapped and single-stranded templates (*96, 150, 156*). The protein is required in an *in vitro* polymerase III replication mix in order to convert bacteriophage G4 single-stranded DNA into its replicative form (*142*); and also to convert the replicative form back to single-stranded circular DNA (*143*).

All these activities may be explained by the ability of SSB protein to bind to single-stranded DNA, trapping the DNA in the open form and

155. C. Christiansen and R. L. Baldwin, *JMB* **115,** 441 (1977).
156. I. J. Molineux and M. L. Gefter, *JMB* **98,** 811 (1975).

melting out double-stranded regions by contiguous cooperative binding. This makes the bases more accessible as templates for polymerases, and thus may increase the fidelity of DNA synthesis as well as the elongation rate. However, there is ample evidence to suggest that *E. coli* SSB protein also interacts specifically with polymerases II and III. The SSB protein also inhibits the exonuclease activity of pol I and the T4 DNA polymerase, and the activity of the single-stranded DNases from *Aspergillus* and *Neurospera crassa*. In contrast it does *not* inhibit, and even stimulates to a small extent, the exonuclease activity of T7 DNA polymerase, pol II, and exonuclease I (*156*). Both pol II and exonuclease I form complexes with SSB protein. While no complex appears to be formed between SSB protein and T7 polymerase, it has been shown that SSB protein can substitute for T7 DNA binding protein (*157*) in a T7 replication system. Other DNA binding proteins, such as fd gene 5, T7 DNA binding protein, and T4 gene 32 protein, cannot substitute for SSB protein in stimulating pol II (*157*). Thus the interaction of *E. coli* SSB protein with pol II is probably physiologically important, whereas the role of the DNA binding protein in the T7 replication system may be to interact primarily with the DNA. Demonstration of meaningful specific interactions with pol III must await a complete characterization of the holoenzyme [see Chapter 3, this volume and Ref. (*140*) for a review of this aspect].

Antibody titration studies have shown that there about 300 copies of *E. coli* SSB protein tetramer per log phase bacterial cell (*142*). Based on the binding information previously presented, and the fact that there are approximately six replication forks per cell (*96*), we can calculate that ~1600 nucleotides of DNA are covered per replication fork.

Both plasmids and transducing phages carrying the *ssb* gene have been used to overproduce SSB protein (*42*). Thus it is unlikely that the synthesis of this protein is autogenously regulated *in vivo*.

b. Recombination and Repair. Further characterization of the *ssb*-1 mutant (*148*) has revealed that the mutant strain is about one-fifth as active in recombination as the wild type, and extremely sensitive to UV damage. The *lexC* gene proposed as a regulator of UV and X-ray inducible repair (*146, 158*) has been tentatively found to be allelic with *ssb*. Both genes map at the same locus, and SSB protein from a temperature-sensitive *lexC* mutant is temperature-sensitive when tested as an accessory protein in bacteriophage G4 replication. While an exact role for the *E. coli* SSB protein in recombination and repair has yet to be established, it has been demonstrated that SSB protein catalyzes *recA*-mediated single-stranded DNA

157. R. C. Reuben and M. L. Gefter, *JBC* **249,** 3843 (1974).

158. J. Greenberg, L. J. Berends, J. Donch, M. H. L. Green, *Genet. Res. (Cambridge)* **23,** 175 (1974).

assimilation into homologous double-stranded DNA *in vitro* (*159, 160*). The concentration of *recA* protein required for such D-loop formation is also reduced in the presence of SSB protein.

D. Bacteriophage T7-Coded Single-Stranded DNA Binding Protein

1. *Protein Properties and Interactions with DNA*

In searching for an analogue of T4-coded gene 32 protein that might serve similarly in phage T7 replication, two groups independently discovered a T7-coded, single-stranded DNA binding protein that stimulates replication by T7 DNA polymerase on either single-stranded or gapped-duplex DNA templates (*161, 162*). Although the original estimates of molecular weight of the two proteins differ considerably [31,000 (*161*) versus 25,000, (*162*)], the proteins appear to be the same on the basis of their other properties. Like gene 32 protein, the T7 HDP does not bind to duplex DNA and carries a net negative charge at neutral pH. NaCl concentrations in excess of 0.25 *M* are required to elute this protein from DEAE-Sephadex; under the same conditions *E. coli* SSB protein (p*I* = 6.0) is eluted at ~0.15 *M* NaCl. The protein appears to be monomeric in 2 *M* NaCl, as judged by its behavior on gel filtration columns.

The T7 HDP denatures duplex poly(dA-dT) (*162*). We estimate, on the basis of incomplete data (*162*), that the T7 protein lowers the T_m of poly(dA-dT) by about 40° in 0.04 *M* NaCl; this ΔT_m is very comparable to that induced by T4 gene 32 protein under the same conditions. The T7 protein also stimulates the exonuclease activity of T7 DNA polymerase (the T7 gene 5 protein–*E. coli* thioredoxin complex) on a duplex DNA template (*163*).

These rather scanty data suggest that the nucleic acid binding properties of the T7 protein may be quite similar to those of the other prokaryotic single-stranded DNA binding proteins described in the preceding sections.

2. *Biological Roles*

The biological properties of the T7 single-stranded binding protein also resemble those of the analogous T4 and *E. coli* proteins. The T7 protein

159. K. McEntee, G. M. Weinstock, and I. R. Lehman, *PNAS* **77**, 857 (1980).
160. T. Shibata, C. DasGupta, R. P. Cunningham, and C. M. Radding, *PNAS* **77**, 2606 (1980).
161. R. C. Reuben and M. L. Gefter, *PNAS* **70**, 1846 (1973).
162. E. Scherzinger, F. Litfin, and E. Jost, *Molec. Gen. Genet.* **123**, 247 (1973).
163. K. Hori, D. F. Mark, and C. C. Richardson, *JBC* **254**, 11598 (1979).

stimulates the polymerization activity of T7 DNA polymerase on single-stranded templates 10- to 15-fold at low temperatures; under these conditions the T7 DNA polymerase is otherwise quite ineffective (*161*). Presumably this stimulation reflects the denaturation of DNA hairpins, which are stable under these conditions, and which, if present, inhibit the polymerase. The T7 protein does not, however, stimulates T4 DNA polymerase or *E. coli* DNA polymerases I, II, or III (*161*). In spite of this demonstration of functional specificity, no physical interaction between the T7 DNA binding protein and T7 polymerase has been observed (*162*). In contrast, the T4 gene 32 protein and the *E. coli* SSB protein *do* bind preferentially to their homologous polymerases in free solution.

In vitro, E. coli SSB protein can substitute for the T7 protein with regard to its stimulatory effect on the activity of T7 DNA polymerase (*162*). Thus, despite the fact that the T7 SSB protein seems to be specific (i.e., it can only stimulate its homologous DNA polymerase), the T7 DNA polymerase can be stimulated by both *E. coli* and T7 SSB proteins. This may also explain why mutations in T7 SSB protein are not lethal to T7 growth; presumably the host SSB protein substitutes for the phage protein, thus "rescuing" the mutant phage.

E. Eukaryotic Single-Stranded DNA Binding Proteins

A large number of proteins that show affinity for either single-stranded or double-stranded DNA have been isolated from a variety of eukaryotic organisms [for reviews, see Refs. (*164–166*)]. In general, these proteins have been isolated from cell extracts on the basis of binding to DNA-cellulose columns; in the absence of genetic information and mutants the physiological role of many of these proteins has been difficult to assess.

The proteins that are included in the following sections have been selected for discussion because they show (i) binding to DNA, (ii) a preferential affinity for single-stranded over double-stranded DNA (thus they are, at least potentially, HDPs), and (iii) a presumed biological activity (usually manifested as a stimulatory effect in an *in vitro* replication system with the homolgous DNA polymerase). As a consequence many interesting proteins have been omitted, but we hope that those discussed will prove to be representative of this potentially important class. Some of the more important properties of these proteins are summarized in Table III.

164. J. E. Coleman and J. L. Oakly, *Crit. Rev. Biochem.* **7**, 247 (1980).
165. J. J. Champoux, *Annu. Rev. Biochem.* **47**, 449 (1978).
166. A. Falaschi, F. Cobianchi, and S. Riva, *Trends Biochem. Sci.* **5**, 154 (1980).

1. *The Lilium (Lily) DNA Binding Protein*

A unique DNA binding protein has been isolated from the meiotic cells of lily plants. In contrast to the other prokaryotic and eukaryotic DNA binding proteins discussed here, this protein is primarily involved in meiotic recombination, rather than in DNA replication (*167–169*). This protein, referred to as R-protein, is synthesized during meiotic prophase and is localized within the nucleus. The fact that it is only present (and active) in germ cells during the portion of the cell cycle corresponding to meiosis strongly suggests that it is somehow involved in the processes of chromosome pairing and recombination (*168*).

The isolated protein has a molecular weight of ~35,000, and forms a stable complex with single-stranded DNA even at 2.0 *M* NaCl. Minimal concentrations (5 m*M*) of Mg^{2+} or Ca^{2+} are absolutely required for binding (*168*). Binding is specific for single-stranded DNA, with no detectable binding to RNA. Furthermore, like T4 gene 32 protein and *E. coli* SSB protein, this protein catalyzes the renaturation of denatured DNA. As previously noted, the ability to catalyze the renaturation of denatured DNA might be quite important in recombination, and is thus consistent with the proposed role of this protein in meiosis.

The lily protein can be phosphorylated by a specific cAMP-dependent protein kinase, and the level to which it is phosphorylated determines its *in vitro* properties (*169*). The native (phosphorylated) protein exhibits a definite preference for binding to single-stranded DNA, and although the dephosphorylated protein has an increased affinity to single-stranded DNA, it shows an *even greater* increased affinity for double-stranded DNA. The properties of the dephosphorylated protein revert to those of the native protein by treatment with the protein kinase, which adds two phosphate groups per protein monomer. The native protein stimulates both duplex DNA denaturation and denatured DNA renaturation, but the dephosphorylated protein shows neither of these activities.

A very similar protein has been isolated from rat spermatocytes (*170*). The DNA binding properties and renaturation activity of the rat protein are also modulated by kinase-driven phosphorylation-dephosphorylation activities. In addition a DNA binding protein of mouse acites cells has been shown to vary in its stimulatory effect on DNA replication, depending on its level of phosphorylation. These results suggest that control of the level of phosphorylation of DNA binding proteins may serve as a

167. Y. Hotta and H. Stern, *Dev. Biol.* **26,** 87 (1971).
168. Y. Hotta and H. Stern, *Nature (London) New Biol.* **234,** 83 (1971).
169. Y. Hotta and H. Stern, *EJB* **95,** 31 (1979).
170. J. Mather and Y. Hotta, *Exp. Cell Res.* **109,** 181 (1977).

general mechanism to modulate the activity of these proteins during the cell cycle in eukaryotes [for a review of eukaryotic protein phosphorylation, see Ref. (*171*)].

2. *Calf Thymus DNA Binding Proteins*

Three DNA binding proteins have been isolated from calf thymus by Herrick and Alberts (*49, 172, 173*) using a general purification protocol designed by these workers for the isolation of eukaryotic HDPs (*49*). The proteins isolated are UP1 [UP for unwinding protein; by current nomenclature this protein would be calf thymus (CT) HDP-I]; a "high salt-eluting fraction" (CT HDP-II), and a "low salt-eluting fraction".

CT HDP-I has a molecular weight of 24,000 and is present in the thymus at ~800,000 copies per cell (*49*). It exists as a monomer in solution, and has an isoelectric point near neutrality. Isoelectric focusing shows that it is composed of four or five subspecies, probably reflecting intrinsic heterogeneity or limited protease action during isolation. These fractions show different affinities for single-stranded DNA; the most acidic fraction binds most weakly. This protein has a marked preference for single-stranded over double-stranded DNA, as demonstrated by its ability to (reversibly) depress the T_m of poly(dA-dT), poly(rA-rU), and *C. perfringens* duplex DNA (*172*). This helix destabilization effect is greatest for poly(dA-dT), less for poly(rA-rU), and least for the *C. perfringens* DNA, suggesting that the protein may have some base- and sugar-binding specificity. From these studies it has been estimated that the affinity for single-stranded DNA is 1500-fold greater than for double-stranded DNA. The acidic subfraction of CT HDP-I shows a much smaller ΔT_m, as might be expected from the smaller affinity of this protein for single-stranded DNA cellulose.

Sedimentation studies (*173*) and optical studies (*174*) have shown that a stoichiometric complex is formed at a ratio of 7 DNA nucleotide residues per protein (CT HDP-I) monomer. The formation of this complex is very dependent on salt concentration, and the data suggest that up to 6 ionic interactions per protein monomer may be involved (*174*). The T_m depression of duplex DNA induced by this protein is salt-dependent, decreasing with increased NaCl concentration (*172*). Although the protein shows a high affinity for single-stranded DNA, unlike the prokaryotic HDPs it does not bind cooperatively (*173*).

Electron microscopy shows an extended DNA complex in which the

171. C. S. Rubin and O. M. Rosen, *Annu. Rev. Biochem.* **44,** 831 (1975).
172. G. Herrick and B. M. Alberts, *JBC* **251,** 2133 (1976).
173. G. Herrick, H. Delius, and B. M. Alberts, *JBC* **251,** 2142 (1976).
174. R. L. Karpel and A. C. Burchard, *Biochemistry* **19,** 4674 (1980).

contour length increases by ~17% (*173*); the comparable increase in length for DNA complexed with T4 gene 32 protein is 46%. Also, as expected, CT HDP-I (and CT HDP-II) produce denaturation "bubbles" in supercoiled SV40 DNA. In addition changes in the polynucleotide circular dichroism and UV absorbance spectra indicate considerable unstacking of the bases (*174*). The accessibility of nucleotide bases in this protein-nucleic acid complex was probed by chemical modification and hydrogen exchange techniques. The results suggest that the bases are as exposed in the complex as in free DNA, but that they are also unstacked in the complex (*175*).

The CT HDP-II preparation contains several protein fractions; the most prominent has a molecular weight of ~33,000 and an isoelectric point (pI) of 5.2–5.6. This protein is very similar to CT HDP-I in that (i) it binds noncooperatively to a single-stranded DNA at a nucleotide residue to protein stoichiometry of ~10 : 1, (ii) it strongly depresses the T_m of poly(dA-dT) and poly(rA-rU), and (iii) it forms an extended complex with fd DNA (*49, 172*).

In contrast, the low-salt fraction (molecular weight ~33,000) does not denature native DNA and does not form a stable complex visualizable in the EM. In addition, this fraction does not stimulate calf thymus DNA polymerase activity.

While the physiological role of these proteins has not been defined, both CT HDP-I and -II stimulate the activity of the calf thymus DNA polymerase α, but not that of polymerase β (*173*). T4 gene 32 protein does not stimulate polymerase α activity; thus stimulation by the calf thymus proteins may be specific. The amount by which this polymerase is stimulated by CT HDP-I depends on the type of DNA template used, and ranges from a 10-fold stimulation on exonuclease-treated DNA to less than a twofold effect on heat-denatured DNA. As found with T4 gene 32 protein, excess CT HDP-I *inhibits* DNA synthesis. At optimal concentrations both CT HDP-I and -II stimulate calf thymus polymerase α activity on oligo(dG) primed-poly(dC) templates more than 5-fold.

CT HDP-I can also catalyze the renaturation of tRNA and 5 S RNA to their active (native) forms (*176*). This renaturation occurs despite the fact that CT HDP-I will not renature denatured DNA; this difference may reflect the fact that the tRNA renaturation is an intramolecular process

175. T. Kohwi-Shigematsu, T. Enomoto, M. Yamada, M. Nakanishi, and M. Tsuboi, *PNAS* **75**, 4689 (1978).

176. R. L. Karpel, N. S. Miller, and J. R. Fresco, *in* "Molecular Mechanisms in the Control of Gene Expression" (D. P. Nierlich, W. J. Rutter, and C. F. Fox, eds.), p. 411. Academic Press, New York, 1976.

and DNA renaturation is intermolecular (*172*). In addition, CT HDP-I seems to be located primarily in the cytoplasm of the cell, suggesting that it may be involved in RNA manipulation (*176*).

By using the complementary approach of isolating a protein that stimulates the activity of polymerase α, and then comparing it to the proteins that have been previously described, a protein very similar to the most basic component of CT HDP-I has been purified (*177*). Its binding is specific for single-stranded DNA, and it stimulates the activity of DNA polymerase α up to 8-fold depending on the template, whereas polymerase β is stimulated less than 2-fold.

3. *Ustilago maydis DNA Binding Protein*

An HDP has been isolated from mitotic cells of the basidiomycete fungus, *U. maydis*. This protein may be involved in DNA replication in this organism in that it stimulates *U. maydis* DNA polymerase (*178, 179*). The protein has a molecular weight of ~20,000, and exists in solution as a monomer. Filter binding experiments show that it binds tightly to single-stranded DNA, but not to double-stranded DNA. The T_m of poly(dA-dT) is lowered by ~50° in the presence of saturating amounts of this protein, and Mg^{2+} was found to increase the rate of renaturation. As with T4 gene 32 protein and the lily R-protein, the *U. maydis* HDP also catalyzes the renaturation of denatured DNA (*178*). The protein stimulates by ~ two-fold the activity on denatured DNA of the only DNA polymerase that has been isolated from *U. maydis*. As with many of the other HDPs, excess binding protein inhibits the activity of the polymerase. Since the *U. maydis* HDP does not stimulate the activity of polymerases from *E. coli*, *M. luteus*, T4, or T7, the stimulation of the *U. maydis* polymerase may be specific. However, no specific interaction of polymerase with binding protein could be detected. In addition to the stimulation of polymerase activity, the *U. maydis* HDP in stoichiometric excess also inhibits the nucleolytic digestion of DNA.

The stimulatory effect of this protein on the *U. maydis* polymerase was found to be due to an increase in the *rate* rather than in the *extent* of DNA replication. This stimulation was found to arise from several effects that were produced by the presence of binding protein (*179*). When denatured DNA was used as a template for *U. maydis* polymerase, the apparent K_m for nucleoside triphosphates was increased 3- to 4-fold in the presence of binding protein. In addition, the apparent affinity of the polymerase for

177. F. Cobianchi, S. Riva, G. Mastromel, S. Spadari, G. Pedrali-Noy, and A. Falaschi, *CSHSQB* **43**, 639 (1979).

178. G. R. Banks and A. Spanos, *JMB* **93**, 63 (1975).

179. G. Yarranton, P. D. Moore, and A. Spanos, *Mol. Gen. Genet.* **145**, 215 (1976).

DNA was increased fourfold by the *U. maydis* HDP. And finally, the V_{max} of the polymerization reaction increased ~50%. All of these effects have been interpreted to indicate that the *U. maydis* binding protein stimulates replication both by removing secondary structure in the DNA template, which may impede the polymerase, and by providing a specific protein–nucleic acid complex with which the DNA polymerase can interact.

4. *Mouse DNA Binding Proteins*

Several DNA binding proteins have been isolated from mouse tissue, including proteins isolated from 3T6 cells (*180*), ascites cells (*181*), and myeloma cells (*182*). Each of these proteins is a helix-destabilizing protein, and the last two stimulate mouse polymerase α. The role of the protein isolated from 3T6 cells is unknown; it is found primarily in the cytoplasm, and more is found in growing than in resting cells.

The HDP found in mouse ascites cells has a molecular weight of ~35,000, and exists as a monomer in solution (*181*). It shows a high affinity for single-stranded DNA or RNA, and only a slight affinity for double-stranded DNA. Sedimentation experiments suggest that the protein–nucleic acid complex is asymmetric and highly extended, and that saturation of the DNA occurs at ~6–10 nucleotide residues per protein monomer. The properties of this protein suggest that it is analogous to the high-salt fraction protein isolated from calf thymus.

The mouse protein also stimulates the activity of mouse DNA polymerase α ~ fourfold, but only on a heat-denatured DNA template. There is no stimulation of activity on pancreatic DNAse-treated (activated) DNA templates, but we note that the activity of the DNA polymerase is already sevenfold greater on the activated than on the denatured DNA template. This helix-destabilizing protein may specifically stimulate polymerase α, since assays of the effect of this protein on mouse polymerase β, *E. coli* DNA polymerase, and T7 DNA polymerase all show only very slight stimulatory effects. The effect on polymerase α has been surmised to be at the elongation step. It has also been showed that DNA binding protein increases the affinity of polymerase α for DNA cellulose (*183*) similar to the effect shown with the *U. maydis* protein. Although no direct association between polymerase and binding protein has been seen, these results suggest that the binding protein may stabilize the polymerase α–DNA complex.

180. R. L. Tasi and H. Green, *JMB* **73**, 307 (1973).
181. B. Otto, M. Baynes, and R. Knippers, *EJB* **73**, 17 (1977).
182. S. R. Planck and S. H. Wilson, *JBC* **255**, 11547 (1980).
183. A. Richter, R. Knippers, and B. Otto, *FEBS* (*Fed. Eur. Biochem. Soc. Lett.*) **91**, 293 (1978).

As with the protein isolated from lily, the mouse DNA binding protein can be phosphorylated (1 phosphate/protein monomer) by a chromatin-associated kinase, resulting in an alteration of its binding properties. Although phosphorylation has no effect on single-stranded DNA affinity, the phosphorylated protein shows a reduced affinity for double-stranded DNA and a greatly reduced ability to stimulate DNA polymerase α. Thus this system seems to represent another example of HDP activity controlled by the level of covalent phosphorylation.

Another protein that has been isolated from mouse cells (myeloma), and is clearly not the same as those obtained from 3T6 or ascites cells, is mouse HDP-I (*182*). This protein is heterogeneous in molecular weight, ranging from 24,000 to 33,000, with a predominant species at ~27,000. This apparent heterogeneity is not due to different levels of phosphorylation, acetylation, or glycosylation; rather peptide mapping and tryptic digestion studies implicate protease activity. This HDP is localized predominantly in the cell nucleus, and is not associated with the chromatin; however ~25% is also found in the cytoplasm. Note that the amino acid compositions and the molecular weights of the mouse HDP-I and CT HDP-I are very similar.

Sedimentation experiments with mouse HDP-I and denatured DNA indicate that the protein binds noncooperatively to single-stranded DNA, with a site size of 5 to 7 nucleotide residues per protein monomer and a binding constant of $\sim 4 \times 10^5\ M^{-1}$. In addition, saturating amounts of this protein depress the T_m of poly(dA-dT) by ~25° in 0.01 *M* salt. The binding of DNA to this protein induces a fluorescence *increase* of ~35%, in contrast to fluorescence quenching typically seen with prokaryotic DNA binding proteins (*182*).

Studies of the products of limited tryptic hydrolysis of the mouse HDP-I have shown that different products are obtained, depending on whether or not the protein is bound to single-stranded nucleic acids in the digestion process. This effect depends on the type of nucleic acid present, with poly(dT) and denatured DNA protecting the protein most effectively. Poly(dI), poly(rA), and poly(dC) alter the digestion pattern to a lesser extent, and poly(dA) and poly(dA-dT) are quite ineffective. Thus mouse HDP-I binding may display some base composition-dependent binding specificity. Short oligonucleotides are much less effective in protecting the ~19,000 molecular weight digestion product, and this spectrum of protection effectiveness has been used to measure the affinity of oligonucleotides for HDP-I. In addition the DNA binding properties of two proteolytic products (molecular weights of about 19,000 and 22,000) that lack the amino terminus are identical to that of native HDP-I, suggesting that ~65

residues at that end of the polypeptide are not essential for DNA binding (*182*).

The mouse HDP-I seems to serve as an accessory protein to mouse DNA polymerase α (*184*). This conclusion is based on the fact that HDP-I increases the processivity of DNA polymerase α in a manner very similar to that observed for T4 gene 32 protein. In the absence of HDP-I, polymerase α can processively extend an RNA-primer, yielding a Poisson distribution of products with a maximum (nucleotides added) length of 8–9 residues. When HDP-I is added to this system, the peak of the distribution of added nucleotides is increased to 21 nucleotide residues, suggesting that eukaryotic binding proteins have similar effects on eukaryotic *in vitro* replication as do the procaryotic proteins (*106, 185*).

5. *Adenovirus DNA Binding Protein*

Early in infection of human cells by oncogenic (DNA) adenoviruses (types 2 and 5), a 72,000 molecular weight viral-coded DNA binding protein (Ad DNA binding protein) is synthesized in very large numbers ($\sim 10^7$ copies per cell) (*186, 187*). The gene coding for this DNA binding protein has been mapped on the adenovirus genome (*188*), and a temperature-sensitive mutant (*H5ts125*) has been isolated (*189, 190*). Adenovirus carrying this mutation is defective in the initiation of viral DNA replication (*191*), and the DNA binding protein isolated from cells infected with the mutant strain is defective, both in binding to single-stranded DNA (*190*) and in complementation assays for *in vitro* replication (*191a*). The protein is phosphorylated *in vivo* (*187, 192–194*). It also appears to be a major

184. S. D. Detera, S. P. Becerra, J. Swack, and S. H. Wilson, *JBC*, in press (1981).
185. C. C. Huang, J. E. Hearst, and B. M. Alberts, *JBC* **256**, 4087 (1981).
186. P. C. Van der Vliet and A. J. Levine, *Nature New Biol.* **246**, 170 (1973).
187. T. Linné, H. Jörnvale, and L. Pullipson, *EJB* **76**, 481 (1977).
188. J. B. Lewis, J. F. Atkins, P. R. Bam, R. Solem, R. F. Gesteland, and C. W. Anderson, *Cell* **7**, 141 (1976).
189. M. J. Ensinger and H. S. Ginsberg, *J. Virol.* **10**, 328 (1972).
190. P. C. Van der Vliet, A. J. Levine, M. J. Ensinger, and H. S. Ginsberg, *J. Virol.* **15**, 348 (1975).
191. P. C. Van der Vliet and J. S. Sussenbach, *Virology* **67**, 415 (1975).
191a. M. S. Horiwitz, L. M. Kaplan, M. Abboud, J. Maritato, L. T. Chow, and T. R. Broker, *CSHSQB* **43**, 769 (1978).
192. N. Axelrod, *Virology* **87**, 366–383 (1978).
193. Y.-H. Jeng, W. S. M. Wold, K. Sugawara, Z. Gilead, and M. Green, *J. Virol.* **22**, 402 (1977).
194. A. D. Levinson, E. H. Postel, and A. J. Levine, *Virology* **79**, 144 (1977).

component of adenovirus replication complexes isolated from infected cells (*195–197*). Whereas this DNA binding protein has been shown to be important in initiation, recent evidence suggests that it may also function during the elongation step of adenovirus DNA replication (*198*).

The Ad DNA binding protein appears to be fibrous (*199, 200*), and to exist as a monomer in solution. It can be cleaved into two subfragments of molecular weights about 26,000 and 44,000 (*201, 202*). The isolated 26,000 fragment is derived from the amino-terminal domain of the protein, contains most of the sites of protein phosphorylation, and does not bind single-stranded DNA (*201*). The 44,000 fragment is derived from the carboxy terminus of the protein, binds to single-stranded DNA as an isolated fragment, and can function in *in vitro* DNA replication (*201, 202*). The defined mutation site (*H5ts125*) is located in the larger fragment.

The DNA binding properties of this protein have only been roughly characterized. Although the protein was initially thought to be a single-stranded specific DNA binding protein (site size ~7 nucleotide residues per protein monomer) on the basis of DNA-cellulose (*186*) and nitrocellulose filter binding assays (*200*), it has since been shown that the protein does not melt poly(dA-dT) and may in fact stabilize this duplex polynucleotide (*203*). In addition, it has been demonstrated that Ad DNA binding protein can bind to double-stranded adenovirus DNA, although it shows a much higher affinity for the termini of duplex DNA molecules (*203*).

As previously indicated, the Ad DNA binding protein has been shown to be phosphorylated *in vivo* at several sites (*187, 192–194*). Although the role of phosphorylation in the function of Ad DNA binding protein has been elusive, it appears that the more extensively phosphorylated species have a lower affinity for single-stranded DNA-cellulose (*201*). Further studies suggest that newly phosphorylated DNA binding protein associates preferentially with replicating viral DNA; after a period of time the protein is also found associated with the mature duplex DNA (*195*). A better understanding of the relationship between phosphorylation, DNA

195. C. H. Shaw, W. C. Russell, and M. K. Rekosh, *Virology* **92,** 436 (1979).
196. C. Kedinger, O. Brison, F. Perrin, and J. Wilhelm, *J. Virol.* **26,** 364 (1978).
197. M. Arens, T. Yamashita, R. Padmanbhian, T. Tsuruo, and M. Green, *JBC* **252,** 7947 (1977).
198. P. C. Van der Vliet, J. Zandberg, and H. S. Jansz, *Virology* **80,** 98 (1979).
199. K. Sugawara, Z. Gilead, and M. Green, *J. Virol.* **21,** 338 (1977).
200. P. C. Van der Vliet, W. Keegstra, and H. S. Jansz, *EJB* **86,** 389 (1978).
201. H. Klein, W. Maltman, and A. J. Levine, *JBC* **254,** 11051 (1979).
202. H. Ariga, H. Klein, A. J. Levine, and M. S. Horwitz, *Virology* **101,** 307–310 (1980).
203. D. M. Fowlkes, S. T. Lord, T. Linne, U. Pettersson, and L. Philipson, *JMB* **132,** 163 (1979).

binding properties, and the role of this protein in DNA replication awaits further experimentation.

V. DNA Binding Proteins as Research Tools

A. Electron Microscopy

Single-stranded DNA binding proteins are, in general, too small to discern as individual molecules in the electron microscope. However complexes formed by these proteins with DNA and RNA cause distinct morphological changes in these entities. This property of the binding proteins has been exploited as a tool for the electron microscopic visualization of single-stranded DNA, which under normal conditions is difficult to see as a "naked" species, and may be severely deformed during spreading and grid preparation. In this way the T4 gene 32 protein has been shown to be useful in the mapping of single-stranded regions of DNA–DNA and DNA–RNA hybrids (*204, 205*). This technique has been applied to the mapping of ribosomal RNA and tRNA genes on ϕ80 phage genomes (*205*), the mapping of histone and ribosomal genes in *Drosophila* (*206, 207*), and the mapping of terminal sequences in the adenovirus genome (*208*).

The *E. coli* SSB protein has been used to stabilize and visualize single-stranded DNA sequences generated by *recBC* enzyme [see Chapter 13, this volume, and Refs. (*209, 210*)] and to visualize single-stranded ends of reconstituted histone-DNA complexes (*211*).

B. Biochemical Assays

Single-stranded DNA binding proteins have been used as traps for sequences of single-stranded DNA that are transiently formed during enzymatic assays. This approach has been particularly useful in studying the (ATP-dependent) unwinding reaction of the class of enzymes known as

204. C. Brack, T. A. Bickle, and R. Yuan, *JMB* **96,** 693 (1975).
205. M. Wu and N. Davidson, *PNAS* **72,** 4506 (1975).
206. M. Wu, D. S. Holmes, N. Davidson, R. H. Cohn, and L. H. Kedes, *Cell* **9,** 163 (1976).
207. M. Pellegrini, J. Manning, and N. Davidson, *Cell* **10,** 213 (1977).
208. M. Wu, R. J. Roberts, and N. Davidson, *J. Virol.* **21,** 766 (1977).
209. J. Rosamond, K. M. Telander, and S. Linn, *JBC* **254,** 8646 (1979).
210. A. Taylor and G. R. Smith, *Cell* **22,** 447 (1980).
211. K. Dunn and J. D. Griffith, *Nucleic Acids Res.* **8,** 555 (1980).

DNA helicases [see Chapter 20 of this volume and Refs. (*212–215*)]. Specifically, either *E. coli* HDP or T4 gene 32 protein have often been used to sequester single-stranded DNA regions formed by the unwinding of duplex DNA by helicases. After stopping the enzymatic reaction, the HDP-nucleic acid complex is dissociated and the free single-stranded DNA is digested by S1 nuclease. The extent of unwinding can then be determined by measuring (typically, radiochemically) the amount of S1 nuclease-resistant duplex DNA that remains. The unwinding activity of *E. coli* helicase III and *rep* protein have been measured in this manner, and it was shown that HDP is required to trap the unwound DNA (*215*). In contrast *E. coli* helicases I and II, and the T4 helicase (at high protein concentration) can unwind duplex DNA in the absence of HDPs (*213*).

VI. Conclusions

Although the specific details of the interactions of the single-stranded DNA binding proteins with nucleic acids, and their exact biological functions, differ somewhat for each protein considered in Section IV, some overriding generalizations emerge. Thus all these proteins seem to operate stoichiometrically (as opposed to catalytically), in that they are present at intracellular levels sufficient to effectively saturate the single-stranded DNA intermediates produced during replication, recombination, and repair. To avoid dissipation over the great excess of double-stranded DNA present in the cell, most of the proteins show appreciably more affinity for single-stranded than for double-stranded DNA. Furthermore this net difference in affinity is (at least for the prokaryotic proteins) amplified by the fact that binding to single-stranded nucleic acids is cooperative. This binding cooperativity is essential in permitting complete coverage of single-stranded sequences, and also in effectively destabilizing the small duplex hairpins formed by intrastrand base pairing in single-stranded DNA. [When present, such hairpins can slow down or stop the utilization of the involved sequence as a template for DNA polymerase [see Refs. (*99, 185*)]. In addition, uncomplexed single-stranded sequences are very susceptible to attack by intracellular endonucleases. It seems most likely that removing hairpins from transient single-stranded DNA sequences, and

212. M. Abdel-Monem, H. F. Lauppe, J. Kartenbeck, H. Durwald, and H. Hoffman-Berling, *JMB* **110,** 667 (1977).

213. B. Kuhn, M. Abdel-Monem, H. Krell, and H. Hoffman-Berling, *JBC* **254,** 11343 (1979).

214. M. Duguet, G. T. Yarranton, and M. L. Gefter, *CSHSQB* **43,** 335 (1978).

215. G. T. Yarranton, R. H. Das, and M. L. Gefter, *JBC* **254,** 12002 (1979).

protecting these sequences against nucleases, comprises the central task of the single-stranded DNA binding proteins *in vivo*.

Of course these proteins must function within multicomponent complexes. As a consequence most appear to have developed some degree of interactional specificity with other proteins of the homologous complex; this is manifested particularly (in *in vitro* assays) by specific stimulation of certain homologous polymerases. These effects might proceed by a variety of mechanisms. For example, interaction with the homologous polymerase could prevent the destabilization of the primer-template complex by single-stranded DNA binding protein. Alternatively, the binding protein could put the single-stranded template into an optimal geometry for utilization as a companion template by the homologous polymerase. Another possibility is that only the homologous binding protein can be effectively removed from the single-stranded DNA by components of the homologous replication (or recombination) complex. Further experiments with complete *in vitro* systems are required to choose between these possibilities or others.

In order to effectively discharge its primary function(s), the binding of binding protein to single-stranded DNA lattices must be relatively nonspecific with respect to nucleotide sequence or composition. At the same time, for effective functioning of replication (and probably recombination and repair) complexes, the presence of either too much or too little single-stranded binding protein could be inhibitory, or even lethal. Thus some mechanism for controlling the intracellular concentration of SSB proteins is probably generally required.

For T4 gene 32 protein this regulation is autogenous at the translational level, and involves differences in affinity for nucleic acid sequences based both on sugar and on nucleotide residue type. These affinity differences are relatively small at the level of the binding of the individual protein molecule, as required to avoid problems of incomplete saturation of single-stranded DNA. Yet these differences are also large enough to be amplified, by binding cooperativity, into control systems of considerable overall specificity (see Ref. 5).

The molecular details of the interactions of single-stranded DNA binding proteins with nucleic acid lattices, resulting in strong overall binding modulated by some binding specificity, are just beginning to emerge. Electrostatic interactions are generally involved; at the same time more specific binding interactions, based on hydrogen bonding and possible stacking interactions of bases with aromatic acid residues in the binding site, may also participate.

The principles outlined above are probably involved, in various combinations, in the *in vivo* functioning of most single-stranded binding proteins.

To the extent that present results permit us to judge, the eukaryotic SSB proteins have many features in common with the better-studied prokaryotic proteins. However some significant differences are seen. For example, some or most of the eukaryotic proteins that have been examined (see Table III) may bind noncooperatively to single-stranded nucleic acid lattices. The significance of this is not clear; perhaps these eukaryotic proteins operate *in vivo* in collaboration with factors or proteins yet to be discovered to achieve the binding saturation brought about by binding cooperativity in the prokaryotic systems. We note also that both the biological activities and the DNA binding properties of several of the eukaryotic binding proteins seem to be modulated *in vivo* by enzymatically catalyzed covalent phosphorylation and dephosphorylation reactions. Such processes have not been observed with prokaryotic single-stranded binding proteins, and when fully understood may turn out to be involved in controlling the effective binding levels and specificities (and binding cooperativity?) of the eukaryotic SSB proteins.

Overall, patterns of single-stranded DNA binding protein properties are starting to emerge (see Tables II and III). However the total range of function in which these proteins participate will not be clear until we understand, in molecular detail, the entire physiological systems of which these proteins form an integral part.

ACKNOWLEDGMENTS

We are pleased to acknowledge many conversations with colleagues, both at the University of Oregon and elsewhere, that have helped to focus our thinking and improve our understanding of the subject matter of this review. We are also grateful to many colleagues, including Drs. Bruce Alberts, Rae Lynn Burke, Joseph Coleman, Malcolm Gefter, Larry Gold, Jack Griffith, Junko Hosoda, Richard Karpel, William Konigsberg, Timothy Lohman, Alexander McPherson, Thomas Record, Kenneth Williams and Samuel Wilson, who provided unpublished data or preprints of their relevant papers prior to publication. We wish to thank Ms. Nancy Caretto, who patiently typed and retyped the many drafts of this review. The research described here that has been carried out in our laboratory has been supported, in part, by U.S. Public Health Services Research Grant GM-15792, American Cancer Society Postdoctoral Fellowship PF-1301 (to S. Kowalczykowski) and USPHS Postdoctoral Fellowship GM-06676 (to D. Bear).