Proc. Natl. Acad. Sci. USA
Vol. 77, No. 8, pp. 4614–4617, August 1980
Biochemistry

Amino acid sequence of the T4 DNA helix-destabilizing protein

(DNA-binding protein/gene 32 protein/partial proteolysis/protein sequencing)

KENNETH R. WILLIAMS, MARY B. LOPRESTI, MASAYUKI SETOGUCHI, AND WILLIAM H. KONIGSBERG

Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510

Communicated by Frederic M. Richards, May 12, 1980

ABSTRACT **The primary structure of the T4 single-stranded DNA-binding protein coded by gene 32 has been determined by manual and automated sequencing of peptides derived from partial proteolysis, cyanogen bromide cleavage, and digestion with trypsin, chymotrypsin, and staphylococcal protease. Tryptic digestion of citraconylated or succinylated gene 32 protein yields five peptides containing 4, 27, 42, 65, and 163 residues, which can be separated by Sephadex chromatography. Each of the tryptic peptides was subjected to automated sequencing and, if necessary, more extensive cleavage. The intact protein contains 301 amino acids, has a molecular weight of 33,487, and can be specifically cleaved at lysines 21 and 253 by limited trypsin digestion. Previous studies have shown that the "B" region (residues 1–21), which has a charge of +4, is important for the protein–protein interactions involved in gene 32 protein self-association and cooperative binding to single-stranded DNA. The "A" region (residues 254–301) has been implicated in controlling the helix-destabilizing "activity" of gene 32 protein and in interacting with other T4 DNA replication proteins. The A region has a charge of −10 and, in addition, contains two unusual stretches of four serine residues separated by glycine 284. The region between positions 73 and 115 contains 75% of the tyrosine residues and may be important for DNA binding.**

Gene 32 of bacteriophage T4 codes for a DNA-binding protein that is required for T4 DNA replication (1), repair (2), and recombination (3). Although the precise function of the gene 32 protein in these processes is not understood, the ability of the protein to bind cooperatively to single-stranded DNA (ss DNA) (4–7) is undoubtedly crucial. This preferential binding to single-stranded nucleic acids results in a decrease of the thermal denaturation temperature of double-stranded polymers such as poly(dA)·poly(dT) when mixed with gene 32 protein. For this reason the gene 32 protein and similar DNA-binding proteins from bacteriophages fd (8) and T7 (9), *Escherichia coli* (10), rat liver (11, 12), and calf thymus (13) have been classified as helix-destabilizing proteins (14).

In addition to binding ss DNA, the gene 32 protein appears to interact directly with several T4 proteins, including DNA ligase (15), exonucleases involved in DNA recombination (16), DNA polymerase (17), and the DNA polymerase "accessory" proteins coded by genes 44 and 62 (18). As a first step in interpreting structure–function relationships, we have determined the amino acid sequence of the T4 DNA helix-destabilizing protein. This information will allow us to interpret more fully the physical-chemical and functional properties of the gene 32 protein.

MATERIALS AND METHODS

The gene 32 protein was isolated from *E. coli* B cells infected with T4 [*amN134, amBL292* (−33, −55)]. Cells (50 g) were resuspended in 200 ml of buffer A (10 mM Tris·HCl, pH 8.0/1 mM EDTA/1 mM 2-mercaptoethanol/50 mM NaCl) and were broken by sonication. After addition of $MgCl_2$ and $CaCl_2$ to final concentrations of 20 mM and 4 mM, respectively, deoxyribonuclease I (20 μg/ml) was added and the suspension was stirred for 2 hr at room temperature. NaCl was added to a final concentration of 0.5 M; after an additional hour, the extract was centrifuged at 4000 × *g* for 20 min. After a high-speed centrifugation (68,000 × *g* for 90 min) the supernate was dialyzed first against buffer A and then against buffer A containing 5 mM EDTA and 10% (vol/vol) glycerol. The dialyzed extract was then applied to a ss DNA-cellulose column. The gene 32 protein was eluted with 2.0 M NaCl (4) and was greater than 95% pure based on the results of $NaDodSO_4$ gel electrophoresis.

The purified protein was extensively dialyzed against water, lyophilized, and then carboxamidomethylated with iodo[^{14}C]-acetamide (19) prior to cleavage at methionine or arginine. For methionine cleavage, 0.15 g of cyanogen bromide was added to 1.8 μmol of carboxamidomethylated protein dissolved in 10 ml of 70% (vol/vol) formic acid. After 24 hr at room temperature, the reaction was stopped by addition of 90 ml of water and lyophilization. The resulting peptides were fractionated by chromatography on Sephadex G-100 (2.5 × 140 cm) in 20 mM Tris·HCl, ph 8.4/6 M guanidine·HCl. CnBr-3 and CnBr-4 (Fig. 1), which coeluted on Sephadex G-100, were separated by chromatography on Bio-Gel P-6 (1.5 × 120 cm) in 0.5% acetic acid. Gene 32 protein was cleaved at arginine by first blocking the lysine residues with succinic anhydride (if the peptides were to be used for automatic sequencing) or citraconic anhydride (if the peptides were to be subjected to further enzymatic digestion) and then digesting with trypsin. In either case, the carboxamidomethylated protein (1.4 μmol) was dissolved in 15 ml of 6 M guanidine·HCl/0.2 M $NaHCO_3$, pH 8.5. Succinic anhydride (5.0 g) or citraconic anhydride (0.3 ml) was added over a period of 30 min while the pH was maintained at 8.5 by addition of 2 M NaOH. The derivatized protein was then dialyzed against 25 mM NH_4HCO_3. Succinylated gene 32 protein was digested for 6 hr and citraconylated gene 32 protein was digested for 1.5 hr at protein:trypsin weight ratios of 50:1 and 30:1, respectively, at room temperature. In both instances, the tryptic peptides were concentrated by lyophilization and then chromatographed on Sephadex G-100 (2.5 × 100 cm) equilibrated with 25 mM NH_4HCO_3 (Fig. 2). The citraconylated peptides were decitraconylated by dissolving them in 70% (vol/vol) formic acid (for peptide A-5) or 5% (vol/vol) acetic acid [for peptides A-(2,3) and A-4] and leaving them at room temperature for 16 hr.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: gene 32 protein, DNA helix-destabilizing protein coded by gene 32 of bacteriophage T4; ss DNA, single-stranded DNA; 32P*-A, tryptic cleavage product of gene 32 protein that lacks the carboxy-terminal "A" region (residues 254–301); 32P*-(A+B), tryptic cleavage product of gene 32 protein that lacks the carboxy-terminal "A" region and the amino-terminal "B" region (residues 1–21); ds DNA, double-stranded DNA.

The decitraconylated tryptic peptides were additionally cleaved with trypsin, chymotrypsin, and staphylococcal protease at 37°C in 50 mM NH_4HCO_3 (pH 7.8). Staphyloccal protease and trypsin digestions were at peptide:enzyme weight ratios of 30:1 and were continued for 12 hr and 6 hr, respectively. Chymotrypsin digestion was for 3 hr at a protein:enzyme weight ratio of 100:1. These digests were complex mixtures which required a size separation step on Bio-Gel P-4 (1.5 × 110 cm; 50 mM NH_4HCO_3) prior to separation by ion-exchange chromatography with Aminex 50W-X4 (20–30 μm, Bio-Rad). Ion-exchange chromatography was at 50°C on a 0.9 × 55 cm column in pyridine acetate buffers. The sample was applied in 50 mM pyridine acetate (pH 3.2), and peptides were eluted with two consecutive 200-ml linear gradients: from 0.05 M to 0.50 M pyridine acetate at pH 4.1 and then from 0.50 M to 1.50 M pyridine acetate at pH 4.7. Strongly adsorbed peptides were eluted by washing with 8.5 M pyridine acetate at pH 5.6. Peptides were detected by reaction with ninhydrin after alkaline hydrolysis.

The major product of proteolysis of gene 32 protein—namely, 32P*-(A+B), which lacks the carboxy-terminal "A" region (residues 254–301) and the amino-terminal "B" region (residues 1–21)—was prepared as described (20) and was carboxamidomethylated prior to automated sequencing. Two chymotryptic peptides, C-5 and C-6 (Fig. 1), were derived from 32P*-(A+B).

Sequences of the larger peptides were determined by automated Edman degradation with a Beckman model 890C Sequencer. Phenylthiohydantoin derivatives were identified by amino acid analysis after hydrolysis with hydriodic acid for 18 hr at 130°C (21). Glutamine, glutamic acid , asparagine, and aspartic acid residues were usually identified by high-pressure liquid or gas chromatography of the phenylthiohydantoin derivatives, but in a few instances, the electrophoretic mobility of short peptides containing these residues was used to assign amides (22). The sequences of smaller peptides were manually determined by the procedure of Tomita *et al.* (23) except that the ethyl acetate extraction of the phenylthiocarbamyl peptide

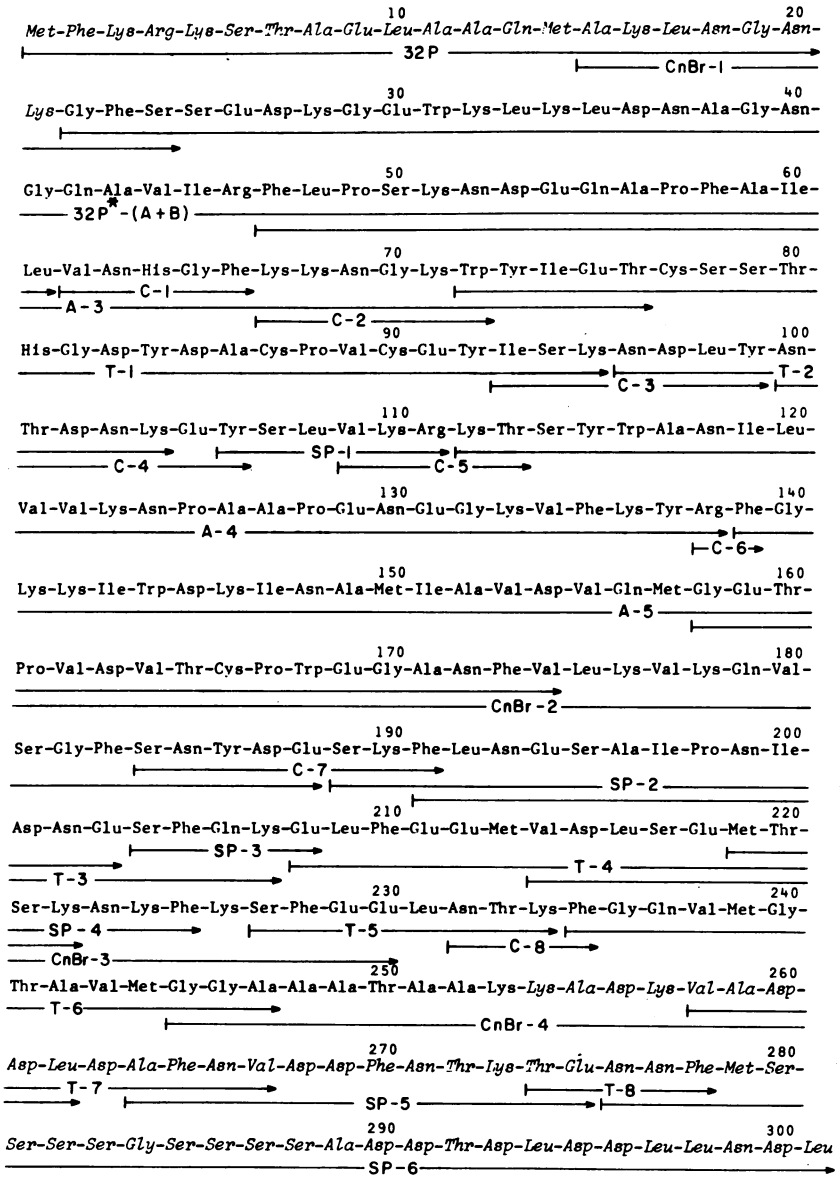

FIG. 1. Amino acid sequence of T4 DNA helix-destabilizing protein. The single-headed arrow indicates the last residue sequenced in each peptide which, because some of the peptides were only partially sequenced, is not necessarily the COOH terminus of the peptide. Cleavage methods used are: CnBr, cyanogen bromide; 32P*-(A+B), partial proteolysis; A, tryptic digestion of succinylated or citraconylated gene 32 protein; T, trypsin; C, chymotrypsin; and SP, staphylococcal protease. The B region (residues 1–21) and the A region (residues 254–301) are italicized.

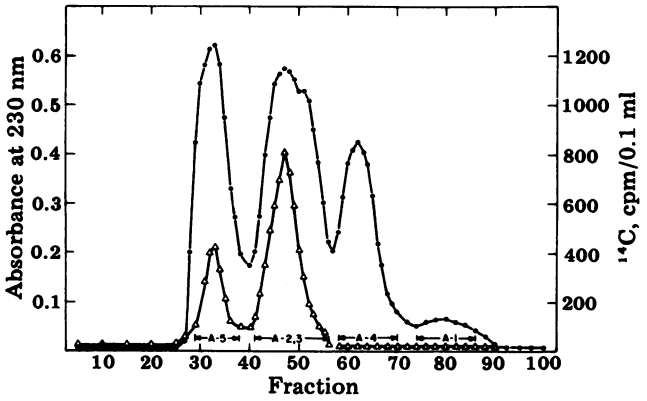

FIG. 2. Sephadex G-100 chromatography of tryptic peptides from succinylated, [^{14}C]carboxamidomethylated gene 32 protein. The derivatized protein (1.4 μmol) was incubated at 37°C with 1 mg of trypsin in a total volume of 25 ml of 25 mM NH_4HCO_3. After 6 hr, the digest was concentrated to 3 ml by lyophilization and applied to a 2.5 × 100 cm column of Sephadex G-100 equilibrated with 25 mM NH_4HCO_3. Fractions, of ≈6.3 ml each were collected. ●, $A_{230\ nm}$; Δ, ^{14}C cpm.

was omitted. Carboxypeptidase digests contained 0.1 unit of carboxypeptidase A and 0.05 unit of carboxypeptidase B per 10 nmol of peptide in a total volume of 0.1 ml of 50 mM $NaHCO_3$ at pH 8.2. After incubation for 30 min at 37°C, the reaction was stopped by addition of 10 μl of 6 M HCl and drying under reduced pressure.

RESULTS AND DISCUSSION

The amino acid sequence of the T4 gene 32 protein is given in Fig. 1. The sequence of the first 60 residues was determined by automated Edman degradation of the protein and its limited proteolysis product, 32P*-(A+B). The sequence of this section confirms our earlier report that the B region of the protein (Fig. 1) comprises the first 21 amino acids (24). After cleavage of succinylated, [^{14}C]carboxamidomethylated gene 32 protein at arginine residues by trypsin and separation of the resulting peptides on Sephadex G-100, four fractions were obtained, A-1, A-(2,3), A-4, and A-5, which contained residues 1–4, 5–46 and 47–111, 112–138, and 139–301, respectively (Fig. 2). Although cleavage occurred at arginine 46, several attempts at separating the two peptides, containing residues 5–46 and 47–111, failed. Therefore, the approximately equimolar mixture of the two peptides was subjected to automated sequencing without further purification. Because the sequence from residues 5–46 was known from previous sequenator runs on gene 32 protein and 32P*-(A+B), it was possible to determine which of the two amino acids at each cycle was derived from the region containing residues 5–46. In addition, the sequence from position 62 to position 72 was confirmed by sequencing the peptides labeled C-1 and C-2 in Fig. 1. From these results the sequence was extended to position 76. Automated sequencing of A-4 and A-5 established the sequence spanning residues 112–174. Automated and manual sequencing of peptides derived from cyanogen bromide cleavage of gene 32 protein and from enzymatic digestion of decitraconylated A-(2,3) and A-5 established the remainder of the sequence shown in Fig. 1.

The calculated molecular weight of gene 32 protein is 33,487, which is very close to the value of ≈35,000 determined by gel filtration (4) or by $NaDodSO_4$ gel electrophoresis (24). In addition, the composition of the protein predicted from the sequence in Fig. 1 agrees with published data on the amino acid content of the protein (24–27) and the sum of the amino acid compositions of the tryptic peptides (Table 1). Based on the

Table 1. Amino acid composition of tryptic peptides from succinylated gene 32 protein*

Amino acid	Peptide† A-1	A-(2,3)	A-4	A-5	Total
Cys	0.3	3.0 (3)		1.1 (1)	4.4 (4)
Asx	0.2	15.2 (17)	3.1 (3)	26.9 (31)	45.4 (51)
Thr	0.3	4.1 (4)	1.0 (1)	8.7 (9)	14.1 (14)
Ser	0.3	7.5 (8)	1.0 (1)	13.6 (16)	22.4 (25)
Glx	0.1	10.5 (10)	2.3 (2)	16.7 (16)	29.6 (28)
Pro	0.3	3.9 (3)	2.4 (2)	3.2 (3)	9.8 (8)
Gly	0.3	8.1 (8)	1.3 (1)	9.1 (9)	18.8 (18)
Ala	0.2	8.3 (9)	2.8 (3)	14.0 (14)	25.3 (26)
Val	0.1	3.9 (4)	2.0 (3)	11.7 (12)	17.7 (19)
Met	0.9 (1)	1.1 (1)	0.2	5.7 (7)	7.9 (9)
Ile	0.1	3.3 (4)	1.0 (1)	4.7 (5)	9.1 (10)
Leu	0.2	8.3 (8)	1.2 (1)	10.5 (10)	20.2 (19)
Tyr	0.1	4.8 (5)	1.7 (2)	1.2 (1)	7.8 (8)
Phe	0.9 (1)	4.2 (4)	0.9 (1)	11.8 (12)	17.8 (18)
His		1.7 (2)			1.7 (2)
Lys‡	1.0 (1)	14.0 (13)	3.8 (4)	15.9 (15)	34.7 (33)
Arg	1.1 (1)	2.1 (2)	1.0 (1)		4.2 (4)
Trp§		1.9 (2)	0.8 (1)	2.3 (2)	5.0 (5)
Yield, μmol	1.45	1.24	1.48	1.21	

* Aliquots were taken from the pools indicated in Fig. 2 and hydrolyzed with 6 M HCl/0.2% phenol for 24 hr at 110°C. Numbers in parentheses were determined from the sequences of the peptides.
† A-1, residues 1–4; A-(2,3), residues 5–46 and 47–111; A-4, residues 112–138; A-5, residues 139–301.
‡ Includes monocarboxymethyllysine produced as a result of carboxymethylating gene 32 protein.
§ Tryptophan was estimated spectrophotometrically.

carboxypeptidase data previously obtained (24), the carboxy terminus of 32P*-A (the cleavage product that lacks the A region) was assigned to lysine 253. The number of sulfhydryl groups in gene 32 protein was estimated by treating the protein in 6 M guanidine·HCl with 5,5′-dithiobis-(2-nitrobenzoic acid) (28). Under these conditions, 3.7 mol of sulfhydryl groups were found per mol of gene 32 protein, thus indicating that the protein contains four cysteine residues and no disulfide bonds.

One of the interesting features of gene 32 protein is the possibility that it contains specific domains for each of the protein–protein and protein–nucleic acid interactions in which it participates. This was suggested because these interactions can be eliminated independently by various point mutations (29–31). *In vitro* studies directed at structure–function relationships in gene 32 protein have also tended to confirm the notion that the protein contains several discrete functional domains. For instance, the NH_2-terminal B region (residues 1–21) has been shown to be essential for cooperative binding to ss DNA (20, 32).

The unusual charge distribution in gene 32 protein (the NH_2-terminal portion has a charge of +8 whereas the COOH-terminal portion has a charge of −17) supports the conclusion drawn from studies with T4 amber mutants in gene 32 (29) that the DNA-binding site is contained within the NH_2-terminal portion of the protein. These studies have also shown that an amber fragment of the gene 32 protein (*amA453*), which has a molecular weight of ≈10,000 and therefore comprises only the first 90–100 residues, contains most or perhaps all of the DNA–protein and protein–protein interaction sites required for initiation of DNA replication and recombination (29). Within the NH_2-terminal half of gene 32 protein, the region spanning residues 73–115 is particularly

unusual in that it contains six of the eight tyrosine residues in the protein. Five of the six tyrosine residues in this region are separated by six to eight amino acids and are therefore almost equally spaced in the primary sequence. Because Anderson and Coleman (27) have suggested that five of the tyrosine residues in gene 32 protein intercalate between the bases in ss DNA, it is tempting to speculate that this region of the protein (between residues 73 and 115) is important for DNA binding.

Cleavage at lysine 253 with trypsin results in removal of the A region (Fig. 1), which has been implicated in controlling the helix-destabilizing "activity" of gene 32 protein (33). Removal of the A region apparently circumvents the "kinetic block" that may account for the failure of the protein to denature native double-stranded DNA (ds DNA) (34). In line with this, 32P*-A is retained by ds DNA-cellulose (33) and can denature T4 ds DNA (35), properties not exhibited by the intact protein. Aside from the observation that the cooperativity of binding of gene 32 protein to ss DNA is enhanced by removal of the A region (20), there is no obvious explanation for the increased helix-destabilizing activity of 32P*-A as compared to gene 32 protein. Moise and Hosoda (33) suggested that interaction of the A region of the protein with other T4 DNA replication proteins might have the same effect as the actual removal of the A region and therefore limit the helix-destabilizing activity of the protein to only the small section of ds DNA just in front of the replication fork. The susceptibility of at least a portion of gene 32 protein (the A region) to proteolysis when it is bound to ss DNA (24) and the high content of polar† amino acids (67%) in the COOH-terminal section suggests that the A region probably is accessible for interactions with other proteins in the T4 DNA replication complex. In fact, Hosoda *et al.* (36) have observed that loss of the A region is accompanied by a loss of affinity of the gene 32 protein for T4 DNA polymerase (gene 43 protein) and the T4 RNA priming protein (gene 61 protein). It is not apparent how the negative charge (−10) or the unusual cluster of serine residues in the A region (residues 280–283 and 285–288) is related to the presumed functions of this domain of gene 32 protein *in vivo*.

Knowledge of the primary structure of gene 32 protein should allow more precise interpretation of structure–function studies to determine the role of various functional domains in the protein. Of particular interest are those regions involved in DNA binding and those that may interact with other proteins in the T4 DNA replication complex. This information, together with work being done on the crystal structure of gene 32 protein (37), should provide a more complete understanding of the role of this helix-destabilizing protein in DNA replication, recombination, and repair.

We thank Dr. Yasutsugu Nakashima for assisting with the initial phase of this work and Robert L. Mittl for determining the cysteine content of the gene 32 protein. We are especially grateful to Gary Davis for running the amino acid sequencer and the gas chromatograph and to Dr. Louis Henderson for high-pressure liquid chromatographic analyses of some of the phenylthiohydantoin derivatives. This investigation was supported in part by National Institutes of Health Grant GM12607.

† Asparagine, aspartic acid, threonine, serine, glutamine, glutamic acid, lysine, arginine, and histidine have been classified here as polar amino acids.

1. Epstein, R., Bolle, A., Steinberg, C., Kellenberger, E., Boy de la Tour, E., Chevalley, R., Edgar, R., Susman, M., Denhardt, G. & Leilausis, A. (1963) *Cold Spring Harbor Symp. Quant. Biol.* **28**, 375–392.
2. Wu, J. & Yeh, Y. (1973) *J. Virol.* **12**, 758–765.
3. Tomizawa, J., Anraku, N. & Iwama, Y. (1966) *J. Mol. Biol.* **21**, 247–253.
4. Alberts, B. & Frey, L. (1970) *Nature (London)* **227**, 1313–1318.
5. Delius, H., Mantell, N. J. & Alberts, B. (1972) *J. Mol. Biol.* **67**, 341–350.
6. Alberts, B., Amodio, F., Jenkins, M., Gutmann, E. & Ferris, F. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33**, 289–305.
7. Kelly, R., Jensen, D. & von Hippel, P. (1967) *J. Biol. Chem.* **251**, 7240–7250.
8. Alberts, B., Frey, L. & Delius, H. (1972) *J. Mol. Biol.* **68**, 139–152.
9. Reuben, R. & Gefter, M. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 1846–1850.
10. Sigal, N., Delius, H., Kornberg, T., Gefter, M. & Alberts, B. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 3537–3541.
11. Patel, G. (1978) *J. Supramol. Struct.* **8**, Suppl. 2, 323.
12. Patel, G. (1978) *J. Cell Biol.* **79**, 141a.
13. Herrick, G. & Alberts, B. (1976) *J. Biol. Chem.* **251**, 2124–2132.
14. Alberts, B. & Sternglanz, R. (1977) *Nature (London)* **269**, 655–661.
15. Mosig, G. & Breschkin, A. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1226–1230.
16. Mosig, G. & Bock, S. (1976) *J. Virol.* **17**, 756–761.
17. Huberman, J., Kornberg, A. & Alberts, B. (1971) *J. Mol. Biol.* **62**, 39–52.
18. Alberts, B., Barry, J., Bittner, M., Davies, M., Hama-Inaba, H., Liu, C., Mace, D., Moran, L., Morris, C., Piperno, J. & Sinha, N. (1977) in *Nucleic Acid-Protein Recognition*, ed. Vogel, H. (Academic, New York), pp. 31–63.
19. Glazer, A., Delange, R. & Sigman, D. (1975) in *Chemical Modification of Proteins*, eds. Work, T. & Work, E. (American Elsevier, New York), p. 104.
20. Williams, K. R., Sillerud, L., Schafer, D. & Konigsberg, W. (1979) *J. Biol. Chem.* **254**, 6426–6432.
21. Smithies, O., Gibson, D., Fanning, E., Goodfliesh, R., Gilman, J. & Ballantyne, D. (1971) *Biochemistry* **10**, 4912–4921.
22. Offord, R. (1966) *Nature (London)* **311**, 591–593.
23. Tomita, M., Furthmayr, H. & Marchesi, V. (1978) *Biochemistry* **17**, 4756–4770.
24. Williams, K. R. & Konigsberg, W. (1978) *J. Biol. Chem.* **253**, 2463–2470.
25. Carroll, R., Neet, K. & Goldthwait, D. (1975) *J. Mol. Biol.* **91**, 275–291.
26. Hosoda, J. & Moise, H. (1978) *J. Biol. Chem.* **253**, 7547–7555.
27. Anderson, R. & Coleman, J. (1975) *Biochemistry* **14**, 5485–5491.
28. Janatova, J., Fuller, J. & Hunter, M. (1968) *J. Biol. Chem.* **243**, 3612–3622.
29. Breschkin, A. & Mosig, G. (1977) *J. Mol. Biol.* **112**, 279–294.
30. Breschkin, A. & Mosig, G. (1977) *J. Mol. Biol.* **112**, 295–308.
31. Mosig, G., Berquist, W. & Bock, S. (1977) *Genetics* **86**, 5–23.
32. Spicer, E., Williams, K. & Konigsberg, W. (1979) *J. Biol. Chem.* **254**, 6433–6436.
33. Moise, H. & Hosoda, J. (1976) *Nature (London)* **259**, 455–458.
34. Jensen, D., Kelly, R. & von Hippel, P. (1976) *J. Biol. Chem.* **254**, 7215–7228.
35. Greve, J., Maestre, M., Moise, H. & Hosoda, J. (1978) *Biochemistry* **17**, 893–898.
36. Hosoda, J., Burke, R., Moise, H., Tsugita, A. & Alberts, B. (1980) *J. Supramol. Struct.* **4**, 304.
37. Chiu, W. & Hosoda, J. (1978) *J. Mol. Biol.* **122**, 103–107.