lated bacteriophages f2 and MS2. Personal communications from W. Konigsberg and H. Fraenkel-Conrat.

[4] Goldthwait, D. A., and G. R. Greenberg, in *Methods in Enzymology*, ed. S. P. Colowick and N. O. Kaplan (New York: Academic Press, 1955), vol. 2, p. 516.

[5] Rabinowitz, J. C., in *The Enzymes*, ed. P. D. Boyer, H. Lardy, and K. Myrbäck (New York: Academic Press, 1960), vol. 2, p. 185.

[6] Huennekens, F. M., P. P. K. Ho, and K. G. Scrimgeour, in *Methods in Enzymology*, ed. S. P. Colowick and N. O. Kaplan (New York: Academic Press, 1963), vol. 6, p. 809.

[7] Marcker, K., *J. Mol. Biol.*, in press.

[8] Sheehan, J. C., and D. H. Yang, *J. Am. Chem. Soc.*, **80**, 1154 (1958).

[9] Sarin, P. S., and P. C. Zamecnik, *Biochim. Biophys. Acta*, **91**, 653 (1964).

[10] Gilbert, W., *J. Mol. Biol.*, **6**, 389 (1963).

[11] Hall, R. H., and G. B. Chheda, *J. Biol. Chem.*, **240**, PC 2754 (1965).

[12] Hall, R. H., *Biochemistry*, **3**, 769 (1964).

[13] Capecchi, M. R., and G. Gussin, *Science*, **149**, 417 (1965).

[14] Nathans, D., *J. Mol. Biol.*, **13**, 521 (1965).

[15] Konigsberg, W., personal communications (1965).

[16] Weber, K., and W. Konigsberg, personal communications (1965).

[17] Webster, W., D. Engelhardt, and N. Zinder, these PROCEEDINGS, **55**, 155 (1966).

[18] Clark, B. F. C., and K. A. Marcker, *Nature*, **207**, 1038 (1965).

[19] Watson, J. D., *Bull. Soc. Chim. Biol.*, **46**, 1399 (1965).

---

# *IN VITRO PROTEIN SYNTHESIS: CHAIN INITIATION**

BY ROBERT E. WEBSTER,† DEAN L. ENGELHARDT, AND NORTON D. ZINDER

THE ROCKEFELLER UNIVERSITY

*Communicated by E. L. Tatum, November 22, 1965*

Nonsense mutations, when not suppressed, result in the premature termination of growing protein chains.[1] In a previous publication[2] we described the appearance of a suppressible protein fragment when RNA from a phage mutant was added to protein-synthesizing extracts of *Escherichia coli*. The fragment was thought to be the amino terminal portion of the phage coat protein.

In this communication, we shall present evidence that this fragment is indeed derived from the amino terminal portion of the coat protein. In addition, we shall show that, although protein isolated from phage particles has alanine with a free amino group at its amino terminus,[3] both the fragment and the whole coat protein molecules when synthesized *in vitro* have masked amino groups. The masking group is N-formylmethionine. A similar finding is reported by Adams and Capecchi.[4]

*Materials and Methods.—Bacteria and phage:* S26, a nonsuppressing (Su−) strain of *E. coli* K12;[5] f2, wild-type phage; sus4A, a mutant of f2 with a nonsense mutation in the gene specifying the coat protein.[6]

*In vitro incorporation:* The S-30 system described by Schwartz[7] was employed, but 0.03 *M* $NH_4^+$ was used in place of $K^+$. When alanine was to be incorporated, 10 mM creatine phosphate and 30 μg/ml of creatine phosphate kinase was added instead of phosphoenolpyruvate and pyruvate kinase. Uniformly $C^{14}$- and $H^3$-labeled L-amino acids were obtained from New England Nuclear. $H^3$ L-phenylalanine was also obtained from Schwarz BioResearch. The specific activities used in all calculations were those specified by the suppliers. The concentration of radioactive label in the reaction mixture was approximately $10^{-5}$ *M*. The amount of phage RNA

added was in the range of 150–300 $\mu$g/ml of reaction mixture. Approximately 10–15 lysine-$C^{14}$ molecules were incorporated into TCA-insoluble material per molecule of phage RNA added.

*Isolation of the "coat fragment":* The fragment was isolated from the reaction mixture by gel-filtration.[2] When necessary, it was further purified by electrophoresis on Whatman 3MM paper under cooled varsol at 20 v/cm for 4 hr at pH 1.9.[6] The fragment was located using a Packard model 7200 radiochromatogram scanner and was eluted from the paper with hot water. Further purification of the fragment containing methionine was accomplished, using a Sephadex G10 column at pH 2.5 (8.7 ml of acetic acid and 2.5 ml formic acid/liter of water). Molecular weight determinations were made by gel-filtration through Sephadex G-10 and Biogel P-2 in 0.05 *M* phosphate buffer at pH 6.8. Size estimates were based on the calibration of the gels which was supplied by the manufacturer.

*Electrophoretic identification of the coat fragment:* Electrophoresis on cellulose acetate strips (5 $\times$ 20 cm Sepraphore III) was performed at pH 1.9 (87 ml of acetic acid and 25 ml formic acid/liter of water), at pH 5.0 (0.03 *M* potassium acetate), at pH 6.5 (0.05 *M* potassium phosphate), and at pH 10 (0.03 *M* glycine). The field strength was 11.5 v/cm. Following electrophoresis for 1 hr, strips were dried by blotting and cut into 0.5-cm segments. Each segment was dissolved in 10 ml of Bray's solution (100 gm naphthalene, 7 gm PPO, and 0.05 gm POPOP/liter of diethylene oxide) and counted in a Nuclear-Chicago model 6804 scintillation counter.

*Unmasking of N-terminal group:* The N-terminal-formyl group of the chymotryptic "tetrapeptide" (*D*) and the "coat fragment" was removed by treatment at room temperature with 1 *N* HCl in methanol.[8] After treatment, the mixture was neutralized with a concentrated solution of ammonium bicarbonate and dried by lyophilization.

*Enzymic digestion and product analysis:* Fingerprinting and radioautography of chymotryptic digests of coat protein and *in vitro* product, as well as counting of carbon and tritium eluted from fingerprint spots, was performed as previously described.[2] Pronase (K. and K. Laboratories) digestion of the "coat fragment" was carried out for 12 hr in 1% ammonium bicarbonate. The products were identified electrophoretically at pH 5.0 (0.03 *M* potassium acetate). Marker $C^{14}$-labeled N-formylmethionine and $H^3$-labeled N-formylalanine were prepared by the method of Waley and Watson[9] and purified by electrophoresis on Whatman 3MM paper at pH 1.9.

*Results.—The N-terminal fragment:* The sequence of the N-terminal portion of the coat protein of f2 is $NH_2$ala-ser-asp$NH_2$-phe-thr-glu$NH_2$-phe-val....[3] The tetrapeptide ala-ser-asp$NH_2$-phe and the tripeptide thr-glu$NH_2$-phe are released from coat protein by chymotryptic digestion (peptides *B* and *A*, respectively, in Fig. 1 of Engelhardt *et al.*[2]). When the nonsense mutation sus4A is suppressed by the bacterial suppressor gene Su-1, a serine residue is substituted for the glutamine residue in the tripeptide (*A′*). These three chymotryptic peptides can be separated and identified in radioautograms of fingerprints of digests of coat protein from purified phage (*in vivo* protein). There is only one other soluble phenylalanine-containing chymotryptic peptide, and it is only slowly released during digestion.

The major product of *in vitro* protein synthesis stimulated by f2 RNA is coat protein.[10] Fingerprints of chymotryptic digests of *in vitro* product would be expected to show peptides homologous with those of *in vivo* protein. However, of the two major spots seen (Fig. 1), one can be identified as the tripeptide (*A*) while the other is a new peptide, *D*, which is uncharged at pH 1.9 and has no *in vivo* homologue. The tetrapeptide, *B*, is not present at its expected position, slightly below the tripeptide. It was surprising to find this peptide missing since, hitherto, a homologue for each peptide derived from *in vivo* protein was found in the digests of *in vitro* product. Peptide *D* was therefore considered likely to be a modified form of the tetrapeptide. Since the tetrapeptide is the N-terminal peptide, its lack of charge at pH 1.9 could be accounted for if its N-terminal amino group were covered.

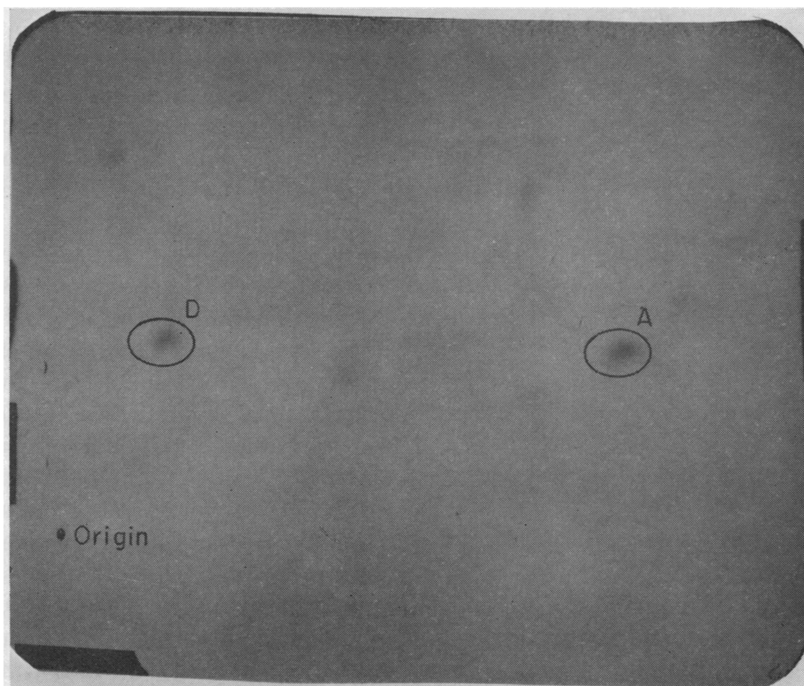


FIG. 1.—Radioautogram of the chymotryptic peptides from $C^{14}$-phenylalanine-labeled *in vitro* product stimulated by f2 RNA. *D* is the "modified tetrapeptide" and *A* is the tripeptide. First-dimension electrophoresis pH 1.9; second-dimension chromatography in butyl acetate.

In order to study in detail the N-terminal portion of the *in vitro* product, advantage was taken of the fact that the nonsense mutation in phage sus4A is in the codon specifying the sixth amino acid from the amino terminus. RNA from this mutant in Su- extracts stimulates the formation of a suppressible fragment[2] which should be the N-terminal portion of the coat protein. The fragment has the same electrophoretic mobility as peptide *D* at pH 1.9, and chymotryptic digestion of the fragment yields material that cochromatographs with peptide *D*.[2] The fragment should therefore be R-ala-ser-asp$NH_2$-phe-thr.

The following experiments on the composition and size of the fragment support this hypothesis. To determine the amino acid content of the fragment, different $C^{14}$ amino acids were added one at a time to Su- extracts with RNA from sus4A. To increase the amount of fragment, the extracts were supplemented with 100 $\mu$g of Su- sRNA per ml. A standard for locating and analyzing the fragment was provided by adding simultaneously $H^3$-phenylalanine to each extract. Figure 2 shows the elution profile of the product made in the presence of $C^{14}$-valine and $H^3$-phenylalanine. The coat fragment, defined by the $H^3$ in the 185th to 193rd ml of effluent appears not to contain valine. The material in these tubes was pooled and the solvent was evaporated. An electropherogram at pH 1.9 of the residue is shown in Figure 3*a;* there is less than 0.25 $\mu\mu$moles of valine for each $\mu\mu$mole of phenylalanine. It is concluded that the fragment does not contain valine. An electropherogram (Fig. 3*b*) of fragment prepared in the presence of $C^{14}$-threonine and $H^3$-phenylalanine gave a different pattern; phenylalanine and threonine are

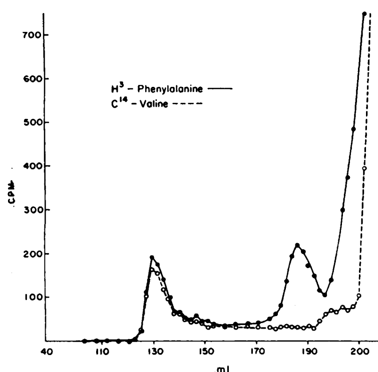


FIG. 2.—Gel-filtration of acetic acid-soluble products of an *in vitro* synthesizing system. Sus4A RNA in Su- extract was used. The radioactive amino acids added were $C^{14}$-valine and $H^3$-phenylalanine. The "coat fragment" is in the 185th to 193rd ml of effluent.

present in almost equimolar concentration (phenylalanine/threonine = 0.95). By repeated use of this procedure, the fragment was shown to contain one molecule of alanine, serine, asparagine, and threonine for each molecule of phenylalanine. It was shown not to contain lysine, arginine, histidine, leucine, or valine. [At the time that these experiments were initiated, the details of the composition and sequence of the N-terminal portion of the coat protein were unknown. The results of these experiments provided a valuable clue for placing in order the six chymotryptic peptides derived from the large N-terminal tryptic peptide.]

The size of the fragment was measured by gel-filtration (see *Methods*), using bovine serum albumin and a synthetic tripeptide (phe-ser-thr) as reference markers. Figure 4 shows the elution profile from Sephadex G-10. The results obtained from gel-filtration give a molecular weight of 600–700 for the fragment. This is closely comparable to the molecular weight (>540) of a masked pentapeptide with the composition R-ala-ser-asp$NH_2$-phe-thr. In a subsequent publication,[11] we will present evidence that the sequence of the fragment is as written above. These results on the size and composition of the fragment together with the fact that production of the peptide is suppressible support the hypothesis that the fragment is indeed the N-terminal portion of the coat protein. Therefore, we conclude that synthesis of the coat protein in Su- extracts with mutant RNA proceeds from the amino terminal end up to the position specified by the nonsense mutation.

*Masking group:* Characterization of the soluble N-terminal fragment of the coat protein permitted further investigation of the nature of the masking group. The formyl group is one such masking group and is unique among acyl masking groups in that it can be removed by mild acid hydrolysis.[12] Removal of a formyl

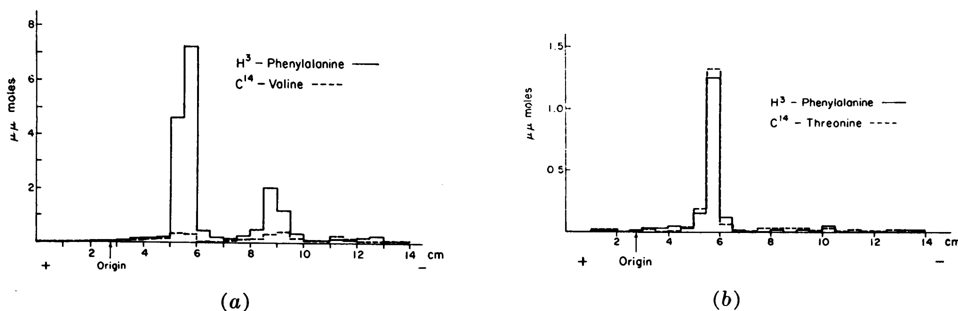


FIG. 3.—Electrophoresis of the "coat fragment" on cellulose acetate at pH 1.9. The fragment is at the position marked 5–6 cm. (*a*) The pattern when $C^{14}$-valine, $H^3$-phenylalanine had been added to the extracts; (*b*) the pattern when $C^{14}$-threonine, $H^3$-phenylalanine had been added to the extracts.

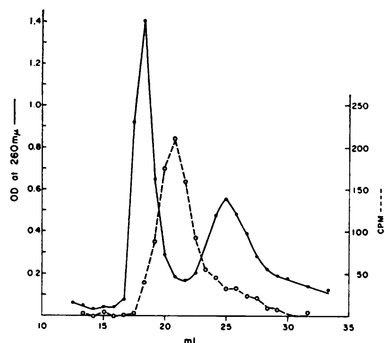


FIG. 4.—Gel-filtration of purified "coat fragment" on Sephadex G-10 for size determination. The fragment is labeled with $C^{14}$-phenylalanine. Bovine serum albumin and a tripeptide (phe-ser-thr) are present as reference markers. The buffer used was 0.05 *M* sodium phosphate pH 6.8.

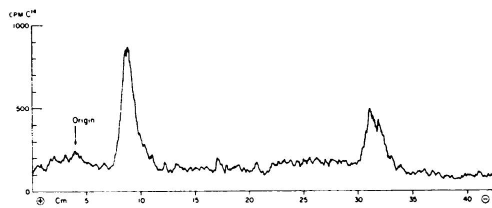


FIG. 5.—Paper electrophoresis of partially deformylated "coat fragment" at pH 1.9. The fragment was labeled with $C^{14}$-alanine and $H^3$-phenylalanine and purified by gel-filtration and paper electrophoresis prior to mild acid hydrolysis. After electrophoresis, the $C^{14}$ pattern was obtained using a Packard radiochromatogram scanner. $H^3$ was counted following elution with hot water.

group from the fragment should be revealed by a more electropositive mobility at pH 1.9. The fragment was labeled with $C^{14}$-alanine and $H^3$-phenylalanine, purified by gel-filtration and paper electrophoresis, and treated with 1 *N* HCl in methanol[9] for 1.5 hr at room temperature. After neutralization with $NH_4HCO_3$, the solvent was evaporated and the residue was dissolved and subjected to paper electrophoresis at pH 1.9. Figure 5 shows that approximately half of the material became positively charged: the amino group had been unmasked. The positive charge did not result from hydrolysis of any of the peptide bonds, since the $C^{14}/H^3$ ratios are the same for the formylated and the deformylated fragments.

Marcker and Sanger[13] have isolated N-formylmethionyl-sRNA from *E. coli* and also found that methionyl sRNA can be formylated in cell-free extracts. Adams and Capecchi[4] present evidence for the incorporation of a formyl group into phage coat protein via this N-formylmethionyl-sRNA. This suggested that the masking group on the coat fragment was N-formylmethionine. Mutant RNA was added to Su- extracts containing $C^{14}$-methionine and $H^3$-phenylalanine and the products separated by gel-filtration (Sephadex G-25). At the position characteristic of the fragment, a large amount of positively charged $C^{14}$-labeled material was found. The fragment was separated from this material (assumed to be S-adenosylmethionine) by paper electrophoresis at pH 1.9. Formylmethionine, also known to be synthesized by the extract, has an electrophoretic mobility at pH 1.9, similar to that of the fragment. Therefore, to make certain that the methionine was in the fragment, further analysis was by gel-filtration. It can be seen (Fig. 6) that methionine and phenylalanine are both in the fragment. Had any formylmethionine been present, it would have been at the position of the reference peptide. In addition, when the purified fragment was subjected to electrophoresis on cellulose acetate at pH 1.9, 5.0, and 10.0, methionine and phenylalanine were found in equimolar ratio.

In an attempt to release free formylmethionine, the fragment was labeled with $C^{14}$-methionine and $H^3$-alanine, purified and digested with pronase. The digestion

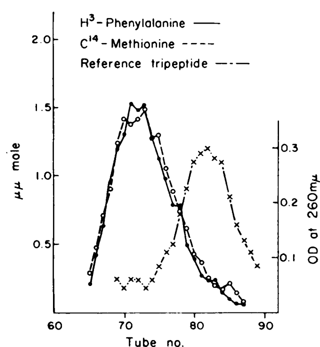


FIG. 6.—Gel-filtration of purified "coat fragment" on Sephadex G-10 at pH 2.5. The fragment is labeled with $C^{14}$-methionine and $H^3$-phenylalanine. The tripeptide phe-ser-thr is present as a reference marker.

was subjected to electrophoresis on cellulose acetate at pH 5.0, a pH at which free amino acids, formylalanine, and formylmethionine are all separable. Neither isotopic label was found at the position corresponding to formylalanine, but a small (5–10%) $C^{14}$ (methionine) peak containing no $H^3$ (alanine) was found at the N-formyl(methionine) position. The majority of the product, however, was composed of free amino acids and incompletely digested fragments that had retained their formyl groups. Further digestion was not profitable because of the deacylase activity of the pronase preparation.

It is probable that the chymotryptic peptide *D* (Fig. 1) also contains formylmethionine when derived from complete protein chains, as well as from the fragment. Complete protein chains were synthesized using wild-type RNA in the presence of $C^{14}$-methionine and $H^3$-phenylalanine, and were isolated by gel-filtration, digested with chymotrypsin, and fingerprinted in the usual manner. Radioautograms revealed a $C^{14}$ (methionine) spot in the position of peptide *D*, containing $C^{14}$ and $H^3$ in an equimolar ratio. Treatment of the chymotryptic digest with 1 *N* HCl in methanol prior to fingerprinting and radioautography yields a new $C^{14}$ and $H^3$ peptide, *D'*, at a position expected for neutral peptides. The formyl group had been removed, probably releasing the pentapeptide meth-ala-ser-asp$NH_2$-phe.

*Summary and Conclusions.*—In extracts of *E. coli*, RNA from the bacteriophage can direct the synthesis of whole phage coat protein molecules. RNA from phage containing a nonsense mutation in the gene specifying the coat protein gives rise instead to a peptide fragment. This fragment has been shown to be the amino-terminal portion of the coat protein, prematurely terminated in its growth at the position specified by the nonsense mutation.

As synthesized in our extracts, both the fragment and the complete protein have the amino group of their N-terminal alanine residue covered by N-formylmethionine. Since this alanine residue has a free amino group when obtained from protein of phage particles, our extracts must be deficient in a way that allows the N-formylmethionine to be retained on protein. Thus, the deficiency of these extracts may have revealed a fundamental principle of protein synthesis—initiation of translation at a unique nucleotide sequence.

Although synthetic polynucleotides of known and random sequence can all stimulate protein synthesis in extracts, it would be advantageous for natural messages such as phage RNA to have a nucleotide sequence that specifies the beginning of protein chains. It would guarantee that the message starts to be read at the right position and thereafter is read in phase.

Several findings, combined with those described above, indicate that there is a unique nucleotide sequence that specifies the amino-terminal portion of all *coli* proteins. (1) Messenger translation proceeds in a $5' \rightarrow 3'$ direction along a polynucleotide chain[14] and most RNA's synthesized using *E. coli* DNA as a template for RNA polymerase have the same $5'$ nucleotide.[15] (2) A particular methionine

sRNA, upon which methionine can be formylated, has been isolated from uninfected *E. coli.*[13] (3) Waller[16, 17] has shown that many *E. coli* proteins have either methionine, alanine, or serine as their N-terminal amino acids. Note that these amino acids are precisely the ones found in the N-terminal peptide of the phage RNA stimulated products. Perhaps after synthesis *in vivo*, the *coli* proteins are deacylated and amino acids removed by specific amino-peptidases, just as the N-formylmethionine must have been removed from the phage coat protein. During or after deacylation how many, if any, of the amino acids are removed could depend on the proximal amino acids in and/or the tertiary structure of the different proteins.

The start signal, implied by the data in this and the accompanying paper,[4] specifies at a minimum N-formylmethionine and perhaps even N-formylmethionyl-alanyl-serine. Its importance lies not only in the fact that there is a unique series of nucleotides meaning "start here," but also in the fact that these nucleotides lead to the incorporation of an amino acid with a blocked amino group. This kind of a start signal has numerous consequences. It would impose a direction on protein synthesis. It would preclude the coupling of polypeptides specified by a polycistronic message, regardless of the nature of "stop" codons. Similarly, it would preclude coupling of polypeptides when nonsense codons, assumed to be homologous to stop codons, are suppressed. It would explain why such nonsense codons, when not suppressed, lead only to the production of an N-terminal fragment, since peptide synthesis can not restart. It would predict the existence of another, perhaps suppressible, class of nonsense mutations—mutations to the N-formylmethionine codon in the middle of a gene. These mutations would probably result in the production of two peptide fragments. By leading to polypeptide chains that may require trimming prior to being able to function, it would offer a new dimension to the control, in time and space, of protein function.

* Supported in part by a grant from the National Science Foundation.

† National Science Foundation postdoctoral fellow.

[1] Sarabhai, A. S., A. D. Stretton, S. Brenner, and A. Bolle, *Nature*, **201**, 13 (1964).

[2] Engelhardt, D. L., R. E. Webster, R. Wilhelm, and N. D. Zinder, these PROCEEDINGS, **54**, 1791 (1965).

[3] Konigsberg, R., G. Notani, and N. D. Zinder, in preparation.

[4] Adams, J., and M. R. Capecchi, these PROCEEDINGS, **55**, 147 (1966).

[5] Weigert, M. G., and A. Garen, *J. Mol. Biol.*, **12**, 448 (1965).

[6] Notani, G. W., D. L. Engelhardt, W. Konigsberg, and N. D. Zinder, *J. Mol. Biol.*, **12**, 439 (1965).

[7] Schwartz, J. H., these PROCEEDINGS, **53**, 1133 (1965).

[8] Waley, S. G., *Chemistry and Industry*, **107** (1953).

[9] Waley, S. G., and J. Watson, *Biochem. J.*, **57**, 529 (1954).

[10] Nathans, D., G. Notani, J. H. Schwartz, and N. D. Zinder, these PROCEEDINGS, **48**, 1424 (1962).

[11] Engelhardt, D. L., R. E. Webster, and N. D. Zinder, in preparation.

[12] Sheehan, J. C., and Ding-Djung H. Yang, *J. Am. Chem. Soc.*, **80**, 1154 (1958).

[13] Marcker, K., and F. Sanger, *J. Mol. Biol.*, **8**, 835 (1965).

[14] Thach, R. E., M. A. Cecera, T. A. Sundararajan, and P. Doty, these PROCEEDINGS, **54**, 1167 (1965).

[15] Maitra, U., and J. Hurwitz, these PROCEEDINGS, **54**, 815 (1965).

[16] Waller, J. P., *J. Mol. Biol.*, **7**, 483 (1963).

[17] *Ibid.*, **10**, 319 (1964).