CHANGE OF A SEQUENCE OF AMINO ACIDS IN PHAGE T4 LYSOZYME BY ACRIDINE-INDUCED MUTATIONS*

BY E. TERZAGHI,[†] Y. OKADA, G. STREISINGER, J. EMRICH, M. INOUYE, AND A. TSUGITA

INSTITUTE OF MOLECULAR BIOLOGY, UNIVERSITY OF OREGON, EUGENE, AND LABORATORY OF MOLECULAR GENETICS, UNIVERSITY OF OSAKA, JAPAN

Communicated by Seymour Benzer, June 20, 1966

Genetic analysis of acridine-induced mutations by Crick et al.¹ has shown that mutations of this kind can be classified into one of two different sets. When combined as a double mutant, mutations of the same set produce the same defective phenotype as either of the single mutations, whereas two mutations belonging to different sets often suppress each other, resulting in a wild or pseudowild phenotype.

A previous proposal by Brenner *et al.*² that acridines induce mutations by causing deletion or insertion of ^a single nucleotide in DNA suggested the identification of the two sets of mutations with, in the one case, deletions, and, in the other case, insertions of single nucleotides. On the basis of this interpretation the ability of mutations of the opposite set to suppress each other could be explained by assuming that the nucleotide sequence of a cistron is translated into its protein product by the sequential reading of triplets of bases (in messenger RNA) from a fixed starting point which corresponds to the beginning of the cistron.¹ The genetic message was thus imagined to be separated into groups of three bases by a "reading frame" which is set in register at the beginning of the message. The addition or deletion of a base would thus cause a shift in the reading frame and would result in a grossly different translation of the message beyond the site of the mutation. On the other hand, if a deletion were followed by an insertion (or vice versa), as would be the case for a double mutant composed of two acridine mutations in different sets, the reading frame, and thus the amino acid sequence of the protein product, would be altered only in the region between the mutations. If the mutations were close to one another, the protein product of the cistron might in some cases be active.

Crick's hypothesis might be tested directly if acridine mutations could be induced in a cistron which controls the primary structure of an isolable protein. In this communication we report the results of such a test based on studies of the lysozyme of bacteriophage T4.

Materials and Methods.³—Bacterial strains Escherichia coli B and $B/1⁴$ and bacteriophage strains T4 $ac q$,⁵ T4 $ac q$ eJ44,⁶ and T4 $ac q$ eJ42⁶ were used. Sodium Fraser medium consisted of NaH2PO4-H20, 4.5 gm; Na2HP04, 10.5 gm; NH4Cl, 3 gm; Difco casamino acids, 15 gm; glycerol, 24 ml; gelatin, 30 mg (in solution); H_2O , 1 liter; the pH is adjusted to 7.0.7

Preparation of lysates for lysozyme isolation:⁸ E. coli B/1 was grown in aerated sodium Fraser medium to a concentration of 4×10^8 per ml at 37°; the medium was supplemented with tryptophan (20 μ g per ml) and infected with about 10⁶ phage per ml. Incubation was continued with aeration at 37° (e^+) or else 30° $(eJ42eJ44)$ for about 4 hr. The lysates were then collected and stored in the cold for at least 18 hr. Lysates were grown in 20-liter batches in an apparatus modified from one first described by Fraser.9

Purification of lysozyme:⁸ Rivanol (6,9-diamino-2-ethoxyacridinelactate \cdot H₂O, 0.25 mg per ml) was added to the lysates. The precipitate (containing DNA, bacterial debris, and acidic proteins) was removed by filtration through cheesecloth (Woolworth) followed by centrifugation. The lysate was diluted twofold with H20 and applied to a column containing IRC 50 (AG, Biorad coarse grade). Lysozyme was eluted from this column with phosphate buffer $(0.1 \, M, pH 7.0)$ containing NaCl $(0.55 \, M)$, and was purified by two further fractionations on IRC 50 columns, elution in each case utilizing phosphate buffer $(0.1 \, M, pH 7.0)$ and NaCl gradients $(0-0.65 \, M)$. The lysozyme was finally fractionated twice on Sephadex G75 columns in phosphate buffer containing NaCl $(0.45 M)$. The purification was carried out at $4^{\circ}.10$

Lysozyme activity was measured by noting the rate of decrease of the optical density of lyophilized preparations of E. coli B.8

Performic acid oxidation: The oxidation was carried out at -7° using the procedure described by Hirs.¹¹

Tryptic digestion: Trypsin $(2) \times$ crystallized, Worthington) was treated with diisopropylfluorophosphate (DFP).12 Digestion was carried out with an enzyme-substrate ratio of 1:100; the pH was maintained at 8.0 with additions of dilute KOH by use of an autotitrator. When the digestion had reached apparent completion, a further aliquot of trypsifi was added and the reaction was finally stopped by adjusting the pH to 2.5 by the addition of formic acid. An insoluble core was removed from the digest by centrifugation.

Peptide column chromatography: Tryptic digests of oxidized lysozyme were applied to 0.9 \times 150 cm columns of Dowex 50-X2 and eluted using a gradient of pyridine-acetate buffer of increasing pH and pyridine concentration established by means of an eight-chamber varigrad apparatus. Final elution was with the last buffer of the system at 50° , followed by 0.5 N ammonium acetate. The composition of the buffers, the temperature of the column, and the entire elution procedure were as described by Canfield."3 Fractions of 4.0 ml were collected and the flow rate was 40 ml per hour.

Assay of eluate fractions: Aliquots of 0.5 ml were hydrolyzed under alkaline conditions as described by Hirs et $al.,¹⁴$ and were assayed by use of the potassium cyanide-ninhydrin method described by Yemm and Cocking,¹⁵ the absorbance at 570 m μ being measured.

Amino acid analysis: Samples of suitable size were hydrolyzed with constant-boiling HCl (at 105° for 24 hr) and were analyzed in a Beckman-Spinco automatic analyzer.

Carboxypeptidase digestion: DFP-treated carboxypeptidase A (Worthington) and carboxypeptidase B (Worthington) were dialyzed ¹² hr at ⁴⁰ against water, and were then stored frozen. Digestion was carried out with 2.5 μ g enzyme in 0.2 ml dilute bicarbonate buffer (NaHCO₃, 1.68) gm; NaCl, 5.85 gm; H₂O, 1 liter) for 5 hr at 37° . The reaction was terminated by the addition of 0.8 ml pH 2.2 buffer (as supplied with the Beckman Spinco amino acid analyzer) and an aliquot was applied to the amino acid analyzer.

Paper chromatography: Whatman No. 1 paper and n-butanol-pyridine-acetic acid-water (15:10:3:12)16 were used.

Chymotryptic digestion: Digestion of peptides was carried out with 5 μ g chymotrypsin (Worthington) in 0.2 ml $N/15$ phosphate buffer, pH 8.0, for 12-18 hr. The cleavage products were separated by means of paper electrophoresis for 3 hr at about 20 v/cm using pyridine-acetic acidwater (10:0.4:90) at pH 6.4.

Edman degradation: An aliquot of the peptide was treated according to the method of Konigsberg and Hill,¹⁷ the recovered peptide (lacking the N-terminal amino acid) being hydrolyzed and analyzed as described above.

Results.—The isolation of the pseudowild double mutant $eJ42eJ44$: In order to recover a pseudowild double mutant, pairwise crosses between a number of closely linked, independently isolated, proflavine-induced mutants were performed. A cross of the mutants eJ42 and eJ44 yielded two recombinant types, wild type and pseudowild.'8 In order to show the pseudowild recombinant type to be a double mutant, crosses of wild type by pseudowild were performed. These crosses yielded recombinants that were mutant; all of these were either eJ42 or eJ44. The pseudowild recombinant (isolated from among the progeny of the cross of $eJ42 \times eJ44$) is thus the double mutant eJ42eJ44.

Lysozyme activity and the amount of lysozyme of e^+ and eJ42eJ44: In order to compare the lysozyme activity in bacteria infected with e^+ and $eJ42eJ44$ phage,

FIG. 1.-Elution patterns on Dowex 50-X2 columns of tryptic digests of performic acid oxidized e^+ and eJ42eJ44 lysozyme. Aliquots from alternate tubes were subjected to alkaline hydrolysis, treated with ninhydrin, and the absorbance at 570 m μ was then measured.

 $#8, 50^{\circ}$ indicates the application of the last buffer of the series to the column, the temperature being raised to 50°

Peak 5 varies in size from preparation to preparation and contains an undigested fragment consisting of the peptides in peaks 2 and 3.

standard lysates of these types, as well as of eJ42 and eJ44 phage, were prepared. The double-mutant phage lysates exhibited approximately half of the lysozyme activity of wild-type lysates; in contrast, the lysates of the single-mutant phage contained no measurable levels of activity (i.e., less than 5% of e^+).

The specific activity of the purified double-mutant lysozyme was about half of that of the wild type. There appears to be no great difference in the amount of lysozyme produced by wild-type and by double-mutant phage.

Separation and the amino acid analysis of the tryptic peptides of e^+ and eJ42eJ44: In order to detect differences between the primary structures of the lysozyme of $e⁺$ and $eJ42eJ44$, tryptic digests of the two lysozymes were chromatographed on Dowex 50-X2 resin as shown in Figure 1. In each digest 18 major peaks were resolved, 17 of these apparently being identical in the two digests. The only significant difference observed in the two chromatograms is the presence of peak 10 in only the wild-type digest and the presence of peak 19 in only the doublemutant digest.

The amino acid composition of material from each of the other peaks was found to be the same. For some of the peaks the material was further resolved into several components by paper chromatography, and in each of these cases the resulting chromatographic patterns were essentially the same for corresponding peaks from the wild-type and double-mutant digests.

The peaks labeled α and β in the double-mutant digests yielded nonintegral ratios for a large number of amino acids, and the molar recoveries of the amino acids (relative to those in other peaks) were very low; these peaks are judged to

TABLE ¹

SEQUENCE OF AMINO ACIDS OF PEPTIDE 10 FROM e^+ LYSOZYME^{*}

* Some of the analyses were performed on peptide 10 isolated from an independently prepared batch of e^+ lysozyme. t Recovered as the acid after acid hydrolysis and as the amide after enzymatic digestion. ^X Molar recovery of liberated amino acids.

TABLE ²

SEQUENCE OF AMINO ACIDS OF PEPTIDE 19 FROM eJ42eJ44 LYSOZYME

* Recovery of liberated amino acids.

† Molar recovery of liberated amino acids.

† This sequence is further confirmed by the following composition of a peptide isolated after chymotryptic

digestion of the whole lysozyme:

represent relatively small amounts of large, contaminating peptides, or else to represent a mixture of several peptides, each present in relatively small amounts.

Amino acid sequence difference in e^+ and eJ42eJ44 lysozymes: Peak 10 and peak 19 could be shown to contain one peptide each. As shown in Table 1, the sequence of amino acids of the peptide that is present only in the e^+ lysozyme is Ser-Pro-Ser-Leu-Asn-Ala-Ala-Lys, while the sequence of amino acids of the peptide that is present only in the double-mutant lysozyme (Table 2) is Val-His-His-Leu-Met-Ala-Ala-Lys.

The lysozyme of the double-mutant strain thus differs from that of the wildtype strain by a sequence of at least five amino acids.

Studies aimed at determining the sequence of amino acids in the whole wild-type lysozyme molecule19 indicate that a major component of peak 18 represents the peptide that is N-terminal to the peak 10 peptide described above. This major component of peak 18 was found to be identical in e^+ and $eJ42eJ44$ digests, suggesting that the change in amino acid sequence is limited to the five amino acids described above.

The amino acid composition and sequence of a number of peptides resulting from the chymotryptic digestion of e^+ and $eJ42eJ44$ lysozyme have been determined. One of these chymotryptic peptides from the e^+ digest was found to have the amino acid sequence Thr-Lys-Ser-Pro-Ser-Leu-Asn, while a chymotryptic peptide from eJ42eJ44 digests was found to have the amino acid composition (Thr-Lys-Val-His-His-Leu). Preliminary results indicate that the amino acid sequence of the peptide is as shown. This establishes that the change in the primary structure of the eJ42eJ44 lysozyme is confined to a sequence of five amino acids.

That the change is in fact confined to the five amino acids described above is further confirmed by the analysis of strains that will be described elsewhere.

 $Discussion$. We have observed that the lysozymes produced by wild-type phage T4 and pseudowild T4 carrying two proflavine-induced mutations differ by a sequence of five amino acids. This result is a verification of a major prediction of the hypothesis advanced by Crick and his associates,' that the translation of the genetic message is initiated at a given point and proceeds by the sequential reading of defined groups of nucleotides. The result also verifies the hypothesis that proflavine acts as a mutagen by inserting or deleting bases.

Nirenberg and his co-workers,²⁰ and Khorana and his co-workers,²¹ on the basis of in vitro studies, have proposed specific triplets that code for each of the 20 amino acids. Using the triplets proposed by these workers, a sequence of triplets can be written that would code for the five relevant amino acids observed in the wild-type lysozyme and that, by the addition and deletion of a single base as indicated below, would generate a new set of triplets that would code for the five new amino acids observed in the double-mutant strain. The two sequences of amino acids and the related sequences of triplets are:

We wish to point out that (1) triplets compatible with the two sequences of amino acids can be generated only if the triplets are oriented such that the ⁵' to 3' polarity of the triplets parallels the N-terminal to C-terminal polarity of the amino acids; (2) the sequence of triplets that codes for the changed amino acid sequence of the double-mutant strain can be generated from the wild-type sequence of triplets only if one of the mutations results in the deletion of a single base near the ⁵' end of the altered region and the other mutation results in the addition of aguanine or adenine residue at the position near the 3' end shown above; (3) the sequences of triplets described above are the only ones that relate the two sequences of amino acids.

To illustrate these points we define two sets of mutations, a " $+$ " set and a "-" set, such that a mutation of either set followed by a mutation of the other set results in the restoration of the reading frame beyond the second mutation. Some members of these sets are:

$u + v$ Set	$u - v$ Set	
$(+ + + +)$	$(+)$	$(+ + + + +)$
$(- -)$	$(- -)$	
$(- - -)$	$(- - - -)$	
etc.	etc.	

where, for example, $(++ + +)$ represents the insertion of four bases and $(- - - -)$ the deletion of four bases. The sets are so composed that a combination of a "+" member with a " $-$ " member always produces a net loss or gain of $3N$ bases, where N is any integer including zero, " $+$ " members producing a frame shift of one base to the left and " $-$ " members a frame shift of one base to the right. Some combinations of " $+$ " and " $-$ " mutations will result in an increase or decrease in the total number of amino acid residues in the product protein. Since our altered lysozyme peptide shows no loss or gain of residues, such combinations can be ignored in the following discussion.

We consider next the two sequences of amino acids in the altered region of the protein and all of the codons that have been proposed for them. These are shown in Table $3A$ where, in the upper part of the table, the bases of the codons have been oriented such that the ⁵' to 3' polarity parallels the N-terminal to C-terminal polarity of the peptide, and in. the lower part of the table, with the opposite polarity. Table 3B lists the amino acid replacements which are incompatible with the possible " $+$ ", " $-$ " combinations (up to the deletion or insertion of four bases) for the two orientations of the nucleotide chain with respect to the peptide. It is seen that the only case which results in no incompatible replacement is the one in which the ⁵' to ³' direction of the codons parallels the N to C direction of the peptide and which involves the deletion and insertion of one base only near the 5' and 3' ends, respectively.

Since it has been shown that the synthesis of polypeptide chains proceeds from the N-terminal towards the C-terminal end, our results demonstrate that the sequence of nucleotides in the mRNA is translated from the ⁵' toward the ³' end. Recent results of other investigators are in accord with this view.22, ²³

The precise location of the base deleted near the ⁵' end of the region specifying the changed amino acids cannot be determined since the deletion of any one of five bases in this region would be compatible with the observed amino acid sequences. The base added near the ³' end of the region specifying the changed amino acids can be only guanine or adenine.

The results presented here indicate that the following codons, determined, or else proposed, on the basis of in vitro studies, are in fact utilized by E . *coli in vivo*:

B. Forbidden Amino Acid Replacements

* A (-) mutation (i.e., the loss of a single base) followed by a (- -) mutation (the loss of two bases) can be excluded because of the amino acid incompatibilities indicated, and, in addition, because no amino acid is los

^t No codon for proline generates any codon for histidine with the indicated direction of frame shift. ^t No pair of codons for serine and proline (in that order) generates any pair of codons for valine and histidine

(in that order).

Ser-AGU, UCA; Pro-CCA; Leu-CUU, UUA; Asn-AAU; Val-GUC; His-CAU, CAC. The codons that are underlined are present in the wild-type T4B strain. It is noteworthy that each of the five codons that we have identified in the wildtype strain ends in either A or U. This is not surprising since the DNA of phage T4 is relatively rich in AT base pairs.

Two different codons can be assigned to serine in the wild-type DNA, and two different codons can be assigned to histidine and leucine, showing degeneracy both for codons in the wild-type DNA and for the in vivo code.

Summary.-The primary structures of the lysozymes produced by wild-type phage and by pseudowild phage carrying two proflavine-induced mutations differ with respect to a sequence of five amino acids.

Utilizing current codon assignments, a unique sequence of nucleotides was found to relate the two sequences of amino acids.

Our results confirm the hypothesis of Crick and his associates regarding the general nature of the genetic code and are compatible only with a 5' to 3' direction of translation of the messenger RNA.

The authors are grateful to F. H. C. Crick for numerous discussions, to A. Novick, F. W. Stahl, and S. Champe for suggestions concerning the manuscript, and to Donald Thomas and Thomas Warner for their help in the preparation and purification of lysozyme samples.

* Part of this work was submitted by E. T. in partial fulfillment of the requirements for the Ph.D. degree, University of Oregon. This investigation was supported by grant 466 GB2261 (National Science Foundation) to G. S., and by a Jane Coffin Memorial Fund for Medical Research grant and grant GM-10982 (National Institutes of Health) to A. T.

^t Predoctoral trainee, grant 66 2T1-GM715-06 (PHS). Present address: Laboratory of Molecular Genetics, University of Osaka, Japan.

¹ Crick, F. H. C., L. Barnett, S. Brenner, and R. J. Watts-Tobin, Nature, $192\frac{1}{2}27$ (1961).

² Brenner, S., L. Barnett, F. H. C. Crick, and A. Orgel, J. Mol. Biol., 3, 121 (1961).

³ Procedures for stock preparation and crosses will be described elsewhere.

4B/1, a strain resistant to phage T1, was used for large cultures in order to avoid otherwise frequent contamination by Tl-like phages.

⁵ The markers ac and q enable phage to grow in the presence of otherwise inhibitory concentrations of acridine dyes and were present in all stocks to facilitate acridine mutagenesis.

⁶ Streisinger, G., J. Emrich, J. Newton, K. Vessey, E. Terzaghi, and K. Fields, manuscript in preparation.

⁷ Modified after Fraser, D., and E. Jerrel, J. Biol. Chem., 205, 291 (1953).

81nouye, M., A. Tsugita, E. Terzaghi, and G. Streisinger, manuscript in preparation.

⁹ Fraser, D., J. Bacteriol., 61, 115 (1951).

¹⁰ About 0.7 mg of purified lysozyme is recovered per liter of original lysate; the recovery of lysozyme activity is about 25%.

¹¹ Hirs, C. H. W., J. Biol. Chem., 219, 611 (1956).

¹² Hartley, B., quoted by Potts, J. T., A. Berger, J. Cooke, and C. B. Anfinsen, J. Biol. Chem., 237, 1851 (1962). DFP was added at a concentration that inactivated about 10% of the trypsin activity; this treatment preferentially inactivates contaminating chymotryptic activity.

¹³ Canfield, R. E., J. Biol. Chem., 238, 2691 (1963).

¹⁴ Hirs, C. H. W., S. Moore, and W. H. Stein, J. Biol. Chem., 219, 623 (1956).

¹⁵ Yemm, E. W., and E. C. Cocking, Analyst, 80, 208 (1955).

16Waley, S. G., and J. Watson, Biochem. J., 55, 328 (1953).

17Konigsberg, W., and R. J. Hill, J. Biol. Chem., 237, 2547 (1962).

¹⁸ Progeny of the crosses were plated on tryptone plates and after 8 hr of incubation the plates were exposed to chloroform vapors. Plaques with large halos, identical to those formed by e^+ phage, were classified as wild type; plaques with small halos were classified as pseudowild. The per cent of recombinants observed and the relationship between the per cent of recombinants and the expected physical distance will be discussed elsewhere.

¹⁹ Tsugita, A., and M. Inouye, manuscript in preparation.

²⁰ Nirenberg, M., P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, and C. O'Neal, these PROCEEDINGS, 53, 1161 (1965).

²¹ S611, D., E. Ohtsuka, D. S. Jones, R. Lohrmann, H. Hayatsu, S. Nishimura, and H. G. Khorana, these PROCEEDINGS, 54, 1378 (1965).

22Thach, R. E., M. A. Cecere, T. A. Sundararajan, and P. Doty, these PROCEEDINGS, 54, 1167 (1965).

²³ Smith, M. A., M. Salas, W. M. Stanley, Jr., A. J. Wahba, and S. Ochoa, these PROCEEDINGS. 55, 141 (1966).