ALTERED AMINO ACID SEQUENCES PRODUCED BY REVERSION OF FRAMESHIFT MUTANTS OF TRYPTOPHAN SYNTHETASE A GENE OF E. COLI*

BY W. J. BRAMMAR,[†] H. BERGER,[†] AND C. YANOFSKY

DEPARTMENT OF BIOLOGICAL SCIENCES, STANFORD UNIVERSITY, STANFORD, CALIFORNIA

Communicated August 21, 1967

Genetic analysis of acridine-induced mutants of phage T4 led to the conclusion that the genetic message is arranged in the form of nucleotide triplets, and is translated sequentially in a given phase from a fixed starting point.' This conclusion has recently been directly confirmed by Streisinger and colleagues in studies with the lysozyme of phage $T_4^{2,3}$. These authors showed that the primary structures of the lysozymes of wild-type phage and of pseudo-wild phage carrying two proflavine-induced mutational changes differed by a sequence of five contiguous amino acids. With current codon assignments, 4.5 a unique sequence of nucleotides was found to relate the two amino acid sequences.

This report concerns frameshift mutations in the tryptophan synthetase A gene of Escherichia coli. Examination of the altered amino acid sequences in the A proteins of revertants of a frameshift mutant revealed that several amino acids, in sequence, were replaced. The altered sequences are entirely consistent with current codon assignments. The findings add to the number of codons known to be utilized in vivo by E. coli.

Materials and Methods.—(a) Bacterial strains: Tryptophan synthetase A gene mutant A9813 was isolated by penicillin selection after ultraviolet irradiation of K12 strain W3110. The spontaneous and mutagen-induced revertants examined are described in Table 1.

(b) Selection of revertants: One of the three reactions catalyzed by the tryptophan synthetase A protein-B protein complex in E. coli is the production of tryptophan from indole and serine.7 This reaction is catalyzed at the wild-type rate by complexes containing mutant A proteins, provided that the B protein component is normal. The B protein itself also catalyzes the reaction, although at ^a rate 30-fold lower than that due to the AB complex. However, since the enzymes of the tryptophan operon can be derepressed up to 50 times the basal level,⁸ B protein activity of all A gene mutants is sufficient to support growth on indole-supplemented plates. When derepression is prevented by the presence of 5-methyl-DL-tryptophan, the activity of the uncomplexed B protein is insufficient for growth on indole-supplemented minimal agar. Thus A gene mutants that do not produce detectable A protein, such as nonsense mutants or frameshift mutants, fail to grow under these conditions, while A gene mutants which produce an A protein capable of complexing with the B protein, and thereby stimulating its catalytic activity, do grow. Minimal-agar plates containing 2 μ g/ml indole and 100 μ g/ml 5-methyl-DL-tryptophan were used to select for reversion of A gene nonsense mutants and frameshift mutants to strains producing A protein.

(c) Muttagenesis studies: Overnight L-broth cultures of the test strain were harvested, washed with saline, and resuspended in saline at the original cell density. One tenth ml was plated on minimal agar supplemented with 0.1 μ g/ml L-tryptophan and 0.05 ml of a solution of the mutagen was applied to the plate as a single drop. NG⁹ was used as a saturated solution, and ICR-191A⁹ at both 1 mg/ml and 100 μ g/ml. The plates were incubated for 2 days at 37° before scoring. A cluster of colonies around the site of application of the mutagen indicates a positive response.

(d) Protein procedures: Procedures for isolating the A proteins from revertants were the same as those described for the wild-type A protein.'0 Methods for oxidation and digestion of the protein have been described.¹¹ Tryptic peptides were isolated following chromatography on Dowex $1-X2.12$ Peptide patterns were preparitively 2-dimensional chromatography and electrophoresis.¹³ In some instances the ascending chromatography procedure and solvent of Weigert and Garen¹⁴ were used. Peptide compositions are expressed in terms of molar ratios of component amino acids based on an average value for one residue determined from the total composition. The subtractive Edman technique, as described by Konigsberg and Hill¹⁵ (Procedure 2), was used. Digestion with chymotrypsin, papain, carboxypeptidase A, and leucine aminopeptidase was carried out as described elsewhere.¹¹

(e) Materials: Trypsin, chymotrypsin, leucine aminopeptidase, and diisopropyl fluorophosphoryl carboxypeptidase A were purchased from Worthington Biochemical Corporation, Freehold, New Jersey. Crystalline mercuripapain was a gift from Dr. A. Light. 2,4,6 collidine, pyridine, and N-ethyl morpholine were redistilled before use. We are grateful to Dr. H. J. Creech for the generous gift of ICR-191A.

Results.—Many UV-induced A gene CRM mutants are not suppressible by any of the known amber or ochre suppressors, and are not reverted by NG. They do revert spontaneously, however, and respond in varying degrees to the acridine half-mustard ICR-191A. The particular mutant employed in this study, A9813, responds weakly but positively to ICR-191A. Mutant 9813 was plated on minimal agar containing 2 γ /ml indole plus 100 γ /ml 5-methyl-DL-tryptophan (see *Materials* and Methods), and several spontaneous revertant colonies that appeared were picked, purified, and characterized (see Table 1). Two spontaneous revertants,

* Characterization as described by Allen and Yanofsky.6 t Expressed as per cent of the colony size of wild-type colonies. In = indole, 5MT = 5-methyl-DoL-tryptophan, Min = minimal medium.

9813 PR8 and 9813 PR11, and one ICR-induced revertant, 9813 ICR-13, were chosen for further study after ascertaining that their A proteins were sufficiently stable to isolate and purify.

Determination of the primary structure changes in the A protein produced by A9813 PR8: Tryptic "fingerprints" of the A protein of ⁹⁸¹³ PR8 showed three differences from that of wild-type A protein. The wild-type peptides TP8 and TP4 were apparently missing from the peptide map of the mutant protein, and a new peptide had appeared near the origin. An amino acid analysis of this peptide, hereafter called TP84, is compared with the known composition of TP8 plus TP4 from the wild-type protein in Table 2. A total of five amino acid residues of the wild-type

AMINO ACID COMPOSITION OF TRYPTIC PEPTIDE TP84 FROM ⁹⁸¹³ PR8

This value for tyrosine is probably for because of insufficient dialysis of the protein prior to perform to add oridation.

Deduced sequence, 9813 ICR-13 TP8

Sequence, wild-type TP8 + TP4:

* Low molar ratio due to some destruction by reaction with ninhydrin.

t Molar recovery of liberated amino acids. \$ After Edman cycle III the residual peptide was electrophoresed on paper at pH 6.5 and stained with nin-

peptides was replaced by five new residues in the mutant peptide. The five altered residues, the last five of TP8, can be seen from Table 2 to be adjacent in the wildtype protein. The replacement of the arginine at the carboxyl terminus of TP8 would account for the absence of TP8 and TP4 from the peptide map, since there would no longer be a trypsin-sensitive bond between the two peptides.

The sequence of TP8-4 was determined by analysis of papain fragments of the isolated peptide, and by Edman degradation of some of these fragments (Table 3). Additional information on the sequence of TP8-4 was derived from analysis of the A protein of strain ⁹⁸¹³ PR8 SR1, ^a spontaneous derivative of ⁹⁸¹³ PR8 selected for better growth on unsupplemented minimal agar (Table 1). The spontaneous mutational event in 9813 PR8 SRi results in the replacement of a glycine by an arginine residue within TP84. This restores a trypsin-sensitive bond withinTP84, resulting in the appearance of two new tryptic peptides on peptide maps. Analysis of the two peptides isolated from a tryptic fingerprint (Table 3, bottom) showed one to have the same composition as wild-type TP4, and the other to contain four of the five amino acid replacements of the PR8 TP84. The replacement of glycine by arginine apparently restores wild-type TP4, and yields a mutant TP8 containing eight residues. This shows that the eighth residue of PR8 TP84 must be glycine, and that the sequence to the right of this position in TP84 is probably the same as that of wild-type TP4. The isolation of TP84 papain peptides Pal and Pa2 (Table 3), which have the same composition and N-terminal residues as two papain peptides from wild-type TP4,¹⁶ supports this conclusion.

Determination of the primary structure changes in the A protein from an ICR-induced revertant, 9813 ICR-13: 9813 ICR-13 is a slow-growing prototrophic strain

TABLE SEQUENCE OF TP8-4

4

FROM 9813 ICR-13

Deduced Sequence

Gly(Tyr,Thr,Phe,Cys,Cys,His, Glu,Glu,Gly,Val, Thr,Gly,Ala, Glu,Asp)Arg (Gly, Tyr, Thr)

(Gly,Tyr,Thr,Phe) (Gly ,Tyr,Thr,Phe,Cys,Cys,His, Glu,Glu,Gly) (Phe,Cys,Cys,His, Glu,Glu,Gly,Val, Thr,Gly)

Phe(Cys,Cys,His, Glu,Glu,Gly,Val, Thr,Gly)

Phe-Cys-Cys-His(Glu, Glu,Gly ,Val,Thr,Gly) (Phe,Cys,Cys,His, Glu,Glu,Gly)

(Phe,Cys,Cys,His ,Glu)Gln-Gly

(Val, Thr,Gly) (Ala, Glu,Asp)Arg

Gly(Tyr,Thr)Phe-Cys-Cys-His-Glu-Gln-Gly(Val, Thr,Gly)(Ala,Glu,Asp)Arg Gly-Tyr-Thr-Tyr-Leu-Leu-Ser- Arg-Ala-Gly-Val-Thr-Gly-Ala-Glu-Asn-Arg

hydrin prior to analysis. It showed a net negative charge at this pH, establishing Glu rather than Gln at posi-
tion five in Pa4. TP8-4 from 9813 ICR-13 A protein was isolated and digested with papain exactly as described

obtained from A9813 by mutagenesis with ICR-191A and selection on minimalglucose agar (see Table 1). The tryptic peptide map of the A protein of ⁹⁸¹³ ICR-¹³ was very similar to that of the PR8 A protein. However, analysis of peptide TP84 from ICR-13, shown in Table 4, revealed additional amino acid replacements. Six consecutive amino acids, at positions 4 through 9 in the peptide, were replaced by six other residues. The relevant sequence data are summarized in Table 4. Several of the papain peptides isolated (Pal, Pa2, and Pa3) have the same composition as peptides isolated from 9813 PR8 (see Tables 3 and 4). The other peptides (Pa4, Pa5, and Pa6) all differ from the corresponding PR8 peptides by re-

TABLE ⁵

SEQUENCE OF TP8 FROM 9813 PR11

Sequence, wild-type TP8: Gly-Tyr-Thr-Tyr-Leu-Leu-Ser-Arg

TP8 was isolated after tryptic digestion and chromatography on Dowex 1-X2. Chymotryptic peptides were isolated following chymotryptic digestion of TP8 in 0.05 M ammonium carbonate, pH 8.3, electro-
phoresis at pH 3.7 an

than Asp.
† N-terminal amino acids give low analytical values due to destruction by reaction with the ninhydrin
stain.

placement of a Gly-Ala-Gly sequence by (Glu, Glu, Gly). The sequence of these residues was determined by digestion of Pa5 with carboxypeptidase A, and by Edman degradation of peptide Pa4.

Determination of the primary structure changes in the Λ protein from A9813 PR11: The tryptic peptide map of the A protein from 9813 PR11 does not show any obvious differences from that of the wild-type A protein. However, peptide TP8 from the partial revertant has two amino acid differences from the wild-type peptide; Thr and Tyr are replaced by Asp (or Asn) and Leu (Table 5). The two residues replaced are at adjacent positions, 3 and 4, in the amino acid sequence of TP8 (see Table 5). The evidence for the amino acid sequence of TP8 from 9813 PRI1 is shown in Table 5.

Discussion.—The data presented here show that the tryptophan synthetase A proteins of spontaneous revertants of a UV-induced CRM $^-$ A gene mutant have amino acid sequences which differ from that of wild-type A protein at several consecutive positions. The relevant sequence changes are:

Using the current triplets derived largely from the work of Nirenberg and Khorana, 4 , 5 a sequence of bases can be written for the wild-type sequence which, with the addition of a single base at one site and the deletion of a single base at another, would code for the new sequence detected in the ⁹⁸¹³ PR8 A protein:

will type --- Thr-Tyr- Leu- Leu- Ser- Arg- Ala---
\n
$$
\begin{array}{c|c}\n & U & U & U \\
\hline\n & 0 & U & U \\
 & 0 & 0 & 0 \\
 & 0 & 0 & 0 \\
 & 0 & 0 & 0 \\
 & - A & 0 & 0 \\
 & 0 & 0 & 0 \\
 & 0 & 0 & 0 \\
 & 0 & 0 & 0 \\
 & 0 & 0 & 0\n\end{array}
$$
\n9813 PR8 --- ACC UUC UGC UGU CAC GGC GCC ---
\n
$$
\begin{array}{c|c}\n & U & U \\
 & 0 & 0 \\
 & 0 & 0 \\
 & 0 & 0 \\
 & 0 & 0\n\end{array}
$$
\n--- Thr-Phe- Cys- Cys- His- Gly- Ala ---

The first position of the second leucine codon shown can be assigned as C rather than U (which would also be consistent with current codon information) because this residue has been shown to be replaced by arginine in mutant A487.17

In a similar manner, codons can be written relating the PR11 TP8 sequence with that of wild-type TP8.

> wild type $-$ -Tyr- Thr- Tyr- Leu --- \sim UA_C ACC UAU CUG $|+ A| - A$ 9813 PR11 -- UA^U_C AAC CUU U_C UG --- $-$ Tyr- Asn- Leu- Leu - $-$ - -

The deletion event cannot be precisely defined from the information obtained with $PR11$, since the loss of the A or any one of the next three bases would be consistent with the observed amino acid sequence. However, since the loss of the A in question is necessary from the PR8 data, and both PR8 and PR11 were obtained from the same mutant, it seems likely that this loss of A from the tyrosine codon must be the mutational event giving rise to mutant 9813. Confirmation for this interpretation is obtained from the observation that A9813 fails to recombine with A446, a missense mutant in which this same tyrosine codon is replaced by a cysteine codon.¹⁸ Note that this change, UAU (Tyr) \rightarrow UGU (Cys) in the messenger RNA, involves the same nucleotide as that proposed for the 9813 mutational event.

The third revertant studied, 9813 ICR-13, appears to contain a frameshift extending further towards the carboxyl-terminus than that of 9813 PR8.

The amino acid replacements in revertant ICR-13 are compatible with the following scheme:

e the mutational event giving rise to mutant 9813. Conformation
station is obtained from the observation that A9813 fails to recomb
missense mutant in which this same tyrosine codon is replaced by a
⁸ Note that this change, UAU (Tyr)
$$
\rightarrow
$$
 UGU (Cys) in the message
the same nucleotide as that proposed for the 9813 mutational event
third revertant studied, 9813 ICR-13, appears to contain a frames
further towards the carboxyl-terminus than that of 9813 PR8.
amino acid replacements in revertant ICR-13 are compatible with
scheme:
wild type -- Thr. Tyr- Leu- Leu- Ser- Arg- Ala-Gly- Val ----
--- ACC UAU $\underset{A}{\text{CUG GUG CUG UGA GGG GUC}}$
 $\begin{array}{\begin{array}{c}\n \text{U} \\
 \text{A} \\
 \text{B} \\
 \text{C} \\
 \text{D} \\
 \text{A}\n \end{array}\n \end{array}\n \times \begin{array}{c}\n \text{U} \\
 \text{U} \\
 \text{A} \\
 \text{B} \\
 \text{C} \\
 \text{D} \\
 \text{D} \\
 \text{D} \\
 \text{E} \\
 \text{D} \\
 \text{E} \\
 \text{D} \\
 \text{E} \\
 \text{E} \\
 \text{D} \\
 \text{E} \\
 \text$

According to this scheme, revertant ICR-13 has the 9813 mutational event in common with the other two revertants. The rephasing event, which is the ICRinduced event, can be any one of several possible single nucleotide additions beyond the C of the wild-type alanine codon.

Since the analysis of ICR-13 identifies the wild-type arginine codon as CGA, the spontaneous reversion event in PR8 can now only be the addition of ^a G to give the GGA glycine codon. Thus the reversion events in PR8 and PR11 both involve the addition of ^a base identical to an adjacent base (i.e., A next to A and G next to G). The same conclusion was drawn from analysis of several addition events in the T4-lysozyme gene."9

This study allows the conclusion that the following codons are utilized in vivo by E. coli:

wild type Thr, ACC; Tyr, UAU; Leu, CUG; Ser, UCA; Arg, CGA
mutant Asn, AAC; Leu, CUU; Cys, UGU, UGC; His, CAC; Glu, GAG; Gly, GGA; Arg, $CGG'(A_{487})$

This small sample demonstrates that ambiguity exists for Leu, Cys, and Arg codons. It is also noteworthy that four of the five wild-type codons have a different third position, which is consistent with the expectation that third positions might be randomized in E. coli, an organism of 50 per cent GC content. This is in marked contrast to the situation with the lysozyme of T4, an AT-rich organism, in which all of five third positions were either A or U.¹⁹

Because acridine dyes are not normally mutagenic in bacteria,²⁰ frameshift mutations have been much less studied in bacteria than in phage. Ames and

Whitfield²¹ have recently presented evidence suggesting the existence of frameshift mutations in the histidine operon of Salmonella typhimurium, and have suggested ways of recognizing such mutants in bacteria. Among these was the lack of response to NG as ^a mutagen, and ^a positive response to acridine half-mustards such as ICR-191A. The present study directly establishes that one such agent, ICR-191A, is a frameshift mutagen, since it reverts A9813 by a nucleotide addition in a nearby codon. In confirmation of this, an A gene mutant induced with ICR-191A has been shown to have ^a frameshift mutation by analysis of the A proteins of its revertants.22

On the basis of other studies we have concluded that the lack of response of spontaneously revertible mutants to NG may be ^a better criterion for frameshift lesions than ICR-induced reversion, since ICR-191A reverts missense mutants by both transition and transversion events.²²

The authors are indebted to Betsy Gillim and Susan Stasiowski for their excellent technical assistance.

* This investigation was supported by grants from the National Science Foundation and the U.S. Public Health Service. W. J. B. was the recipient of a Wellcome Research Travel Grant.

^t Present address: Department of Molecular Biology, Edinburgh University, Scotland. ^t Present address: Department of Biology, The Johns Hopkins University, Baltimore, Mary-

land.

¹ Crick, F. H. C., L. Barnett, S. Brenner, and R. J. Watts-Tobin, Nature, 192, 1227 (1961).

² Terzaghi, E., Y. Okada, G. Streisinger, J. Emrich, M. Inouye, and A. Tsugita, these PRO-CEEDINGS, 56, 500 (1966).

³ Okada, Y., E. Terzaghi, G. Streisinger, J. Emrich, M. Inouye, and A. Tsugita, these PRO-CEEDINGS, 56, 1692 (1966).

⁴ Nirenberg, M., P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, and C. O'Neal, these PROCEEDINGS, 53, 1161 (1965).

⁵ Soll, D., E. Ohtsuka, D. S. Jones, R. Lohrmann, H. Hayatsu, S. Nishimura, and H. G. Khorana, these PROCEEDINGS, 54, 1378 (1965).

⁶ Allen, M. K., and C. Yanofsky, Genetics, 48, 1065 (1963).

⁷ Smith, 0. H., and C. Yanofsky, in Methods in Enzymology, ed. S. P. Colowick and N. 0. Kaplan (New York: Academic Press, 1963), vol. 5, p. 794.

⁸ Ito, J., and I. P. Crawford, Genetics, 52, 1303 (1965).

⁹ Abbreviations used: NG, N-methyl-N'-nitro-N-nitrosoguanidine; ICR-191A, 3-chloro-7 methoxy-9-(3-[chloroethyl] aminopropylamino) acridine dihydrochloride; CRM, cross-reacting material.

¹⁰ Henning, U., and C. Yanofsky, these PROCEEDINGS, 48, 1497 (1962).

l1 Guest, J. R., and C. Yanofsky, J. Biol. Chem., 240, 679 (1965).

 12 Ibid., 241, 1 (1966).

¹³ Helinski, D. R., and C. Yanofsky, Biochim. Biophys. Acta, 63, 10 (1962).

¹⁴ Weigert, M. G., and A. Garen, J. Mol. Biot., 12, 448 (1965).

¹⁵ Konigsberg, W., and R. J. Hill, J. Biol. Chem., 237, 2547 (1962).

¹⁶ Ito, J., and C. Yanofsky, unpublished observation.

¹⁷ Carlton, B. C., and C. Yanofsky, J. Biol. Chem., 240, 690 (1965).

¹⁸ Helinski, D. R., and C. Yanofsky, J. Biol. Chem., 238, 1043 (1963).

¹⁹ Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye, in Cold Spring Harbor Symposia on Quantitative Biology, vol. 31 (1966), p. 77.

²⁰ Orgel, L. E., in Advances in Enzymology, ed. F. F. Nord (New York: Interscience Publishers, 965), vol. 27, p. 289.

²¹ Ames, B. N., and H. J. Whitfield, Jr., in Cold Spring Harbor Symposia on Quantitative Biology, vol. 31 (1966), p. 221.

²² Berger, H., W. J. Brammar, and C. Yanofsky, in preparation.