*THE COMPLETE AMINO ACID SEQUENCE OF THE TRYPTOPHAN SYNTHETASE A PROTEIN (α SUBUNIT) AND ITS COLINEAR RELATIONSHIP WITH THE GENETIC MAP OF THE A GENE**

BY CHARLES YANOFSKY, GABRIEL R. DRAPEAU,† JOHN R. GUEST,† AND BRUCE C. CARLTON†

DEPARTMENT OF BIOLOGICAL SCIENCES, STANFORD UNIVERSITY

Communicated December 19, 1966

Previously we presented findings demonstrating the existence of a colinear relationship between gene structure (the genetic map) and protein structure.[1, 2] The altered tryptophan synthetase A proteins produced by a group of mutants of *Escherichia coli* were examined for primary structure changes and each mutant protein was found to differ from the wild-type protein by a change of a single amino acid.[3–6] The colinear relationship was then established by showing that the order of the positions at which these single amino acid changes occurred in the A protein was the same as the order of the respective mutational sites on the genetic map.[1, 2] It was also observed in these investigations that distance on the genetic map was reasonably representative of distance in the polypeptide chain.[1, 2] Colinearity of gene structure and protein structure has also been convincingly demonstrated as a result of studies with very different experimental material, viz., nonsense mutants[7, 8] and frame-shift mutants.[9]

Recently the complete sequence of the 267 amino acid residues in the tryptophan synthetase A protein has been determined,[10] and consequently the relationship between the genetic map of the A gene and the changes in mutationally altered A proteins can be reconsidered in terms of the primary structure of the entire protein. The purpose of this report is to re-examine this relationship.

Results and Discussion.—The amino acid sequence of the A protein, shown in Figure 1, was determined by analysis of fragments derived by treating the protein with various proteolytic enzymes or with cyanogen bromide. The details of the sequence studies will be described elsewhere.[10]

The genetic map of the relevant mutationally altered sites in the A gene is presented in Figure 2. Mutants A38 and A96 do not produce detectable altered A proteins; the map locations of their alterations are included because these sites presently represent the most distant sites in the A gene. The A38 site maps closest to the B gene and to the operator end of the tryptophan operon.[11, 12] The other altered sites on the map are the genetic locations of mutational alterations which lead to the single amino acid substitutions in the A protein that are indicated. Of these altered sites, the positions of the A3 and A33 changes are closest to the A38 site. As can be seen in Figure 2, the A3 and A33 mutational alterations lead to amino acid changes at position 48 in the protein, the closest position to the amino-terminal end of the protein at which amino acid changes are observed. The other mutationally altered sites shown correspond to amino acid changes at positions in the protein which are in the same relative order as the respective altered sites on the genetic map, as reported previously.[1, 2] The A169 mutational alteration is closest to the A96 site on the genetic map and the affected position in the protein is

Met-Gln-Arg-Tyr-Glu-Ser-Leu-Phe-Ala-Gln-Leu-Lys-Glu-Arg-Lys-Glu-Gly-Ala-Phe-Val-
10 20

Pro-Phe-Val-Thr-Leu-Gly-Asp-Pro-Gly-Ile-Glu-Gln-Ser-Leu-Lys-Ile-Asp-Thr-Leu-Ile-
30 40

Glu-Ala-Gly-Ala-Asp-Ala-Leu-Glu-Leu-Gly-Ile-Pro-Phe-Ser-Asp-Pro-Leu-Ala-Asp-Gly-
50 60

Pro-Thr-Ile-Gln-Asn-Ala-Thr-Leu-Arg-Ala-Phe-Ala-Ala-Gly-Val-Thr-Pro-Ala-Gln-Cys-
70 80

Phe-Glu-Met-Leu-Ala-Leu-Ile-Arg-Gln-Lys-His-Pro-Thr-Ile-Pro-Ile-Gly-Leu-Leu-Met-
90 100

Tyr-Ala-Asn-Leu-Val-Phe-Asn-Lys-Gly-Ile-Asp-Glu-Phe-Tyr-Ala-Gln-Cys-Glu-Lys-Val-
110 120

Gly-Val-Asp-Ser-Val-Leu-Val-Ala-Asp-Val-Pro-Val-Gln-Glu-Ser-Ala-Pro-Phe-Arg-Gln-
130 140

Ala-Ala-Leu-Arg-His-Asn-Val-Ala-Pro-Ile-Phe-Ile-Cys-Pro-Pro-Asn-Ala-Asp-Asp-Asp-
150 160

Leu-Leu-Arg-Gln-Ile-Ala-Ser-Tyr-Gly-Arg-Gly-Tyr-Thr-Tyr-Leu-Leu-Ser-Arg-Ala-Gly-
170 180

Val-Thr-Gly-Ala-Glu-Asn-Arg-Ala-Ala-Leu-Pro-Leu-Asn-His-Leu-Val-Ala-Lys-Leu-Lys-
190 200

Glu-Tyr-Asn-Ala-Ala-Pro-Pro-Leu-Gln-Gly-Phe-Gly-Ile-Ser-Ala-Pro-Asp-Gln-Val-Lys-
210 220

Ala-Ala-Ile-Asp-Ala-Gly-Ala-Ala-Gly-Ala-Ile-Ser-Gly-Ser-Ala-Ile-Val-Lys-Ile-Ile-
230 240

Glu-Gln-His-Asn-Ile-Glu-Pro-Glu-Lys-Met-Leu-Ala-Ala-Leu-Lys-Val-Phe-Val-Gln-Pro-
250 260

Met-Lys-Ala-Ala-Thr-Arg-Ser

FIG. 1.—Amino acid sequence of the tryptophan synthetase A protein of *E. coli*.[10] The underlined residues are at the positions in the protein at which amino acid changes have occurred in mutants.

only 33 residues from the carboxy-terminus of the protein. Thus, for almost the entire length of the map of the A gene, the existing evidence indicates that it is colinear with the structure of the A protein. The established orientation of the A protein relative to the A gene and consequently to the operator region of the operon, in conjunction with the orientation of the operon on the *E. coli* chromosome,[14] permits the orientation of the amino acid sequence of the A protein relative to the chromosome. The nucleotide sequence corresponding to the A protein runs in a clockwise direction from the region specifying the COOH-terminal end; i.e., the order is *thr-gal*—A gene region specifying the COOH-terminal end of the A protein—A gene region specifying the amino-terminal end of the A protein—*his*.

A representative value relating distance on the genetic map to distance in the polypeptide chain can be calculated by dividing the map distance separating the A3 and A169 sites by the number of amino acid residues in between the positions of the

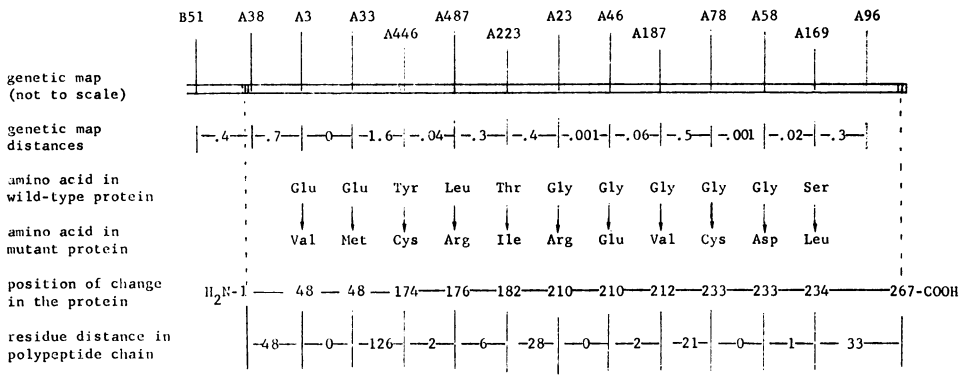

FIG. 2.—Genetic map of the A gene and the corresponding amino acid changes in the A protein. The positions of these changes in the amino acid sequence are also indicated. Mutants A3 and A33, and their amino acid replacements, will be described in detail elsewhere.[13]

corresponding amino acid changes. The value so obtained is about 0.015 map units per amino acid residue. Using this value we can estimate that the genetic map of the A gene should extend some 0.7 units to the left of the A3 site and about 0.5 units beyond the A169 site. This would give a total length of approximately 4.2 map units for the A gene, and would place the A38 site at or very near the beginning of the A gene.

The authors are pleased to acknowledge the assistance of Virginia Horn, Donald Vinicor, Deanna Thorpe, and George Pegelow at various stages of this investigation.

* These investigations were supported by grants from the U.S. Public Health Service and the National Science Foundation.

† Present addresses: G. R. D., Department of Microbiology, University of Montreal, Montreal, Quebec, Canada; J. R. G., Department of Microbiology, Sheffield University, Sheffield, England; B. C. C., Department of Biology, Yale University, New Haven, Connecticut.

[1] Yanofsky, C., in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 28 (1963), p. 296.

[2] Yanofsky, C., B. C. Carlton, J. R. Guest, D. R. Helinski, and U. Henning, these PROCEEDINGS, **51**, 266 (1964).

[3] Helinski, D. R., and C. Yanofsky, these PROCEEDINGS, **48**, 173 (1962).

[4] Henning, U., and C. Yanofsky, these PROCEEDINGS, **48**, 183 (1962).

[5] Guest, J. R., and C. Yanofsky, *J. Biol. Chem.*, **240**, 679 (1965).

[6] Carlton, B. C., and C. Yanofsky, *J. Biol. Chem.*, **240**, 690 (1965).

[7] Sarabhai, A. S., A. O. W. Stretton, S. Brenner and A. Bolle, *Nature*, **201**, 13 (1964).

[8] Fowler, A. V., and I. Zabin, *Science*, **154**, 1027 (1966).

[9] 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, in press.

[10] Guest, J. R., B. C. Carlton, and C. Yanofsky, in preparation; Carlton, B. C., J. R. Guest, and C. Yanofsky, in preparation; Drapeau, G., and C. Yanofsky, in preparation; and Yanofsky, C., J. R. Guest, G. Drapeau, and B. C. Carlton, in preparation.

[11] Guest, J. R., and C. Yanofsky, *Nature*, **210**, 799 (1966).

[12] Matsushiro, A., K. Sato, J. Ito, S. Kida, and F. Imamoto, *J. Mol. Biol.*, **11**, 54 (1965).

[13] Drapeau, G., W. Brammar, and C. Yanofsky, in preparation.

[14] Signer, E. R., J. R. Beckwith, and S. Brenner, *J. Mol. Biol.*, **14**, 153 (1965).