The Amino Acid Sequence of *Escherichia coli* Cyanase*

(Received for publication, July 12, 1982)

Christopher C. Q. Chin‡§, Paul M. Anderson¶, and Finn Wold‡§

From the Departments of Biochemistry, University of Minnesota, St. Paul, Minnesota 55108 and Duluth, Minnesota 55812

The amino acid sequence of the enzyme cyanase (cyanate hydrolase) from *Escherichia coli* has been determined by automatic Edman degradation of the intact protein and of its component peptides. The primary peptides used in the sequencing were produced by cyanogen bromide cleavage at the methionine residues, yielding 4 peptides plus free homoserine from the NH2-terminal sequences. The complete sequence of the cyanase subunit consists of 156 amino acid residues (Mr = 16,350). Based on the observation that the cysteine-containing peptide is obtained as a disulfide-linked dimer, it is proposed that the covalent structure of cyanase is made up of two subunits linked by a disulfide bond between the single cysteine residue in each subunit. The native enzyme (Mr = 150,000) then appears to be a complex of four or five such subunit dimers.

The enzyme cyanase (cyanate hydrolase EC 3.5.5.3) has been reported to be present in animal tissues (1), plants (2, 3) and bacteria (4, 5). In a recent investigation of this enzyme it was not possible to confirm its presence in animal tissues (6), but as reported (5) the enzyme was found to be present in *Escherichia coli*, and after induction with cyanate it could be isolated and characterized (7). The pure *E. coli* enzyme is a multimer (Mr = 150,000) made up of subunits of about 15,000 daltons; it catalyzes the hydrolysis of cyanate to ammonia and bicarbonate in a rather complex reaction that requires bicarbonate and in which the relatively slow release of ammonia strongly suggests that an intermediate such as carbamate is the initial product of the reaction. In order to explore the molecular and catalytic properties of this enzyme in greater depth, it was felt important to determine its covalent structure. In this paper we report the amino acid sequence of the apparently identical Mr = 16,354 subunits that make up the native enzyme.

EXPERIMENTAL PROCEDURES

RESULTS AND DISCUSSION

Summary and Evaluation of the Sequence Data—The complete structure of *E. coli* cyanase is given in Fig. 1. An examination of the fragments used in deducing the sequence should establish that the reported sequence is documented in terms of peptide overlaps and that it is unlikely that segments may have been missed. The weakest point in this regard is the sequence Ala-58-Arg-59-Leu-60 (the COOH terminus of a chymotryptic peptide, C3) in which the single Leu residue establishes the connection between the two Arg peptides R1 (12-59) and R2 (60-81). The sequence of CNBr peptide M5 (95-100) was determined by ambiguous results because of the combination of Glu and homoserine at the COOH-terminal end; these two amino acids were not resolved by amino acid analysis and the carboxypeptidase treatment consequently gave equivocal results. The ideal way to solve this problem would be to sequence peptide R2 (97-141), but we were unable to achieve a satisfactory separation of the two large Arg peptides R1 and R2. Since R1 had been determined from the large CNBr peptide M3, however, and its sequence had been established, R1 and R2 were sequenced as a mixture, and the initial 8 residues of R2 were thus deduced to establish the important overlap of M3, R1, and M5. Another part of the sequence, the carboxyl terminus, is based on minimal direct sequence data but is supported by carboxypeptidase A release of a single Phe from both the intact enzyme and the COOH-terminal peptides M4 and R6, the amino acid analyses of the two COOH-terminal peptides exclude the presence of an additional short peptide terminating in a second Phe residue. Some useful peptides were obtained by serendipity from unexpected enzyme cleavages. Thus, our preparation of clostripain yielded several unexpected peptides, including one from M3 which could only result from a cleavage between Ala-46 and Leu-47 and which along with the chymotryptic peptide 49-59 established the COOH-terminal segment of R3.

As in previous work (8, 9) we used strong acid for many of the peptide fractionations, and the identification of Asn and Gln could consequently be ambiguous if the exposure to acid was sufficient to hydrolyze the amides. Another similar problem is related to our finding low Tyr yields in CNBr peptides both on amino acid analysis and sequencing. Because of these concerns, we collected a good deal of redundant sequence data as a means to eliminate these possible ambiguities. As a rule with the acid-treated peptides we assigned any residue which had a readily identifiable amide component (10-20 of the total...
Amino Acid Sequence of E. coli Cyanase

**FIG. 1.** The amino acid sequence of *E. coli* cyanase. The lines indicate the extent to which sequence information was obtained from each fragment: dotted line, intact enzyme; broken lines, peptides produced by cyanogen bromide cleavage, designated Ms, M\(_d\), and M\(_S\) (M\(_I\) is free homoserine and sequence data for M\(_Z\) are not included in this report); solid lines, peptides produced by tryptic cleavage of acetylated protein or peptides, designated R\(_S\), R\(_4\), R\(_S\), R\(_e\), R\(_7\), and R\(_o\); alternating dots and lines, 2 peptides produced by chymotryptic cleavage of peptides M\(_P\) and M\(_S\), respectively, designated C\(_i\) and C\(_Z\); and one of several (unexpected) peptides produced by clostripain cleavage of M\(_I\), designated CL.

Asx or Glx) to Asn or Gln, and in many cases these could be confirmed by high yield of amides in peptides that had been treated more gently.

It is our experience that one of the greatest sources of variation in automatic sequencing with the model 890C sequencer without a condenser (cold trap) for the vacuum system is the rapidly declining quality of the pumps. Even with frequent exchange with rebuilt pumps, both the degree of evacuation and the rate with which it is achieved vary over relatively short periods of time, and the repetitive yields were consequently also subject to substantial variation (87-97%). Frequent monitoring of the sequencer performance was carried out, and most of the reported data were obtained with repetitive yields from 93 to 97%.

**Implications of the Sequence on Cyanase Structure**—It may be interesting to note that in assessing the cyanase sequence according to Levitt’s (10) statistical analysis of conformational preferences of amino acids in globular proteins, no sequence of more than 4 amino acids favoring (or neutral to) either α-helix, sheet, or reverse-turn structures was found. One aspect of the data warrants detailed consideration here, namely the role of the single Cys residue in cyanase. In the original report on *E. coli* cyanase (7) the absence of any free sulfhydryl groups was noted. This feature was established by titration with 5,5′-dithiobis-(2-nitrobenzoate) before and after treatment with sodium dodecyl sulfate or in the presence of 6 M guanidine hydrochloride or 8 M urea. Since each subunit contains only a single Cys residue, the most plausible explanation for the absence of free sulfhydryl groups is that pairs of subunits are covalently linked through disulfide bonds, and that the covalent structure of cyanase thus is a dimer. Consistent with this picture is the finding (Fig. 2 in Miniprint) that the CNBr treatment of cyanase yielded a substantial amount (50-70%) of the peptide M\(_S\) (78-94, containing the single Cys-83 residue) as a dimer which could be converted to the monomer by performic acid oxidation. We assume that under the acidic conditions of the CNBr treatment and the subsequent peptide separation, no disulfide formation would
take place, and that the dimeric peptide thus represents a structure preformed in the intact, active enzyme. With this model the 150,000-dalton enzyme should have the general structure \((P-S-S-P)_n\), where \(P\) represents the cyanase subunit of \(M_r = 16,350\), and \(n\) has a value of either 4 or 5.

**Acknowledgments**—We wish to express our gratitude to Tom Krick for performing the mass spectrographic analyses. The sequencing facility used for most of this work was established with contributions from the Agricultural Experiment Station, the Medical School, and the Graduate School at the University of Minnesota. The final experiments were carried out at the Sequencing Laboratory at the University of Texas Medical School at Houston established by support from the Robert A. Welch Foundation. All these sources of support are gratefully acknowledged.

**REFERENCES**

1. Holtham, S. B., and Schutz, F. (1948) *Experientia* (Basel), 4, 398-399
2. Lisanti, L. E. (1963) *Ann. Sper. Agrar.* 17, 45-51
3. Lotti, G. (1963) *Chim. Ind. (Milan)* 45, 561-563
4. Taussig, A. (1960) *Biochim. Biophys. Acta* 44, 510-519
5. Taussig, A. (1965) *Can. J. Biochem.* 43, 1063-1069
6. Collins, C. A., and Anderson, P. M. (1977) *Biochem. Biophys. Res. Commun.* 79, 1253-1260
7. Anderson, P. M. (1980) *Biochemistry* 19, 2882-2888
8. Chin, C. C. Q., Brewer, J. M., Eckard, E., and Wold, F. (1981) *J. Biol. Chem.* 256, 1370-1376
9. Chin, C. C. Q., Brewer, J. M., and Wold, F. (1981) *J. Biol. Chem.* 256, 1377-1384
10. Levitt, M. (1978) *Biochemistry* 17, 4277-4285

Additional references are found on p. 280.
Amino Acid Sequence of E. coli Cyanase

Christopher C. O. Chiu, Paul M. Anderson and Finn Wold

EXPERIMENTAL PROCEDURES

RESULTS AND DISCUSSION

Preparation of Peptides: The general strategy for peptide separation was to (a) isolate the active cyanase by column chromatography, (b) perform tryptic digestion of the cyanase and isolate the resulting peptides by column chromatography, (c) purify the isolated peptides by high-performance liquid chromatography, (d) determine the amino acid composition of the purified peptides, (e) perform automated Edman degradation of the purified peptides, and (f) compare the Edman degradation data of each peptide with the theoretical Edman data for a tryptically digested peptide. The results of this strategy are presented in Table I and Table II.

TABLE I. Sequencing Data for Intact Cyanase (110 mM)

| Cycle | Peptide | N-Terminal Sequence | C-Terminal Sequence |
|-------|---------|----------------------|---------------------|
| 1     | Cys     | Glu                   | Leu                 |
| 2     | Asp (Ser) | Ala                  | Leu                 |
| 3     | Asp     | Glu                   | Leu                 |
| 4     | Ala     | Glu                   | Leu                 |
| 5     | Asp (Ser) | Ala                  | Leu                 |
| 6     | Asp     | Glu                   | Leu                 |
| 7     | Asp (Ser) | Ala                  | Leu                 |
| 8     | Asp     | Glu                   | Leu                 |
| 9     | Ala     | Glu                   | Leu                 |

Peptides were isolated from the cyanase by column chromatography, and the amino acid sequence of each peptide was determined by automated Edman degradation. The results are presented in Table I. The N-terminal sequence of the cyanase was confirmed by the Edman degradation of the peptide isolated from the cyanase. The C-terminal sequence was determined by the Edman degradation of the peptide isolated from the cyanase. The results are consistent with the N-terminal and C-terminal sequences determined by other methods.
**Amino Acid Sequence of E. coli Cyanase**

#### TABLE VI. Sequencing Data for Peptide Rh (100 nmol)

| Residue Identified | Identifiable by Amino Acid Analysis after Hydrolysis |
|---------------------|-----------------------------------------------------|
| Glu | Gln |
| Asp | Asp |
| Leu | Leu |
| Glu | Glu |
| Asp | Asp |
| Ala | Ala |
| Val | Val |
| Asp | Asp |
| Ser | Ser |
| Gly | Gly |
| Val | Val |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Glu | Glu |
| Asp | Asp |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Gly | Gly |
| Val | Val |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Glu | Glu |
| Asp | Asp |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Gly | Gly |
| Val | Val |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Glu | Glu |
| Asp | Asp |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Gly | Gly |
| Val | Val |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Glu | Glu |
| Asp | Asp |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Gly | Gly |
| Val | Val |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Glu | Glu |
| Asp | Asp |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Gly | Gly |
| Val | Val |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Glu | Glu |
| Asp | Asp |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Gly | Gly |
| Val | Val |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Glu | Glu |
| Asp | Asp |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Gly | Gly |
| Val | Val |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Glu | Glu |
| Asp | Asp |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
| Gly | Gly |
| Val | Val |
| Ala | Ala |
| Leu | Leu |
| Asp | Asp |
| Ser | Ser |
Amino Acid Sequence of E. coli Cyanase

Fig. 4. Purification of peptide 4. A peptide (purified by salting-out fractionation of peak A in Fig. 1) was digested with trypsin and digested A 20% TFA in 99% ethanol. The digestion mixture was applied to a 6 X 25 cm column (50 mM HCOOH, 50% ethanol). A separation of 3 peaks was observed when the column was eluted with a linear gradient of ethanol. Amino acid analysis of peak 2 indicated it was a single peptide corresponding to Fig. 2.

Fig. 5a. Purification of peptide 5. Peptide 4 was digested with trypsin and the digestion mixture was digested and attempted lyophilized. The lyophilized material was redissolved in 99% ethanol and a second separation was observed. Amino acid analysis indicated that fractions 31-38 (labeled A) contained arginine. These fractions were pooled and lyophilized.

Fig. 5b. Fraction A from Figure 5a was subjected to gel filtration on Sephadex G-25 (1.5 X 60 cm) with 1.0 M ammonium acetate. Amino acid analysis of the pooled fractions 30-38 showed a high arginine content. This fraction was lyophilized and used for further purification.

Fig. 5c. Fraction A from Figure 5b was subjected to a phenylboronate column (conditions given in Fig. 5). Fractions 40-48 (labeled A) was lyophilized, lyophilized and desalted. This fraction contained a single peptide (I).

Fig. 6a. Fraction A from Figure 6 was subjected to a phenylboronate column (conditions given in Fig. 5). Fractions 40-48 (labeled A) was lyophilized, lyophilized and desalted. This fraction contained a single peptide (I).

Fig. 6b. Fraction A from Figure 6b was subjected to a phenylboronate column (conditions given in Fig. 5). Fractions 40-48 (labeled A) was lyophilized, lyophilized and desalted. This fraction contained a single peptide (I).

Fig. 7a. Purification of peptide 2. Peptide 1 (1000 n mole) was digested with chymotrypsin and the digestion mixture was applied to a phenylboronate column (1.5 X 60 cm). The digestion mixture was applied to a phenylboronate column (1.5 X 60 cm). A separation of 3 peaks was observed when the column was eluted with a linear gradient of ethanol. Amino acid analysis showed that most of the arginine was located in pooled Fraction 2-18 (B), and this fraction was lyophilized and used for further purification.

Fig. 7b. Purification of peptide 3. Peptide 1 (1000 n mole) was digested with chymotrypsin and the digestion mixture was applied to a phenylboronate column (1.5 X 60 cm). A separation of 3 peaks was observed when the column was eluted with a linear gradient of ethanol. Amino acid analysis showed that most of the arginine was located in pooled Fraction 2-18 (B), and this fraction was lyophilized and used for further purification.

Fig. 7c. Purification of peptide 4. Peptide 3 (1000 n mole) was digested with chymotrypsin and the digestion mixture was applied to a phenylboronate column (1.5 X 60 cm). A separation of 3 peaks was observed when the column was eluted with a linear gradient of ethanol. Amino acid analysis showed that most of the arginine was located in pooled Fraction 2-18 (B), and this fraction was lyophilized and used for further purification.

Fig. 7d. Purification of peptide 5. Peptide 4 (1000 n mole) was digested with chymotrypsin and the digestion mixture was applied to a phenylboronate column (1.5 X 60 cm). A separation of 3 peaks was observed when the column was eluted with a linear gradient of ethanol. Amino acid analysis showed that most of the arginine was located in pooled Fraction 2-18 (B), and this fraction was lyophilized and used for further purification.

Fig. 8. Purification of peptide 6. Peptide 5 (1000 n mole) was digested with chymotrypsin and the digestion mixture was applied to a phenylboronate column (1.5 X 60 cm). A separation of 3 peaks was observed when the column was eluted with a linear gradient of ethanol. Amino acid analysis showed that most of the arginine was located in pooled Fraction 2-18 (B), and this fraction was lyophilized and used for further purification.
Amino Acid Sequence of E. coli Cyanase

| Cyanase (1-156) | CNBr - Peptides | Tryptic Peptides (Arg-cleavage) | Other Peptides |
|-----------------|-----------------|---------------------------------|----------------|
|                 | M2              | M3                             | M4             | M5             | R3             | R4             | R5             | R6             | R8             | C1             | C2             | CL             |
| Cysteic Acid    | 0.7(1)          | - (0)                          | 0.7(1)         | - (0)          | - (0)          | 0.9(1)         | - (0)          | - (0)          | - (0)          | - (0)          | - (0)          | - (0)          |
| Aspartate       | 17.1(18)        | 9.9(10)                        | 3.0(3)         | 0.2(0)         | 5.4(5)         | 4.6(6)         | 3.0(3)         | 2.5(2)         | 1.4(1)         | 1.2(1)         | 1.3(1)         | 2.1(2)         | 1.2(1)         |
| Threonine       | 7.6(6)          | 1.8(2)                         | 2.0(2)         | 0.2(0)         | 3.2(4)         | 1.5(2)         | 0.4(0)         | 0.2(0)         | 2.1(2)         | 1.6(2)         | 0.1(0)         | 1.0(1)         | - (0)          |
| Serine          | 5.6(6)          | 2.9(5)                         | - (0)          | 0.2(0)         | 1.0(1)         | 1.5(2)         | 1.1(2)         | 0.6(0)         | 0.2(0)         | 0.2(0)         | 0.1(0)         | 0.3(0)         | - (0)          |
| Glutamate       | 13.8(13)        | 8.7(9)                         | 1.0(1)         | 1.6(2)         | 4.8(4)         | 3.8(4)         | 2.5(2)         | 0.6(0)         | 0.1(0)         | 0.2(0)         | 1.4(2)         | 2.3(2)         | 2.5(2)         |
| Proline         | 7.9(7)          | 1.0(1)                         | 3.7(3)         | - (0)          | 2.6(2)         | 0.8(1)         | 1.0(1)         | - (0)          | 1.8(2)         | 1.8(2)         | 0.9(1)         | 1.0(1)         | 0.7(1)         |
| Glycine         | 10.8(11)        | 4.2(4)                         | 1.5(1)         | - (0)          | 6.5(6)         | 3.4(3)         | 1.8(1)         | 1.6(1)         | 0.3(0)         | 1.1(1)         | 0.9(1)         | 2.4(2)         | 1.4(1)         |
| Alanine         | 18.3(18)        | 14.0(14)                       | 0.3(0)         | - (0)          | 4.6(4)         | 10.9(13)       | 2.8(1)         | 0.2(0)         | - (0)          | 1.1(1)         | 3.1(4)         | 2.0(2)         | 3.9(4)         |
| Valine          | 6.6(7)          | 2.0(2)                         | - (0)          | - (0)          | 4.4(5)         | 1.1(1)         | 0.9(1)         | - (0)          | - (0)          | 1.0(1)         | - (0)          | 2.7(3)         | - (0)          |
| Methionine      | 4.0(4)          | - (0)                          | - (0)          | - (0)          | - (0)          | - (0)          | - (0)          | - (0)          | - (0)          | - (0)          | - (0)          | - (0)          | - (0)          |
| Isoleucine      | 11.4(13)        | 5.2(6)                         | 2.9(3)         | - (0)          | 2.6(4)         | 1.8(2)         | 2.0(2)         | 1.1(1)         | 1.6(1)         | 1.0(1)         | 0.4(1)         | 1.0(1)         | - (0)          |
| Leucine         | 21.8(21)        | 14.0(14)                       | 1.4(1)         | 0.1(0)         | 7.0(6)         | 7.3(9)         | 6.3(6)         | 0.3(0)         | - (0)          | 2.1(2)         | 1.9(2)         | 2.0(2)         | 2.2(3)         |
| Tyrosine        | 3.8(4)          | - (0)                          | - (0)          | 1.4(2)         | 1.9(2)         | - (0)          | - (0)          | - (0)          | 0.7(1)         | 0.9(1)         | - (0)          | - (0)          | - (0)          |
| Phenylalanine   | 6.2(6)          | 2.0(2)                         | 0.1(0)         | 1.0(1)         | 3.3(3)         | 1.8(2)         | - (0)          | - (0)          | - (0)          | 0.8(1)         | - (0)          | 0.5(0)         | - (0)          |
| Lysine          | 10.9(11)        | 4.0(4)                         | 0.4(0)         | - (0)          | 6.4(7)         | 2.9(3)         | 1.7(1)         | 0.2(0)         | - (0)          | 1.8(2)         | - (0)          | 2.7(3)         | - (0)          |
| Histidine       | 1.1(1)          | - (0)                          | - (0)          | - (0)          | 1.3(1)         | - (0)          | - (0)          | 0.1(0)         | - (0)          | - (0)          | - (0)          | - (0)          | - (0)          |
| Arginine        | 7.0(7)          | 3.1(3)                         | 2.2(2)         | 1.0(1)         | 1.3(1)         | 1.0(1)         | 1.0(1)         | 1.0(1)         | 0.7(0)         | 1.0(1)         | 1.0(1)         | 1.0(1)         | 1.0(1)         |

a The numbers in parenthesis represent the numbers of residues deduced from the reported sequence. The analysis for cyanase is the average of the analysis of three different preparations of the enzyme. A dash signifies less than 0.05 residues.

b Homoserine, the product of the CNBr-treatment of Met, elutes at the Glu position. The glutamate values for the CNBr peptides thus represent Glu and homoserine (Met) both in the analysis values and the theoretical number of residues.

c C1 and C2 represent two chymotryptic peptides from M2 and M5, respectively; CL represents a peptide isolated after treatment of M2 with clostripain.