Amino Acid Sequence of Escherichia coli Biotin Carboxyl Carrier Protein (9100)*

(Received for publication, January 25, 1977)

MICHAEL R. SUTTON, R. RAY FALL,‡ ANDRAI M. NERVI,* ALFRED W. ALBERTS,¶ P. ROY VAGELOS,¶ AND RALPH A. BRADSHAW

From the Department of Biological Chemistry, Washington University School of Medicine, St. Louis, Missouri 63110

The amino acid sequence of a proteolytic fragment of Escherichia coli biotin carboxyl carrier protein was determined from the structures of overlapping tryptic, thermolytic, and staphylococcal protease peptides together with automated sequenator analyses on the intact protein. The fragment, 82 residues in length, contains the single residue of biocytin of the protein. The relationship of the $M_r = 9100$ fragment to the native $M_r = 22,500$ subunit is discussed.

Biotin carboxyl carrier protein plays a central role in the acetyl-CoA carboxylase complex of Escherichia coli. The enzyme complex catalyzes the carboxylation of acetyl-CoA to malonyl-CoA, the first committed step in the synthesis of long chain fatty acids. The enzyme has been resolved into three dissimilar subunits (1, 2). In addition to BCCP, there is a biotin carboxylase and a transcarboxylase, each catalyzing one step in the overall reaction:

$$\begin{align*}
\text{HCO}_3^- + \text{ATP} &\rightarrow \text{Mn}^{2+} + \text{ADP} + \text{P}, \\
\text{biotin carboxylase} &\rightarrow \text{BCCP} \\
\text{BCCP} &\rightarrow \text{BCCP-CO}_2^- \\
\text{transcarboxylase} &\rightarrow \text{CO}_2\text{CH}_2\text{CO}_2\text{CoA} \\
\text{CH}\text{COSCoA} &\rightarrow \text{CH}\text{COSCoA}
\end{align*}$$

BCCP, which contains a biotinyl prosthetic group covalently attached to a specific lysyl residue, is carboxylated by bicarbonate in the first partial reaction. The reaction is catalyzed by the biotin carboxylase subunit and depends on ATP and Mn$^{2+}$ (2, 3). Transcarboxylase then catalyzes the transfer of the carboxyl group to an acetyl-CoA acceptor, forming malonyl-CoA and regenerating BCCP for further carboxylation (4, 5). The stoichiometry of the subunits in the native enzyme is unknown since the complex spontaneously dissociates in vitro. Several species of BCCP ranging in molecular weight from 8,900 to 45,000 have been isolated from E. coli. The native form of BCCP is a dimer of similar polypeptide chains, each with a molecular weight of 22,500 and containing one covalently bound biotin residue (6, 7). Two smaller fragments of E. coli BCCP have also been isolated, having molecular weights of 10,400 and 9,100 (8). Although conditions for the isolation of these fragments could not be reproduced, they are thought to have resulted from the action of endogenous proteolytic enzymes present during the isolation (9). Treatment of crude preparations of BCCP with subtilisin Carlsberg produced a fragment BCCP$_{9100}$, identical in all respects to BCCP$_{10400}$ except that amino acid analysis showed it to contain 2 fewer alanyl residues (9).

The BCCP fragments each contain 1 mol of biotin per mol of polypeptide and all are active to approximately the same extent in both partial reactions of acetyl-CoA carboxylase, although with $K_m$ values 50 to 100 times that of native BCCP (8, 9). BCCP$_{9100}$ and BCCP$_{9100}$ readily crystallize (8, 9) and three-dimensional structural analyses are presently under investigation.

The complete amino acid sequence of fragment BCCP$_{9100}$ has now been determined from overlapping tryptic, thermolytic, and Staphylococcus aureus V8 protease peptides, together with automatic sequenator analyses on the intact fragment.

EXPERIMENTAL PROCEDURES

Materials—BCCP$_{9100}$ and BCCP$_{10400}$, both containing $^{14}$H-biotin, were prepared as described previously (8, 9). 1,1-Tosylamido-2-phenylethyl chloromethyl ketone (TPCK)-treated trypsin and diisopropyl fluorophosphate-treated carboxypeptidase A were purchased from Worthington Biochemical Corp., thermolysin from Calbiochem, diisopropyl fluorophosphate-treated carboxypeptidase B were purchased from Worthington Biochemical Corp., thermolysin from Calbiochem, diisopropyl fluorophosphate-treated carboxypeptidase B from Sigma Chemical Co., and Staphylococcus aureus V8 protease from Miles Laboratories, Inc. Dithiothreitol, iodoacetic acid, and trifluoroacetic acid were purchased from Sigma. Iodoacetic acid was recrystallized from petroleum ether and trifluoroacetic acid was redistilled prior to use. Iodo$^{13}$C-acetic acid was purchased from New England Nuclear, $\beta$-mercaptoethanol from Eastman Kodak Co., phenylisothiocyanate from Beckman Instruments, Inc., and dimethylallylamine from Pierce Chemical Co. Fluorescamine was obtained from Roche Diagnostics and guanidine·HCl was ultrapure grade from Mann. Polymide thin layer plates were obtained from the Chen Ching Trad-
ing Ca. and silica thin layer plates from Analytichem, Inc. All other materials were the highest grade commercially available.

Reduction and Carboxymethylation—BCCP (3 to 5 μmol) was dissolved in 0.5 ml of Tri/i/Cl (0.3 M, pH 7.1) containing guanidine hydrochloride (500 mg). Dithiothreitol (10 μl of 100 mg/ml aqueous solution) was added and the pH adjusted to 6.7 with 0.2 M NaOH. Beddingfield was allowed to proceed for 10 min in the dark at room temperature. Recrystallized iodoacetic acid (40 mg) was added to ido[14C]acetic acid (3 mg, 0.22 mCi) and dissolved in 0.5 ml of 0.2 M NaOH. The iodoacetic acid solution was added to the reduced protein prior to the addition of excess β-mercaptoethanol (100 μl). The solution was dialyzed against water (three changes of 2 liters each) in Spectrapore 6000 to 8000 dialysis tubing. The reduced carboxymethylated protein was lyophilized.

Enzymatic Digests—Enzymatic digests of carboxymethylated BCCP with trypsin and thermolysin were carried out under the conditions described previously (10, 11). S. aureus V8 protease digestion was carried out on carboxymethylated BCCP C3. The protein (36 mg) was dissolved in 2 ml of NH₄HCO₃, pH 8.0, and the enzyme was added (1 mg dissolved in 0.1 ml of 0.1 M NH₄HCO₃). The reaction was allowed to proceed at 37°C for 18 h. The pH of the solution was checked at time zero and after 30 min. Precipitated material was centrifuged and the supernatant adjusted to 2.0 with 6 M HCl. The precipitated material was dissolved in 0.5 ml of Tris/Cl (0.2 M, pH 7.1) containing guanidine hydrochloride (500 mg). Dithiothreitol (10 μl of 100 mg/ml aqueous solution) was added and the pH adjusted to 6.7 with 0.2 M NaOH. Beddingfield was allowed to proceed for 10 min in the dark at room temperature, the reaction was stopped by the addition of excess β-mercaptoethanol (100 μl). The solution was dialyzed against water (three changes of 2 liters each) in Spectrapore 6000 to 8000 dialysis tubing. The reduced carboxymethylated protein was lyophilized.

RESULTS

The proposed structure for BCCP C3 is shown in Fig. 1. The structure was constructed from peptides derived from trypsin (T), thermolytic (Th), and Staphylococcus aureus protease (SP) hydrolyses of the S-¹⁴C-carboxymethylated protein. Only the S. aureus protease hydrolysis provided a complete set of peptides.

The proposed structure contains 82 and 80 amino acids for BCCP C3 and BCCP C1, respectively, and is in exact agreement with the amino acid compositions reported previously (8, 9). The calculated molecular weights from the sequence of the two proteins are 9084 (BCCP C3) and 8942 (BCCP C1). As shown in Fig. 1, the 2 additional residues of alanine distinguishing BCCP C3 from BCCP C1 are located at the NH₂ terminus.

The peptides isolated in this study were derived from both forms of the protein, primarily due to the limited amounts of BCCP C1 available. One trypsic hydrolysis and the thermolysin hydrolysis were carried out on the longer protein; the second trypsin and the S. aureus protease hydrolyses were carried out on the subtilisin derivative. As shown in the figure, sequenator analyses were also carried out on both forms of the protein. Residue numbering is based on BCCP C3.

Sequenator Analysis—Sequenator analysis of the two S-¹⁴C-carboxymethyl BCCP derivatives identified 23 and 22 residues of BCCP C3 and BCCP C1, respectively, giving a total identification of 24 residues. These analyses, which are summarized in Table I, yield the location of the single histidyl, the single tyrosyl, and both arginyl residues of the protein. They also provided assignments for Glu 3 and Arg 24.

Isolation of Enzymatic Fragments—Tryptic hydrolysates of both forms of BCCP were obtained, with markedly different results. The digestion of BCCP C3 yielded a large acid-insoluble fraction which contained both the S-¹⁴C-carboxymethyl cysteinyl and the S-¹⁴C-hydroxycetyl residues. This material was not successfully fractionated. Fractionation of the acid-soluble material on a column of Dowex 50-X8 (see miniprint supplement) produced four peptides: T1, T2, T3, and T7. As indicated in Fig. 1, peptide T1 was sequenced partially and peptide T2 completely by manual methods. They contribute the first two fragments of the protein and as such are completely contained within the region structured by the automated sequenator analyses. Peptide T3 was sequenced partially and peptide T7 completely by manual Edman degradations.

The second trypsic hydrolysis, carried out on BCCP C1, inexplicably produced nearly quantitative chymotrypsin-like cleavages in addition to the normal trypsic cleavages. This digest was carried out overnight using a different preparation of trypsin than that used in the first hydrolysis, which was of 4 h duration. As a result of the more extensive digestion, a considerably smaller acid-insoluble core was generated and a greater percentage of the molecule was solubilized. The NH₂-terminal peptide, in this case without the NH₂-terminal alanyl residues, was recovered as T1a. Partial hydrolysis at His 7 produced in addition T1b and T1c. Peptide T2 was recovered from this digest as T2a, which was foreshortened by 2 residues at the COOH terminus as a result of cleavage at Phe 17. A similar situation was encountered with peptide T3, which was isolated both as the octapeptide and as a hexapep-
Sequence of E. coli Biotin Protein

FIG. 1. Amino acid sequence of Escherichia coli biotin carboxyl carrier protein 9100. Peptides are designated as tryptic (T), thermolytic (Th), or staphylococcal protease (SP) and are numbered sequentially in the order they occur in the protein. Residues positioned by automated Edman degradation (→), by manual Edman degradation (→), or by hydrolysis with carboxypeptidases (→) are indicated. Bet (residue 48) = biocytin.

tide, T3a. Peptide T7 was recovered in the same form as that obtained in the first digest.

Four additional peptides, all arising as a result of chymotryptic-like cleavages, were obtained. These were T4, T5, T6, and T8. The first three were completely sequenced by the dansyl procedure. T5 contained the single cysteinyl residue and T6 the single biocytinyl residue of the protein. As the result of limited amounts of material, only the NH2-terminal residue of T8 was determined. Peptides corresponding to residues 20 to 26, 48 to 57, and 52 to 57 could be tentatively identified in impure fractions, but were not obtained in a homogeneous form.

The thermolytic digest of BCCP9100 produced negligible acid-insoluble material, and hence yielded an almost complete set of peptides. For the five regions (residues 35 to 36, 43 to 46, 49, 61 to 63, and 81 to 82) for which no peptides are listed (Fig. 1), impure fractions were again obtained in which the missing peptides could be tentatively identified, albeit in low yield. Amino acid and NH2-terminal analyses of Th5 showed it to be heterogeneous, with partial phenylalanine at the NH2 terminus. No attempt was made to further purify the components of this pool.

The S. aureus protease hydrolysis of S-[14C]carboxymethyl BCCP9100 produced a complete set of peptides in generally high
Table I

| Position | Residue | Yield* |
|----------|---------|--------|
| 1        | Alanine | 116    |
| 2        | Alanine | 103    |
| 3        | Glutamic acid | 58 |
| 4        | Isoleucine | 151 |
| 5        | Serine  | 72     |
| 6        | Glycine | 117    |
| 7        | Histidine | 92  |
| 8        | Isoleucine | 72  |
| 9        | Valine  | 31     |
| 10       | Arginine | 51   |
| 11       | Serine  | 31     |
| 12       | Proline | 31     |
| 13       | Methionine | 51  |
| 14       | Valine  | 31     |
| 15       | Glycine | 31     |
| 16       | Threonine | 51 |
| 17       | Phenylalanine | 50 |
| 18       | Tyrosine | 55    |
| 19       | Arginine | 31    |
| 20       | Threonine | 31  |
| 21       | Proline | 31     |
| 22       | Serine  | 31     |
| 23       | Proline | 31     |
| 24       | Aspartic acid | 31 |

* Yields from analysis of BCCP<sub>primo</sub>.

Table II

| Residue | Peptide/protein | Method | Assignment |
|---------|-----------------|--------|------------|
| Glx 3   | BCCP<sub>primo</sub> and SC | Sequencer* | Glutamic acid |
| Asx 24  | BCCP<sub>B</sub> | Sequencer* | Aspartic acid |
| Glx 30  | SP2b | CPA* | Glutamic acid |
| Glx 33  | T3a | CPA and B* | Glutamine |
| Glx 45  | SP3 | PTH | Aspartic acid |
| Asx 36  | T4 | HVE* | Aspartic acid |
| Glx 45  | SP3 | CPA* | Glutamic acid |
| Asx 39  | T4 | HVE* | Aspartic acid |
| Glx 45  | SP4 | CPA* | Glutamic acid |
| Glx 70  | SP6 | PTH | Aspartic acid |
| Glx 73  | SP6 | CPA* | Glutamic acid |
| Asx 75  | SP7 | PTH | Aspartic acid |
| Glx 76  | SP7 | PTH | Glutamic acid |
| Glx 82  | BCCP<sub>sc</sub> | CPA* | Glutamic acid |

* Identification of phenylthiohydantoin by gas-liquid chromatography from automated Sequencer analysis.

The amide assignments of all of the aspartyl and glutamyl residues of the protein are listed in Table II. Glx 3 and Asx 24 were assigned as acids from the sequenator analyses. Glx 30 was assigned as an acid based on the electrophoretic mobilities of SP2, position 30 was assigned as glutamine from digestion of T3a with carboxy-peptidase A and B. Asx 36 and Asx 75 were identified from the electrophoretic mobilities of T4 and Th7. Residue 45 was identified as glutamic acid from carboxypeptidase A hydrolysis of SP3 and residues 51 and 52 were identified as amides from the neutral electrophoretic behavior of Th10 at pH 6.5. Carboxypeptidase A hydrolysis of SP4 released 1 mol of glutamic acid per mol of protein. Since residue 53 is glutamic acid, it follows that residue 54 is glutamic acid. Asp 56 was identified from the thiohydantoin. Residues 67 and 73 were identified as glutamic acid by carboxy-peptidase A hydrolysis of SP5 and SP6, respectively. Glx 70, Asx 75, and Glx 76 were identified from the thiohydantoins. The COOH terminus of the protein was identified as glutamic acid by carboxypeptidase A digestion of BCCP<sub>sc</sub>.

Carboxypeptidase A Digestion — Digestion of BCCP<sub>sc</sub> with carboxy-peptidase A for 1 h at pH 8.5 released the following residues (mol/mol of protein): glutamic acid (1.00), valine (2.04), methionine (0.37), isoleucine (1.00), leucine (1.06).

Discussion

An unambiguous structure for the 82 residues comprising BCCP<sub>primo</sub> can be constructed from the peptides described. Sequenator analysis provides identification of the first 24 residues. Tryptic, thermolytic, and Staphylococcus aureus protease peptides derived from this region are consistent with the structure with respect to both composition and to residues positioned by sequence analysis. The extension of this region to peptide SP2b is provided mainly by the compositional overlap of SP2. The structure of SP2a has been partially determined by subdigestion with trypsin and is readily added to the completely structured SP2b to yield SP2. SP2b overlaps both T3 and T3a as well as Th6 and Th6a, which extends the sequence to Lys 34. These same peptides are in turn overlapped by SP3 which extends the sequence to Asn 39 in directly structured residues and by composition to Glx 45. This provides the information to overlap peptides T4 and T5, which are both fully sequenced. The compositional overlap provided by the COOH-terminal portion of SP3 is secured by the fact that it contains the only cysteinyl residue in the protein. The AlaMet sequence in the COOH-terminal position of T5 is overlapped by SP4, which extends the sequence through the biocytinyl residue at position 48 to Glx 54. The overlap of SP4 and SP5 is provided by Th11. This 2-residue overlap of Ile-Glu deserves further comment. Two other Ile-Glu sequences occur in the protein, at positions 29-30 and 51-52. However, the sequence at 29-30 is rigorously extended by peptides T3 and Th6 and therefore cannot be confused with the Ile-Glu at position 51-52. The sequence at 51-52, which forms the COOH terminus of the protein, is contiguously extended from position 53-54, as described below, thus making an overlap impossible. SP5, which was completely structured, extends the sequence to residue 67. Th12 provides a firm overlap to SP6. By composi-
tional analysis, Th12 overlaps the completely structured SP7, which provides the COOH-terminal fragment of the protein. Carboxypeptidase A hydrolysis of intact BCCPc confirms the COOH-terminal sequence.

The thermolytic and S. aureus protease digestions adhere tightly to the reported specificities of these enzymes (11, 20, 21), although, in the case of theromyosin, a number of potentially susceptible sites were not cleaved by the enzyme. With S. aureus protease, only one partial cleavage at aspartic acid was noted. It is of interest that the NH2-terminal glutamic acid residue of BCCPc was quantitatively removed, in contrast to the findings of Austen and Smith (21). Failure to cleave the Glu 76—Pro 77 bond is, however, in keeping with their findings.

The unusual chymotryptic-like cleavages observed in the second tryptic digest were probably associated with technical (contamination) problems rather than with any unusual properties of BCCP. This is supported by the results of the first tryptic digest, which apparently resulted in normal hydrolysates. The same batch of trypsin, however, when applied to other proteins, showed no unusual specificity.

The most interesting feature of the sequence is the cluster of methionyl residues surrounding the biocytin. In the absence of three-dimensional data, it is premature to assign a role to these residues. However, it seems plausible to suppose that they act to anchor the biocytin by extending inwards to the hydrophobic interior of the molecule. They may also act, at least in part, as a recognition site for the holoenzyme synthetase that specifically attaches the biotin cofactor to the lysyl residue (22).

The relationship of the BCCP fragments to the 22,500 molecular weight native polypeptide can only be surmised. At least in the case of BCCPc, some proteolysis clearly occurs at the NH2 terminus of the native protein. However, it is possible that additional proteolysis occurs at the COOH terminus. In this regard, it is of interest to note that carboxypeptidase A hydrolysis of intact BCCPc produced, in addition to quantitative release of residues 78 to 82, a partial residue (0.37 residue/mol) of methionine. Curiously, the composition of BCCPc shows 4.1 residues of methionine. In view of the fact that methionine values from acid hydrolysates are usually underestimates, this supports the possibility of a partial residue of methionine in the COOH-terminal position, and implies the existence of an additional COOH-terminal fragment in the 

\[ M_c = 22,500 \]

protein. Complete structural analysis of the native polypeptide will be required to elucidate this point.

Acknowledgments—We would like to thank Ms. Penny Shampaine and Ms. Sally Chandler for able technical assistance with the sequence analyses and Mr. Owen J. Bates and Dr. John S. Garavelli for assistance with the Sequencer analyses.

REFERENCES

1. Alberts, A. W., and Vagelos, P. R. (1968) Proc. Natl. Acad. Sci. U. S. A. 59, 561-568
2. Alberts, A. W., Nervi, A. M., and Vagelos, P. R. (1969) Proc. Natl. Acad. Sci. U. S. A. 66, 1319-1329
3. Dimroth, P., Guchhait, R. B., Stall, E., and Lane, M. D. (1970) Proc. Natl. Acad. Sci. U. S. A. 67, 1353-1360
4. Alberts, A. W., Gordon, S. G., and Vagelos, P. R. (1971) Proc. Natl. Acad. Sci. U. S. A. 68, 1289-1293
5. Guchhait, R. B., Moss, J., Sokolski, W., and Lane, M. D. (1971) Proc. Natl. Acad. Sci. U. S. A. 68, 653-657
6. Fall, R. R., and Vagelos, P. R. (1972) J. Biol. Chem. 247, 8005-8015
7. Fall, R. R., Nervi, A. M., Alberts, A. W., and Vagelos, P. R. (1971) Arch. Biochem. Biophys. 143, 401-411
8. Fall, R. R., and Vagelos, P. R. (1973) J. Biol. Chem. 248, 2078-2088
9. Shearer, W. T., Bradshaw, R. A., Gurd, F. R. N., and Peters, T., Jr. (1967) J. Biol. Chem. 242, 5451-5459
10. Bragshaw, R. A. (1969) Biochemistry 8, 3871-3877
11. Bragshaw, R. A., Garner, W. H., and Gurd, F. R. N. (1969) J. Biol. Chem. 244, 9169-9178
12. Hill, R. L., and Delaney, R. (1967) Methods Enzymol. 11, 339-351
13. Nakai, N., Lai, C. Y., and Horecker, B. L. (1974) Arch. Biochem. Biophys. 58, 563-570
14. Pisano, J. J., and Bronzert, T. J. (1969) J. Biol. Chem. 244, 5597-5607
15. Jacobs, J. W., and Niall, H. D. (1975) J. Biol. Chem 250, 3629-3636
16. Gray, W. H. (1972) Methods Enzymol. 25, 333-344
17. Hartley, B. S. (1970) Biochem. J. 119, 805-822
18. Sauer, R. J., Niall, H. D., Hogan, M. L., Keutmann, H. T., O'Riordan, J. L. H., and Potts, J. T., Jr. (1974) Biochemistry 9, 1994-1999
19. Houmard, J., and Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 3506-3509
20. Austen, B. M., and Smith, E. L. (1976) Biochem. Biophys. Res. Commun. 72, 411-417
21. McAllister, H. C., and Coon, M. J. (1966) J. Biol. Chem. 241, 2555-2561
22. Hirs, C. H. W. (1966) Methods Enzymol. 11, 197-199
**Sequence of E. coli Biotin Protein**

![Graph Image](http://www.jbc.org/)

**Table 1**

| Amino Acid | Position | Residues |
|------------|----------|----------|
| | 11-13 | Leu, Val, Leu |
| | 14-16 | Asp, Asp, Glu |
| | 17-19 | Gly, Thr, Ala |
| | 20-22 | Leu, Val, Leu |
| | 23-25 | Asp, Asp, Glu |
| | 26-28 | Gly, Thr, Ala |

In this table, the following amino acid compositions are given as part of a standard set of residues, and are based on 11-3 and 12-5 residues as described under "Correlation Parameters". Figures in parentheses show amino acid residues. By standard residue tests were determined to the desired sequence.

**Diagram 1**

![Graph Image](http://www.jbc.org/)

**Diagram 2**

![Graph Image](http://www.jbc.org/)

**Diagram 3**

![Graph Image](http://www.jbc.org/)

**Diagram 4**

![Graph Image](http://www.jbc.org/)
Amino acid sequence of Escherichia coli biotin carboxyl carrier protein (9100).
M R Sutton, R R Fall, A M Nervi, A W Alberts, P R Vagelos and R A Bradshaw
J. Biol. Chem. 1977, 252:3934-3940.

Access the most updated version of this article at http://www.jbc.org/content/252/11/3934

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/252/11/3934.full.html#ref-list-1