Vol. 85, No. 1, 1978 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 14, 1978 Pages 309-316

ASSIGNMENT OF THE ENDS OF THE β-CHAIN OF *E. COLI* TRYPTOPHAN SYNTHASE TO THE F_1 AND F_2 DOMAINS

Irving P. Crawford, Monique Decastel and Michel E. Coldberg

Department of Microbiology, University of Iowa, Iowa City, Iowa 52242, USA, and Unité de Biochimie Cellulaire, Département de Biochimie et Génétique :robienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

Received September 18, 1978

SUMMARY : Mild proteolytic treatment of the β_2 subunit of E. coli tryptophan synthase was previously reported to produce two large, non-overlapping polypeptide fragments, F_1 and F_2. Analyses of the N- and C-terminal sequences of the β chain and these proteolytic fragments now indicates that the F_1 fragment contains the intact N-terminal end and the F_2 fragment the intact C-terminal end of the native polypeptide chain. Therefore, any small peptide(s) released upon limited proteolytsis of the native enzyme comprise a connecting region between the F_1 and F_2 domains. The N-terminal sequence of the β-chain of E. coli tryptophan synthase is : Thr-Thr-Leu-Leu-Asn-Pro-Tyr-Phe-Gly-Glu-Phe-Gly-Gly-Met-Tyr-Val-Pro-Gln-Ile-Leu-Met-?-Ala-Leu-Leu(?)-Gln.

INTRODUCTION : Selective cleavage of polypeptide chains by limited proteolysis is proving a powerful tool in the study of protein structure. It is often used successfully in sequence determination. For example, in the recent case of rabbit muscle phosphorylase partial digestion by subtilisin and staphylococcal protease as well as trypsin proved very useful in ordering the cyanogen bromide peptides (1). Selective proteolysis has also recently turned out to be a very efficient way to approach deduction of the organization of the polypeptide chain in globular proteins. Among the more striking examples are the studies of Cohen and his colleages showing that each of the bifunctional Escherichia coli aspartokinase-homoserine dehydrogenases, I and II, consists of two domains per protomer, each domain carrying out one of the enzymatic functions of the whole protein (2,3).

Selective enzymatic cleavage of the β_2 subunit of E. coli tryptophan synthase was previously shown to produce two non-overlapping fragments, F_1 and F_2 (4), which should be useful in the determination of the amino acid sequence of the β-chain. These fragments have been shown to correspond to independently folding domains in the intact protein (5). Similarly, the polypeptide fragments produced by chain-terminating or deletion mutants in the gene for β-galactosidase have been shown to correspond to independently folding regions of the native enzyme (6). These polypeptides complement each other to give rise to enzymatically active molecules in vivo or in vitro (7). It could be envisaged

0006-291X/78/0851-0309$01.00/0
Copyright © 1978 by Academic Press, Inc.
All rights of reproduction in any form reserved

that the F_1 or F_2 fragments obtained from the β_2 subunit of tryptophan synthase might in a similar way complement missense mutant proteins in the trpB gene to produce active enzymes. For the design and interpretation of such experiments it is necessary to know which region of the trpB gene encodes a given domain, and specifically which domains correspond to the N- and C-terminal portions of the intact β-chain. This should be easily accomplished if after limited proteolytic treatment the N- and C-terminal ends of the β-chain remained intact or only slightly degraded, since the C-terminal sequence of the β-chain is known for 36 residues (8) and the N-terminus of the molecule has been reported to be blocked. Unfortunately, no chemical information is available concerning the blocking group or the N-terminal amino acid sequence (9).

In addition to allowing assignment of the N-terminal extremity to the F_1 or F_2 domains, knowledge of the N-terminal sequence would be of considerable value in interpretation of the nucleotide sequences near the beginning of the trpB gene under investigation in E. coli (T. Platt, personal communication) and Salmonella typhimurium (E. Selker and C. Yanofsky, personal communication).

For these reasons we began a study of the extremities of the intact polypeptide chain and its fragments. The results reported here demonstrate unambiguously that F_1 contains the N-terminus and F_2 the C-terminus of the native polypeptide and establish the sequence of 25 of the first 26 residues of the β-chain.

MATERIALS and METHODS : The β_2-subunit of E. coli tryptophan synthase was prepared by any of three very similar methods previously described (5,10,11). Identical proteolytic fragments F_1 and F_2 were obtained from each preparation and were purified as described earlier (5).

N-terminal residues were determined using the dansyl-chloride technique (12,13). Methods of peptide separation by ion exchange chromatography and composition analysis were described earlier (8).

Automated sequence analysis was accomplished with samples of 100 nmoles or more in a Beckman model 890C automatic sequencer using programs recommended by the manufacturer or described by Hermodson et al. (14). PTH-amino acids were obtained from the thiazolinones as described by Laursen et al. (15); they were identified either conventionally by gas-liquid chromatography and amino acid analysis following HI hydrolysis or by thin layer chromatography on silica plates and high pressure liquid chromatography according to Frank and Strubert (16) using a Siemens chromatograph.

RESULTS : The sequence of the C-terminal 35 residues of the β-chain, comprising cyanogen bromide fragment CB-3, was reported earlier (8). The molecular weight of this segment can be calculated to be 3975, close to the amount missing from the sum of the F_1 and F_2 fragments (4). Thus two possibilities exist ; either the peptide(s) missing from F_1 and F_2 should correspond very closely to CB-3 or the C-terminal sequence of either F_1 or F_2 should overlap at least part of the CB-3 sequence. In the latter case the portion over-

lapping CB-3 might contain at least some of the tryptic peptides (T-15, T-19, T-29, T-14) known to comprise 34 of the 35 residues of CB-3 (8). Therefore we undertook an analysis of the tryptic peptides of the F_1 and F_2 fragments.

Tryptic peptides of the F_2 fragment. Two experiments were performed using F_2 fragments that had been prepared and purified as described previously (5). Trypsinization was halted when 10-15 % of the enzymatic activity still remained. After dialysis in Spectrapor No. 3 tubing against water to remove urea and buffer salts, the F_2-containing fractions were performic acid oxidized by the Hirs method (17). The performic acid oxidized F_2 fragment (about 1.5 μmoles) was digested with 1 % (w/w) trypsin for 2 hr at 37° at pH 8. The digest was fractionated by cation-exchange chromatography as described for digests of the entire β-chain (18). About 10 ninhydrin reactive peaks were seen in the elution diagram (Fig. 1). When contamination was indicated by thin layer chromatography or composition analysis, the peptides in these peaks were further purified by anion exchange chromatography (18).

The yield and composition of the tryptic peptides obtained from F_2 are indicated in Table 1. All four of the peptides previously identified in CB-3 (T-15, T-19, T-29, and T-14) were found in high yeild. One additional peptide appears by composition and N-terminal analysis (Ala) to represent an incompletely digested T-29 - T-14 compound peptide. From these results we conclude that the F_2 fragment contains CB-3 and hence the known C-terminal sequence of the β-chain.

N-terminal sequence of the F_2 fragment. A single automatic sequencing attempt with unoxidized F_2 gave the results shown in Table 2. Of the first 26 residues, only residue 15 could not be identified with good confidence. It is noteworthy that this sequence contains no basic residues (hence should be found in T-40, the only tryptic peptide exceeding it in size) and contains as residues 3 and 4 a vicinal methionine pair, preceeding a second similar Met-Met sequence near the beginning of CB-3.

The smallest cyanogen bromide fragment (except for free homoserine) found in the intact β-chain (unpublished results) has the sequence Lys-Ala-Pro-Hse. Clearly this could arise from the 3 N-terminal residues of F_2 plus an immediately adjacent lysyl residue needed to provide the cleavage point for trypsin in the formation of the F_2 fragment. The immediately proximal segment of the β-chain should therefore have Met-Lys for a C-terminus.

N-terminal residue of the β-chain and the F_1 fragment. Having shown that the F_2 fragment contains the C-terminus of the β-chain, we compared the N-terminal residues of the F_1 fragment and the β-chain to learn whether the N-terminus of the native protein corresponds to F_1 or to a smaller peptide re-

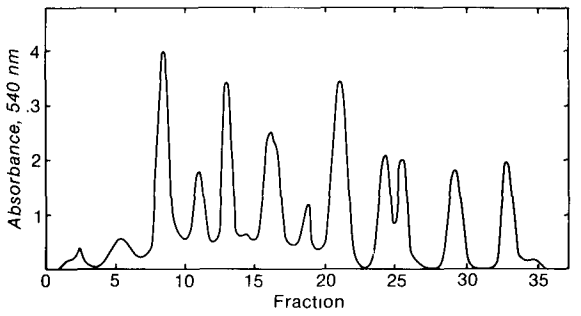

Figure 1. Elution diagram of trypsin digest of performic acid-oxidized F_2 fragment. A small portion of the eluent stream was reacted with nihydrin and the absorbance at 540 nm was continuously recorded. Unreacted fractions were dried and redissolved in 0.2 ml of water; examination of 1% of each fraction aided in deciding how the fractions should be analyzed for amino acid content. No significant amino acids were found in fractions 2 or 7; fractions 18 and 19 were not investigated.

Table 1

Tryptic peptides recovered from performic acid-oxidized F_2

Fraction	5-6	8-9	11	13	15-17[a]	21	24	25-26	29	32-33
Peptide	T-40[b]	T-41[b]	T-14	T-15	T-16	T-19	T-24	T-29-T-14	T-28	T-29
Yield[c]	150	151	460	236	515	549	607	450	697	384
Amino Acid										
Asx	4	2		3		3				
Thr	3	1			1	1				
Ser	5	1		1	2					
Glx	8	2	1	3	2			1		
Pro	4			1	1					
Gly	5	1	1	1	1	1	2	1		
Ala	4	2			4			1		1
$Cys.SO_3$					1					
Val	2	1		1		1	1			
$Met.SO_2$	2						1		2	
Ile	3	1	1		1	2	1	1		
Leu	3	2		3	7	1				
Tyr	2						1			
Phe	1	1				1	1			
His	1				3	1				
Lys	1			1	1	2	1			
Arg		1		1				1	1	1

a - This peptide required refractionation by Dowex-1 chromatography.

b - Compositions reported for these peptides are only approximate.

c - Yield expressed in nmoles is the amount actually observed in the fractions indicated (see Fig. 1).

Table 2

N-terminal sequences of the β-chain, the F_1 and the F_2 fragments

Residue Number	β-chain [a]	F_1 fragment [a]	F_2 fragment [d]
1	Thr	Thr	Ala
2	Thr	Thr	Pro
3	Leu	Leu	Met▲
4	Leu	Leu	Met▲
5	Asn	Asn	Glx△
6	Pro	Pro	Thr
7	Tyr	Tyr	Glx
8	Phe	Phe	Asx
9	Gly	Gly	Gly
10	Glu	Glu	Glx△
11	Phe	Phe	Ile
12	Gly	Gly	Glx△
13	Gly	Gly	Glx△
14	Met	Met	Ser[e]
15	Tyr	Tyr[b]	?
16	Val	Val or Phe[b]	Ser[e]
17	Pro	?[b]	Ile▲
18	Gln	Gln[b]	Ser[e]
19	Ile		Ala
20	Leu		Gly
21	Met		Leu
22	?		Pro
23	Ala		Phe
24	Leu		Pro
25	Leu[c]		Ser[e]
26	Gln		Val

a - Unless otherwise indicated, PTH amino acids were identified by both high pressure liquid chromatography and thin layer chromatography.

b - Identified by thin layer chromatography alone.

c - This residue, though not apparently ambiguous in any way by the methods used, is not in agreement with the initial tentative nucleotide sequence for this portion of the trpB gene (T. Platt, personal communication). Its assignment should be considered provisional until confirmed by additional amino acid or nucleotide sequence data.

d - PTH amino acids were identified either by gas-liquid chromatography▲ or by amino acid analysis△ or by both methods if not indicated.

e - Although alanine was found after HI hydrolysis, a corresponding amount of PTH-alanine was not detected by gas chromatography, leading to the conclusion that the residue in question is serine.

leased during limited proteolysis. Many earlier attempts to establish the N-terminal residue of the β-chain had failed or given equivocal results, suggesting that the amino group was blocked. Notwithstanding, generous samples of the intact protein and the F_1 fragment (12 nmoles each) were dansylated and the N-terminal residues determined by thin layer chromatography. In each case single spots of equivalent intensity were obtained, corresponding in position to serine or threonine (these two dansyl amino acids can barely be distinguished from each other in the solvent systems used) (12). Thus it appeared likely that the F_1 fragment and β-chain have the same N-terminal amino acid and that F_1 might comprise the N-terminal portion of the intact protein.

N-terminal sequence of the β-chain and F_1 fragment. To verify this conclusion the N-terminal sequences of the intact protein and the F_1 fragment were compared by automatic sequence determination. The first run was continued in each case as long as the amino acids could be determined without amibiguity (Table 2). A second run was performed for both the β-chain and F_1 to the sixth residue to remove all doubt about the identity of the sequences.

Because the first two residues of each chain are threonine and this amino acid yields several peaks on high pressure liquid chromatography (16), the yield at the first two steps could not be determined directly. The amount of leucine released at step 3 could be determined in three of the runs, however. This yield was found to be above 40 % for the β-chain, while it was 35 % and 70 % in the two runs with the F_1 fragment. That the yields found for F_1 and the β-chain were similar indicates that the N-terminal sequence observed for the β-chain did not result from some contamination with F_1 fragment and vice versa. These data clearly establish that F_1 and the β-chain of tryptophan synthase have the same N-terminal sequence for 18 residues, and establish this sequence as Thr-Thr-Leu-Leu-Asn-Pro-Tyr-Phe-Gly-Glu-Phe-Gly-Gly-Met-Tyr-Val-Pro-Gln.

DISCUSSION : The results reported above clearly demonstrate that the F_1 fragment is the N-terminal part and the F_2 fragment the C-terminal part of the β-chain. Hence, the fragment(s) of low molecular weight released during limited proteolysis of the β_2 subunit correspond to the "connecting region" between the F_1 and F_2 domains in the native enzyme.

A systematic analysis of the cyanogen bromide and tryptic peptides of the F_1 and F_2 fragments should allow establishment of three classes among the peptides obtained in the intact protein, those belonging to F_1, F_2 and the connecting region. For instance, in addition to T-14, T-29, T-15, and T-19 which had already been shown to be at the C-terminal end of β, the tryptic peptides T-28, T-16, T-24, T-41, and T-40 have also been identified as belonging to the F_2 region. Similarly, the N-terminal sequence of the β-chain shows that two cyanogen bromide peptides, CB-4 and CB-7, previously isolated from the β-chain and partially sequenced by one of us (unpublished results) correspond to residues 1-14 and 15-21 of the intact protein. No doubt the possibility of assigning other tryptic and cyanogen bromide peptides to a given region of the β-chain will be of considerable help in sequencing the protein.

At this time little can be said about the "blocked" N-terminus of the β-chain observed earlier. Probably different preparations vary in the degree of substitution, if any, of the amino group of the threonyl N-terminus. Neither can it be established how large a segment of the polypeptide chain in addition to the initiating formyl-methionyl residue is removed in processing the poly-

peptide. Two likely approaches to establishing the extent of this processing would involve in vitro protein synthesis, where processing may not occur at all, or the establishment of the nucleotide sequence preceeding the two threonine codons in the trpB gene. It would be interesting to know why among the 5 polypeptides produced by the trp operon only the trpB gene product, as obtained after purification, lacks the initial amino acid(s).

Finally, the results obtained in the present study give some information about the functional importance of connecting region in the native enzyme. The nicked protein which, as shown here lacks only the connecting region, retains the binding sites for substrates and coenzymes (4,19) but is no longer able to associate with the α-chain (4). It is tempting therefore to conclude that the binding site for α is related to this connecting region, which would then appear to be heavily involved in the functional activities of the β_2 subunit either directly or by constraining the topology of the F_1 and F_2 domains.

ACKNOWLEDGMENT. We are grateful for assistance in automatic sequence determination from M. De Wolf, Institut Pasteur, Paris, France, and P. Siegrist, Scripps Clinic and Research Foundation, La Jolla, California. This work was supported by funds from the Centre National de la Recherche Scientifique, the Délégation Générale à la Recherche Scientifique et Technique, the Université Paris VII, the Commissariat à l'Energie Atomique, and the Fondation pour la Recherche Médicale Française (M.G.), as well as National Institutes of Health Grant GM-19098 (I.P.C.).

REFERENCES.

1. Titani, K., Koide, A., Hermann, J., Ericsson, L.H., Kumar, S., Wade, R.D., Walsh, K.A., Neurath, H., and Fischer, E.H. (1977) Proc. Natl. Acad. Sci. USA, 74, 4762-4766.
2. Veron, M., Falcoz-Kelly, F., and Cohen, G.N. (1972) Eur. J. Biochem., 28, 520-527.
3. Dautry-Varsat, A., and Cohen, G.N. (1977) J. Biol. Chem., 252, 7685-7689.
4. Högberg-Raibaud, A., and Goldberg, M.E. (1977) Proc. Natl. Acad. Sci. USA, 74, 442-446.
5. Högberg-Raibaud, A., and Goldberg, M.E. (1977) Biochemistry, 16, 4014-4020.
6. Celada, F., Ullmann, A., and Monod, J. (1974) Biochemistry, 13, 5543-5547.
7. Ullmann, A., and Perrin, D. (1970) The Lactose Operon (J. Beckwith and D. Zipser, eds.), Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y., pp. 143-172.
8. Cotton, R.G.H., and Crawford, I.P. (1972) J. Biol. Chem., 247, 1883-1891.
9. Yanofsky, C., and Crawford, I.P. (1972) The Enzymes, 3rd Edition (P.D. Boyer, ed.), Academic Press, New York, pp. 1-31.
10. Wilson, D.A., and Crawford, I.P. (1965) J. Biol. Chem., 240, 4801-4808.
11. Miles, E.W., and Moriguchi, M. (1977) J. Biol. Chem., 252, 6594-6599.
12. Zanetta, J.P., Vincendon, G., Mandel, P., and Gombos, G. (1970) J. Chromatog., 51, 441-458.
13. Gray, W.R. (1967) Methods Enzymol., 11, 469-475.
14. Hermodson, M.A., Ericsson, L.H., Titani, K., Neurath, H., and Walsh, K.A. (1972) Biochemistry, 11, 4493-4502.
15. Laursen, R.A., Bonner, A.G., and Horn, M.J. (1975) In : Instrumentation in amino acid sequence analysis, (R.N. Perham, ed.), Academic Press, p. 92.
16. Frank, G., and Strubert, W. (1973) Chromatographia, 6, 522-524.

17. Hirs, C.H.W. (1956) J. Biol. Chem. 219, 611-621.
18. Fluri, R., Jackson, L.E., Lee, W.E., and Crawford, I.P. (1971) J. Biol. Chem. 246, 6620-6624.
19. Decastel, M., and Goldberg, M.E. (1978) FEBS Letters, 89, 64-68.