PRIMARY STRUCTURE OF CHOLERA TOXIN B-SUBUNIT

C.Y. Lai, E. Mendez[1], D. Chang and M. Wang[2]

Roche Institute of Molecular Biology
Nutley, New Jersey 07110

Received November 5,1976

SUMMARY: The primary structure of cholera toxin B-subunit, responsible for the binding of the toxin to cell surfaces, has been elucidated. The polypeptide contains 103 amino acid residues and one intra-chain disulfide bridge between Cys 9 and Cys 86. The molecular weight is calculated to be 11,637, 15-20% higher than the values estimated by physicochemical methods. This value is consistent with a structure containing five moles of B-subunits per mole of cholera toxin.

Cholera toxin, a diarrheagenic protein isolated from the culture filtrate of Vibrio cholerae, has been shown to contain two types of subunits, A and B, in a molar ratio of one A-subunit to five or six B-subunits. The B-subunit has a high and specific affinity to ganglioside GM1 found on most cell surfaces, and is considered to be responsible for the attachment of cholera toxin to intestinal cell walls during intoxication (4,5). To provide information on the structural basis for this affinity as well as the interactions between subunits in the holotoxin or choleragenoide (6), the sequence analysis of the B-subunit was undertaken. This communication reports the conclusion of these studies[3].

The sequence of 30 amino acids around the single disulfide bridge in the B-subunit, 23 of which correspond to the NH_2-terminal segment of the polypeptide, suggested the existence of some structural analogy between these regions of cholera toxin B-subunit and glycoprotein hormones (7,8). Inspection of the complete sequence, reported here, did not reveal any additional analogies.

MATERIALS AND METHODS

Purified cholera toxin was prepared in our laboratory from the culture filtrate of Vibrio cholerae provided by the Geographic Medicine Branch of the

[1] Present address: Centro de Investigaciones Biologicas, Instituto de Biologia del Desarrollo, Madrid, Spain.

[2] Present address: Department of Medicine, Albert Einstein College of Medicine, Bronx, New York

[3] Full details of the experiments will be published elsewhere.

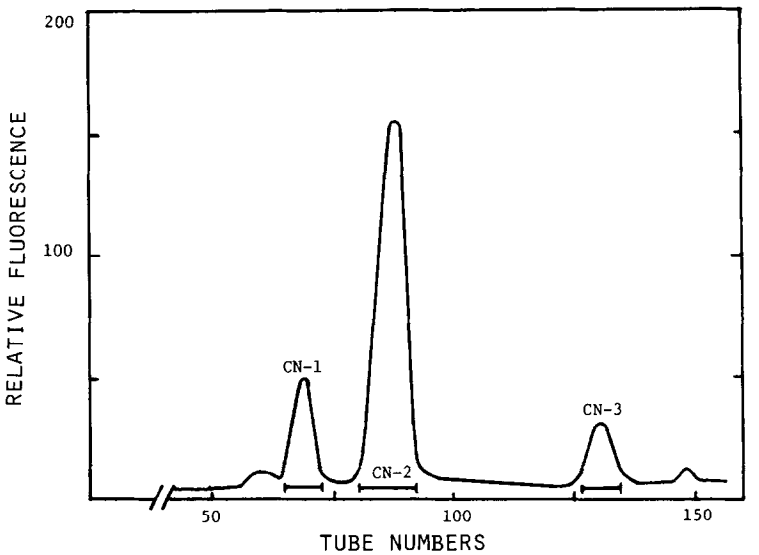

Figure 1. Sephadex G50F gel filtration of BrCN-peptides from the B-subunit. S-carboxymethyl B-subunit (20 mg) in 0.5 ml 70% HCOOH was treated with 0.15 ml of BrCN (50 mg/ml in 70% HCOOH) for 18 hrs at room temperature. After lyophilization, the cleavage product was dissolved in 3 ml of 0.1 M NH_4-formate containing 4% (v/v) HCOOH, and gel filtered in the same solution on a Sephadex G50F (1.5 x 190 cm) at a flow rate of 14 ml/hr. Aliquots (20 µl) from every other tube (2.3 ml) were reacted with fluorescamine (11) to detect peptides. Fractions indicated with bars were combined and lyophilized. The CN-3 fraction contained the dipeptide, Ala-Asn, in approximately 80% yield.

National Institutes of Health through the courtesy of Dr. Carl E. Miller. Subunits were separated by gel filtration in 5% HCOOH as previously reported (3).

Proteolytic enzymes used in the sequence analyses were obtained from Worthington Biochemical, Freehold, N.J. All reagents were of the highest purity available commercially and used without further purification. Solvents such as pyridine, benzene and ethyl acetate were distilled and used within 6 months.

Isolated B-subunit of cholera toxin, reduced and S-carboxymethylated according to Crestfield (9) in 6 M guanidine HCl, was the starting material. Peptides obtained by proteolytic digestion of S-carboxymethyl B-subunit were fractionated by gel filtration on Sephadex G25F, then purified by Dowex 50 and/or Dowex 1 chromatography essentially as described by Schroeder (10), and by high voltage electrophoresis on Whatman 3 MM paper at the indicated pH in a Savant apparatus (Savant Instruments, Hicksville, N.Y.). Peptides were detected by methods previously described using fluorescamine (11,12). Amino acid analyses were carried out according to Spackman, et al. (13), with the JEOL 6AH or 5AH acid analyzer.

Edman degradation (14) was carried out essentially as described previously (15). The residue removed as thiazolinone at each degradation step was identified by analysis of the regenerated amino acid (16). Portions of the peptides were removed after each degradation and amino acid analysis performed in case the above identification was not successful.

The presence of side chain amide groups on aspartic or glutamic acids were determined by the method of Offord (17).

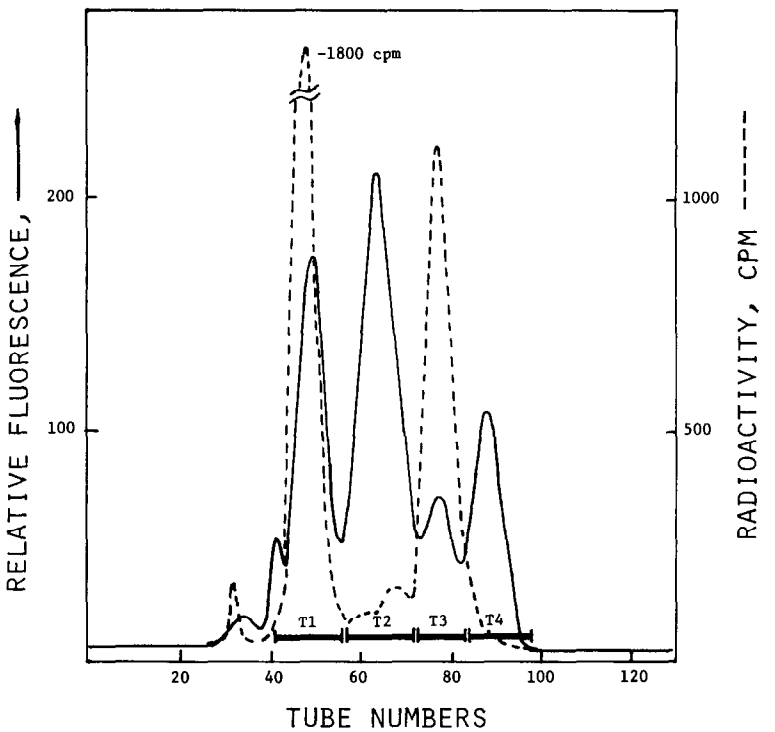

Figure 2. Fractionation of the tryptic digest of S-carboxymethyl B-subunit. S-carboxymethyl B-subunit (12.5 mg) in 2 ml H_2O was treated with 0.05 ml of TPCK-trypsin (2 mg/ml) as described in the text and applied to a Sephadex G25F column (2.2 x 90 cm) in 0.1 M NH_4HCO_3. Gel filtration was carried out at a flow rate of 40 ml/hr. Aliquots (50 µl) from every other tube (2.4 ml) were assayed with the fluorescamine method (11). Fractions T1-T4 were lyophilized and fractionated further as indicated in Table 1.

RESULTS

NH_2- and COOH-terminus of the cholera toxin B-subunit: Two steps of Edman degradation performed on the B-subunit successively yielded threonine (as α-aminobutyric acid, 42%) and proline (75%). This indicated that the B-subunit was a single chain polypeptide with NH_2-terminal sequence of Thr-Pro-, as previously reported (7,18,19).

No amino acid was released on hydrazinolysis of the B-subunit, suggesting that the COOH-terminus of this peptide was either glutamine or asparagine (20). The COOH-terminal sequence was established as Met-Ala-Asn when dipeptide, Ala-Asn, was isolated in near quantitative yield after treatment of the B-subunit with BrCN (Fig. 1), and its sequence determined by Edman degradation.

TABLE 1. Summary Data on the Purified Tryptic Peptides

Separation Procedures[1]	DEAE-Cellulose Chromatography, pH 8			Dowex 50 x 2									Dowex 50 x 2			
	NH_4HCO_3 = 0.05 M - 0.15 M Linear Gradient						Dowex 1						Paper Electr. pH 3.5			Paper Electr. pH 3.5
Peptides	T1a	T1b	T1c	T2a,b	T2d,e	T2f	T2g1	T2g2	T2h	T2i	T2j,k	T2l	T3a1	T3c	T3d	T4a
Lysine[2]	1.9	1.1	1.1			1.0	1.0	1.0	1.1		0.9	1.0	1.0	1.1	1.0	
Histidine	0.9	0.9	1.9		0.9											
Arginine										1.0	0.9	0.8				1.0
CM-Cystein			0.5											0.8		
Aspartic Acid	2.3	2.2	5.1	1.2	1.3						0.9			2.2		
Threonine	0.9	0.9	3.6		0.8	0.7		1.0	0.9		1.1	0.7				
Serine	1.9	1.9		1.1	0.9				1.9			1.7				
Glutamic Acid	3.8	4.2	3.2			1.0	1.0	1.0	1.0	1.0		1.1				
Proline	1.0	1.0	1.0		1.0											
Glycine	1.8	2.0							1.0			1.0				
Alanine	1.1	1.1	1.0	3.6	3.6	1.0		1.8	1.0	1.0		1.1				
Valine	1.8	2.0					0.9							1.0		
Methionine				1.0	1.0	0.5					0.9				1.0	
Isoleucine	1.0	1.03	2.0	1.8	1.7	1.54		0.9	0.9	0.9		0.9				
Leucine			2.2					0.9	0.9		0.9	1.1		1.0		
Tyrosine			0.9					0.6	0.9			0.7				
Phenylalanine	0.8	0.9				0.8			0.9			0.9				
Tryptophan														(1)[3]		
Total Residues	20	19	23	9	12	8	3	8	11	4	6	12	1	7	2	1
Yield, %[4]				24	42	33	56	52	52	90	46	25	57	37	4	65

[1] Four fractions obtained from gel filtration on Sephadex G25F, T1, T2, T3 and T4 (Fig. 2), were further fractionated by the procedures indicated. Peptides were designated as described in the text with small letters followed with numbers.

[2] Numbers of residues per mole of peptide as determined by amino acid analysis are shown.

[3] The numbers in parentheses are values estimated by other methods.

[4] Overall yields of peptides, without correction for the loss during manipulation.

Tryptic peptides from the B-subunit: S-carboxymethyl B-subunit was digested for 3 hr at room temperature with TPCK-trypsin (15) (substrate/enzyme = 100/1,w/w) at pH 8. The digest was filtered on a Sephadex G25F column yielding four fractions (Fig. 2), designated T1-T4 in order of their elution from the column. Separation of peptides in these fractions was achieved by DEAE-cellulose chromatography,

TABLE 2. BrCN Peptides and Chymotryptic Peptides

Separation Procedures	Sephadex G50		Sephadex G25[2]		Sephadex G25[3]					
	DEAE[1] Urea		Dowex 50		Dowex 1		Dowex 50			
Peptides	CN2d,e	CN3	C2i	C2k	cC1b	cC1f	cC2a	cC2d	cC2f	cC2i
Lysine	2.1[4]		1.2	1.1	3.4	2.0			1.0	0.9
Histidine	1.9				1.0					1.0
Arginine	1.0			1.0	1.0					
CM-Cystein						0.6				
Aspartic Acid	5.1	1.2	2.2		1.9		1.0		1.2	2.1
Threonine	4.6		0.8			0.9		0.8	1.0	0.9
Serine	2.1				1.7		0.9			
Glutamic Acid	5.5			1.1	5.2	2.0				
Proline	0.9				1.0					
Glycine	1.1			1.2	1.1				1.0	
Alanine	2.4	1.0		1.0	1.1	1.0	4.0	1.0	0.9	
Valine					1.5	1.3				
Methionine	(1)			(1)	0.4					
Isoleucine	2.8		1.0		2.0		1.9	1.1		
Lucine	2.9		1.0		0.6	1.6				
Tyrosine	1.7									
Phenylalanine	1.0		1.0					0.9	0.9	
Tryptophan						(1)				
Total Residues	37	2	7	6	44	12	9	5	6	5
Yields, %[5]	42	80	45	27	43	66	23	50	18	23

[1] The fraction, CN-2, from Sephadex G50F (Fig. 1), was further separated by DEAE-Sephadex A-50 chromatography in 0.02 M Tris HCl, pH 8.5 containing 3 M urea, with a NaCl gradient.

[2] Chymotryptic peptides of the B-subunit were first fractionated on Sephadex G25F. Fraction C2 was separated by Dowex 50 x 2 chromatography (10) to yield C2i and C2k.

[3] Chymotryptic peptides from the citraconylated B-subunit. Citraconylation was carried out in 0.1 M $NaHCO_3$ containing 4 M guanidine HCl (10 mg protein/ml), with pH maintained at 8.5 by addition of 0.5 M NaOH. After desalting on a Sephadex G25F column (1 x 26 cm) in 0.1 M NH_4HCl_3, chymotrypsin was added (substrate/enzyme = 40) and digestion carried out for 4 hrs at 35°. Fraction cC1 was separated on Dowex 1 and cC2 on Dowex 50. Peptides were designated with the cC prefix.

[4] Numbers are expressed in residues per mole peptide. Those in parentheses were estimated values.

[5] Overall yields of peptides, without correction for the loss during manipulation.

Dowex 50 chromatography or high voltage paper chromatography as indicated in Table 1. Peptides obtained from each of the four Sephadex fractions (Fig. 2) were designated with small letters (e.g. T1a, T1b, T2a, T2b, etc.) in order of their emergence from the column or electrophoretic mobilities on paper. When

further separation was necessary, paper electrophoresis was employed. The peptides so obtained were then designated by numbers affixed to the designate of the fraction from which they were isolated, e.g. T2g1, etc. Amino acid compositions and the overall yields of the tryptic peptides are summarized in Table 1. Peptides apparently derived by the chymotrypsin-like activity of trypsin, i.e. cleavage of peptide bonds at carboxyl ends of residues other than lysine and arginine, have been omitted from the table. The total number of amino acid residues was found to be about 15% larger than that originally calculated from the assumed molecular weight of 10,000 for this protein (3).

Peptides from BrCN or chymotryptic cleavages: Gel filtration of the BrCN cleavage products from the B-subunit yielded one major peak (CN-2) and dipeptide, Ala-Asn (Fig. 1). Edman degradation performed on CN-2 revealed the presence of three peptides with NH_2-terminus of Thr, Ala and Lys, respectively, in approximately equal quantities. Attempts to separate all three peptides were unsuccessful, but resulted in the isolation of the NH_2-terminal segment of 37 residues, CN-2d,e (Table 2). This peptide yielded T1c, T2h, free Arg (Table 1) and dipeptide, Glu-Ser, on tryptic digestion.

Chymotryptic peptides from the S-carboxymethylated B-subunit were separated essentially as described above for the tryptic peptides. Those containing amino acid residues which overlap with two or more tryptic peptides are listed in Table 2. Useful information on the arrangement of the tryptic peptides as well as on the overall sequence was provided by digestion of the citraconylated S-carboxymethyl B-subunit with chymotrypsin. Peptides isolated from this digest were designated with the prefix "cC" (Table 2).

Complete primary structure of cholera toxin B-subunit: Based on the sequence analyses of the peptides listed in Tables 1 and 2, the results of which are summarized and included in Fig. 3, the complete sequence of 103 amino residues in subunit B of cholera toxin has been deduced[4]. The molecular weight calculated from the sequence data was 11,637.

[4] While this manuscript was in preparation, we learned that the sequence analyses on this protein had been completed independently by Kurosky and coworkers. The results presented here have been confirmed (Personal communication: A. Kurosky, University of Texas Medical Branch, Galveston, Texas).

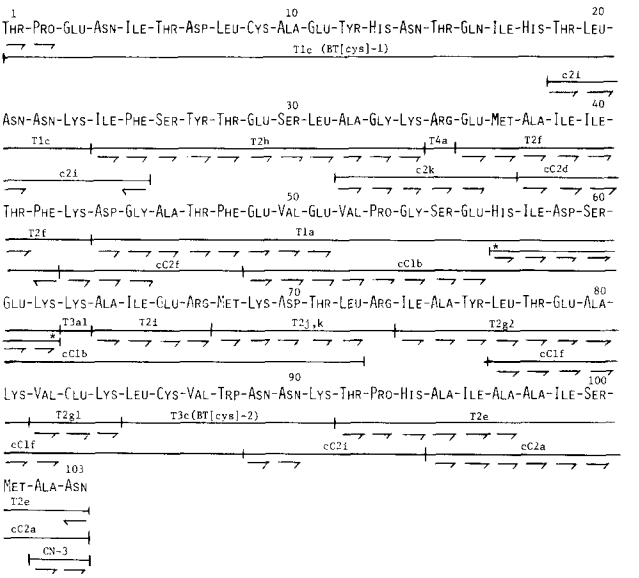

Figure 3. Primary structure of cholera toxin B-subunit. Information which has led to the elucidation of the complete sequence of 103 amino acid residues has been summarized here. Sequence of the cysteine-containing peptides T1c and T3c have been reported previously (7) and not explained here. The peptide 57-62 (*) was isolated from the chymotryptic digest of T1a. Arrows pointing to the right denote the sequence determined by Edman degradation procedure. Those pointing to the left, denote the COOH-terminal residue identified by hydrazinolysis of the peptide and amino acid analysis.

DISCUSSION

Some peptides obtained by chymotryptic digestion contained Lys or Arg at the NH_2-termini, providing insufficient overlap for the alignment of the tryptic peptides. Citraconylation of the B-subunit prior to the chymotrypsin digestion circumvented this problem.

The molecular weight of the B-subunit has been calculated as 11,637 from the completed sequence information. The figure is nearly 20% higher than the values estimated by physicochemical methods (1-3). Since the molecular weight of cholera toxin and the A-subunit have been estimated to be 84,000 and 28,000-30,000, respectively, the figure obtained in this study for the B-subunit supports the notion that there are five B-subunits in a molecule of the native toxin (21). We have previously proposed that the subunit composition of cholera toxin could be A_1B_5 (3).

Visual inspection of the amino acid sequence of the cholera toxin B-subunit and that of the β-chain of thyrotropin (22) did not reveal further analogies other than those pointed out by Ledley, et al. (8). Functional implication of the primary structure is of great interest, and a careful comparison between the structure of cholera toxin and that of receptor binding proteins using a computer program are in order. Meanwhile, studies on the "active site" of the cholera toxin B-subunit using chemical modifications are in progress in our laboratory.

REFERENCES

1. Lönnroth, I., and Holmgren, J. (1973) J. Gen. Microbiol. 76, 417-427.
2. Finkelstein, R.A., Boesman, M., Neoh, S., LaRue, M.K., and Delaney, R. (1974) J. Immun. 113, 145-150.
3. Lai, C.Y., Mendez, E., and Chang, D. (1976) J. Infect. Dis. 133, S23-S30.
4. Holmgren, J., Lönnroth, I., and Svennerholm, L. (1973) Scand. J. Inf. Dis. 5, 77-78.
5. King, C.A., and Van Heyningen, W.E. (1973) J. Infect. Dis. 127, 639-647.
6. Finkelstein, R.A., and LoSpalluto, J.J. (1969) J. Exp. Med. 130, 185-202.
7. Mendez, E., Lai, C.Y., and Wodnar-Filipowicz, A. (1975) Biochem. Biophys. Res. Commun. 67, 1435-1443.
8. Ledley, F.D., Mullin, B.R., Lee, G., Aloj, S.M., Fishman, P.H., Hunt, L.T., Dayhoff, M.O., and Kohn, L.D. (1976) Biochem. Biophys. Res. Commun. 69, 852-859.
9. Crestfield, A.M., Moore, S., and Stein, W.H. (1965) J. Biol. Chem. 238, 622-627.
10. Schroeder, W.A. (1967) in Methods in Enzymology, Vol. XI (C.H.W. Hirs, ed.), pp. 361-369, Academic Press, New York.
11. Nakai, N., Lai, C.Y., and Horecker, B.L. (1974) Anal. Biochem. 58, 563-570.
12. Mendez, E., and Lai, C.Y. (1975) Anal. Biochem. 65, 281-292.
13. Spackman, D.H., Stein, W.H., and Moore, S. (1958) Anal. Chem. 30, 1190-1206.
14. Edman, P. (1956) Acta Chem. Scand. 10, 761.
15. Lai, C.Y. (1975) Arch. Biochem. Biophys. 166, 330-338.
16. Mendez, E., and Lai, C.Y. (1975) Anal. Biochem. 68, 47-53.
17. Offord, R.E. (1966) Nature (London) 211, 591-593.
18. Jacobs, J.W., Niall, H.D., and Sharp, G.W.G. (1974) Biochem. Biophys. Res. Commun. 61, 341-345.
19. Kurosky, A., Markel, D., Touchstone, B., and Peterson, J.W. (1975) Proc. Amer. Soc. Microbiol. 75, 23.
20. Fraenkel-Conrat, H., and Tsung, C.M. (1967) in Methods in Enzymology, Vol. XI (C.H.W. Hirs, ed.), pp. 151-155, Academic Press, New York.
21. Gill, M. (1976) Biochemistry 15, 1242-1248.
22. Dayhoff, M.O., Hunt, L.T., McLaughlin, P.J., and Barker, W.C. (1972) Atlas of Protein Sequence and Structure 5, p. D200, Nat. Biochem. Res. Foundation, Washington, D.C.