SEQUENCE SIMILARITY BETWEEN CHOLERA TOXIN AND GLYCOPROTEIN HORMONES: IMPLICATIONS FOR STRUCTURE ACTIVITY RELATIONSHIP AND MECHANISM OF ACTION

Fred D. Ledley,[1] Brian R. Mullin,[1] George Lee,[1] Salvatore M. Aloj,[1] Peter H. Fishman,[2] Lois T. Hunt,[3] Margaret O. Dayhoff,[3] and Leonard D. Kohn[1]

[1] Section on Biochemistry of Cell Regulation, Laboratory of Biochemical Pharmacology, National Institute of Arthritis, Metabolism, and Digestive Diseases, National Institutes of Health, Bethesda, Maryland 20014, U.S.A.

[2] Developmental and Metabolic Neurology Branch, National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health.

[3] National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C. 20007, U.S.A.

Received February 3,1976

SUMMARY: The B chain of cholera toxin and the β subunits of thyrotropin, luteinizing hormone, human chorionic gonadotropin, and follicle-stimulating hormone are shown to have a region of sequence analogy believed to correlate with their ability to bind to receptors on cell membranes. A possible sequence analogy is also defined in the α subunits of these glycoprotein hormones and a region of the cholera toxin A_1 chain believed to be responsible for adenylate cyclase activation.

The native cholera toxin molecule is reported to be a polymer of six B chains and one A subunit; the A subunit is believed to consist of A_1 and A_2 chains which are linked by a single disulfide bond (1-5). At present it is believed that the B chain of cholera toxin binds specifically to a membrane ganglioside, G_{M1},* which serves as its receptor (1-5); that the binding of the B chain to G_{M1} causes a change in the molecular conformation of the intact toxin molecule with the resultant formation of an "active" A subunit; and that the "active" A subunit translocates within the cell membrane and activates adenylate cyclase by direct interaction (5).

---

* Abbreviations: TSH, thyrotropin; LH, luteinizing hormone; HCG, human chorionic gondatropin; FSH, follicle-stimulating hormone; G_{M3}, N-acetylneuraminyl galactosylglucosylceramide; G_{M2}, N-acetylgalactosaminyl-(N-acetylneuraminyl)-galactosylglucosylceramide; G_{M1}, galactosyl-N-acetylgalactosaminyl-(N-acetylneuraminyl)-galactosylglucosylceramide; G_{D1a}, N-acetylneuraminylgalactosyl-N-acetylgalactosaminyl-(N-acetylneuraminyl)-galactosylglucosylceramide; G_{D1b}, galactosyl-N-acetylgalactosaminyl-(N-acetylneuraminyl)-galactosylglucosylceramide; G_{T1}, N-acetylneuraminylgalactosyl-N-acetylgalactosaminyl-(N-acetylneuraminyl-N-acetylneuraminyl)-galactosylglucosylceramide.

In a recent report (6) we suggested that an analogous mechanism holds for the glycoprotein hormone, thyrotropin (TSH). Thus, specific ganglioside-TSH binding interactions were demonstrated and shown to correlate with conformational changes in the TSH molecule (6). These interactions were critically affected by the number and position of the sialic acid residues on the carbohydrate portion of the ganglioside structure, the efficacy of binding having an order distinct from that exhibited by cholera toxin, *i.e.*, $G_{D1b} > G_{T1} > G_{M1} > G_{M2} = G_{M3} > G_{D1a}$. Last, gangliosides with the chromatographic properties of G_{D1b}, G_{T1}, and G_{M1} were present in thyroid plasma membranes in higher quantities than had been previously found in extraneural tissue (6).

The present report supports the hypothesis that cholera toxin and TSH are similar in their mode of action on the membrane, by documenting a sequence analogy between the B chain of cholera toxin and the β subunit of TSH and between the A_1 chain of cholera toxin and the α subunit of TSH. Similar analogies are demonstrated in the other glycoproteins of the TSH-LH-HCG-FSH superfamily (7); however, differences in the FSH sequence adjacent to the sequence analogy on the β subunit may explain the relative absence of FSH interactions with TSH receptors (8-12).

METHODS

The sequence of cholera toxin used included the 42 amino terminal residues of the B chain as reported by Kurosky *et al.* (13)† and the A chain sequences recently reported by Mendez *et al.* (14). All other sequences were from the tapes of the *Atlas of Protein Sequence and Structure* (15-20) updated to December 5, 1975. The computer methods were those previously reported for sequence analysis (15) and involved the use of the programs SEARCH, ALIGN, and RELATE. The computer work was done on the IBM 360/44 computer at the National Biomedical Research Fouundation.

The program SEARCH compares a given peptide sequence with every possible peptide of similar length in the data library, and prints out those sequences having the greatest similarities. This is used as a first approximation of possible relationships and to highlight significant analogies for more detailed analysis. The program ALIGN, based on the algorithm of Needleman and Wunsch

† We are also indebted to Dr. R. A. Finkelstein, University of Texas Southwestern Medical School, Dallas, Texas 75235, for providing us with his currently available sequence data for the A and B components of cholera toxin and for his helpful advice during the course of our initial observations in this area.

(21), determines the highest possible score for any alignment (including gaps) of two sequences based either on a unit matrix (identity comparison) or mutation data matrix (mutational probability). This score is then compared with the highest possible scores obtained by aligning pairs of randomized sequences having the same amino acid composition as the two real sequences. The scores of the randomized sequences form a normal distribution from which the mean and standard deviation are computed. The probability that the score for the real sequence is derived from this normal (random) distribution can then be obtained and expressed as the number of standard deviations of the real score from the mean random score.

The fragment comparison program RELATE compares all fragments of a given length from one sequence with all possible fragments of the same length from the second sequence. A score is computed for each trial based either on the number of absolute identities (unit matrix) or relative mutational identities (mutation data matrix). If the proteins are related, there will be two populations of scores: a large population from trials between unrelated fragments of the molecules, and a smaller population from related sequences. The average of the latter scores represents the score of the whole molecule comparison. An identical analysis is performed on randomized sequences having the same amino acid composition and, as with the program ALIGN, the probability that the score for the real sequences is derived from the normal (random) distribution is computed. The results are expressed as the number of standard deviations of the real score from the mean random score.

RESULTS AND DISCUSSION

Only the first 42 amino acid residues from the amino terminal end of the B chain of cholera toxin have been sequenced (13). An initial screening was performed for sequences with any possible similarity to this portion of the cholera toxin B chain using the program SEARCH and the data banks of the *Atlas of Protein Sequence and Structure* (15-20). The results of this evaluation showed that the CAEY region of cholera toxin B chain (residues 9-12) was analogous to a CAGY region of the β subunits of the glycoprotein superfamily (TSH residues 27-30, LH residues 34-37, HCG residues 34-37, and FSH residues 32-35). An alignment of peptides based on the superposition of these regions (Fig. 1, top) gives a qualitative first approximation of the analogy of this region.

A quantitative measure of the significance of the analogy was computed using the fragment comparison program RELATE to compare residues 1 to 42 of the cholera toxin B chain with residues 19-60, 26-67, 26-67, and 24-65, respectively of the β subunits of TSH, LH, HCG, and FSH. The results confirm the significance of the analogies between cholera toxin, TSH, LH, and HCG (Table I) but did not indicate any significant analogy between cholera toxin and FSH.

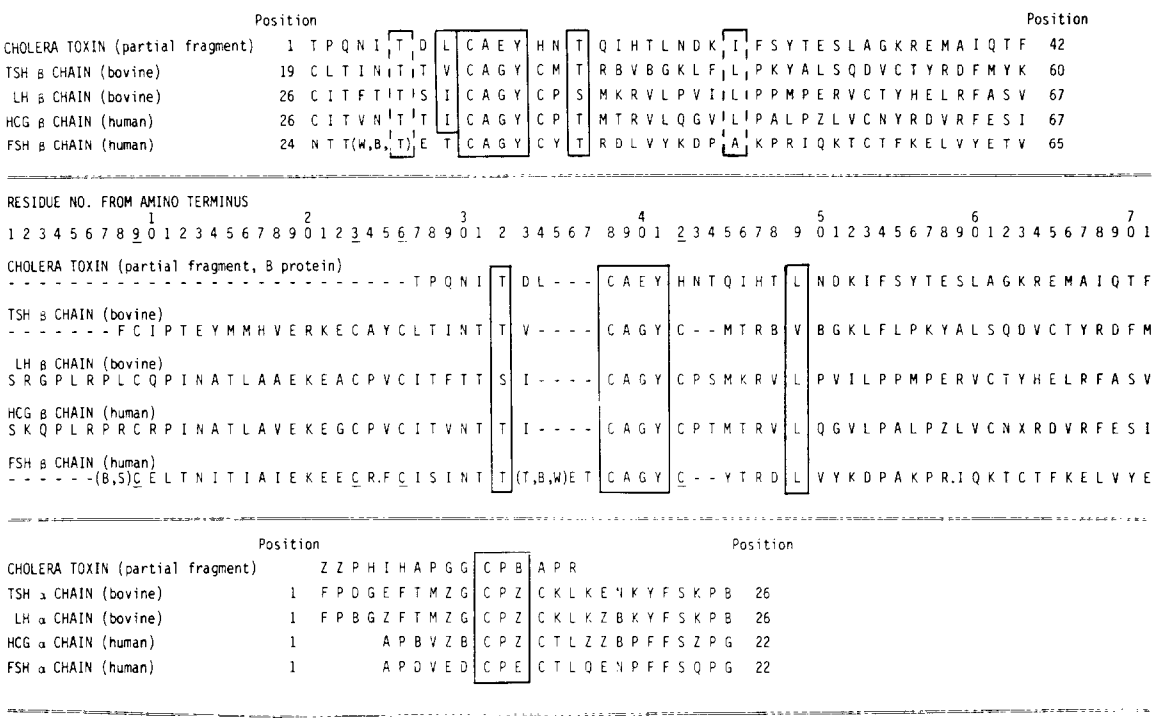

FIG. 1. Top: *Sequences of the amino terminal 41 residues of B chain of cholera toxin and of portions of the β subunits of TSH, LH, and HCG defined by positions from the amino terminal residue of each molecule.* The residue symbols corresponding to the standard residue abbreviations are as follows (15): A, Ala; P, Pro; D, Asp; T, Thr; S, Ser; L, Leu; I, Ile; V, Val; E, Glu; G, Gly; H, His; C, Cys; N, Asn; M, Met; Q, Gln; Z, Glx; B, Asx; R, Arg; F, Phe; Y, Tyr; K, Lys; W, Trp; X, undetermined. Middle: *The sequence of the amino terminal portions of the the B chain of cholera toxin and the β subunits of TSH, LH, HCG, and FSH aligned to maximize homologies by the insertion of gaps resulting from potential deletions or insertions in the course of evolution.* The alignment was generated by the computer program ALIGN (see "Materials and Methods"). The underlined residues show the half-cystine alignment which results from this analysis. Bottom: *The sequences of the amino terminal portions of the α subunits of TSH, LH, HCG, and FSH aligned to show an analogy with a fragment of the A_1 chain of cholera toxin known to be responsible for adenylate cyclase activation (25).*

The program ALIGN similarly identified the significant analogies between the amino-terminal regions of the cholera toxin B chain and the β subunits of TSH, LH, and HCG, while failing to note any particular analogy between cholera toxin and FSH (Table II). Fig. 1 (middle) shows the highest analogy configurations determined by the program ALIGN when interhormonal analogies are also maximized. Various other proteins were similarly compared to the cholera toxin

TABLE I. *Quantitative measure of the similarity between the B chain of cholera toxin and the β subunits of TSH, LH, HCG, and FSH using the fragment comparison program, RELATE, and the sequences noted in Fig. 1*

	(Residues)	Cholera toxin	TSH	LH	HCG	FSH
Cholera toxin . .	(1-42)	n.c. [a]	2.2 [b]	4.4	4.0	0.6
TSH	(19-60)	2.2	n.c.	4.7	6.6	4.5
LH	(26-67)	4.4	4.7	n.c.	13.8	4.4
HCG	(26-67)	4.0	6.6	13.8	n.c.	3.5
FSH	(24-65)	0.6	4.5	4.4	3.5	n.c.

[a] Not compared.

[b] The number of standard deviations of the real score above a random score generated from 100 random runs, using regions of TSH, LH, HCG, and FSH overlapping the known portion of cholera toxin, and measured with fragment lengths of 12 residues in the mutation data matrix.

B chain. These included the α subunits of the glycoprotein hormones, corticotropin, glucagon, insulin, prolactin, growth hormone, and parathyroid hormone. In no case was a significant deviation above the random scores noted.

Since the CAGY region in the β subunit of the TSH-LH-HCG-FSH superfamily represents an important homology which has been highly preserved through the structural and functional differentiation of these hormones, its presence implies the existence of a locus which is critical for function and resistant to successful mutation. Because the sequence analogy exists on that portion of both the TSH and toxin molecules (the B chain or β subunit) which carries the primary determinants for receptor binding (1-5, 22) and because the β subunits of the glycoprotein hormones determine their target organ specificity (23), this region of sequence similarity as well as residues in its immediate spatial proximity may constitute the "active site" concerned with the binding of these molecules to receptors on the cell membrane.

The RELATE and ALIGN programs fail to detect a significant analogy between

TABLE II. *Quantitative measure of the similarity between the B chain of cholera toxin and the β subunits of TSH, LH, HCG, and FSH using the alignment scores for the sequences shown in Fig. 1, middle*

	(Residues)	Cholera toxin	TSH	LH	HCG	FSH
Cholera toxin . .	(1-42)	n.c. [a]	3.7 [b]	3.2	5.0	0.4
TSH	(19-60)	3.7	n.c.	13.7	15.6	10.8
LH	(26-67)	3.2	13.7	n.c.	20.3	12.2
HCG	(26-67)	5.0	15.6	20.3	n.c.	12.2
FSH	(24-65)	0.4	10.8	12.2	12.2	n.c.

[a] Not compared.

[b] The number of standard deviations of the real score above a random score generated from 200 random runs. The mutation data matrix was used with a bias of 1; the gap penalty was sufficiently large to preclude the insertion of gaps additional to those which are shown in Fig. 1, middle, and which were entered manually without penalty.

the β subunit of FSH and the B chain of cholera toxin (Tables I and II) despite the fact that the β subunit of FSH exhibits a highly significant level of homology between itself and the β subunits of the other members of the glycoprotein superfamily (LH-HCG-TSH) (Tables I and II) and despite the fact that the CAGY region is obviously present (Fig. 1, top and middle). This finding and the best alignment derived by the ALIGN program (Fig. 1, middle) may explain why FSH is the only member of the glycoprotein hormone superfamily which is unable to prevent or reverse [^{125}I]TSH binding to TSH receptors (8-12). The best alignment of the β subunit of FSH with its sister molecules requires a four-residue gap in the β subunits of LH, HCG, and TSH just before the CAGY sequence and a three-residue gap in the cholera toxin B chain. Gaps of this size should alter spatial location of the CAGY region with regard not only to other regions of the β subunit but also to the receptor. It is not unreasonable to assume that the consequence of such an alteration might be a loss of competitive binding activity toward receptors specific for TSH, LH,

and HCG, and that the inverse situation will be equally valid.

A comparison of a recently described peptide fragment of the A_1 chain of cholera toxin (14) with the α subunits of TSH, LH, HCG, and FSH suggests that the structure function analogy may extend to these subunits (Fig. 1, bottom). Mutation matrix data analogous to those of Table II yield values between 2 and 5 standard deviations above random for the analogy between the cholera toxin A_1 chain fragment and the noted regions of the α subunits of TSH and LH. The outlined analogy in the α subunits of TSH and LH is presumed to exist in a highly restricted conformational state defined by intrachain disulfide bonds involving half-cystine residues 11, 14, 35, and 36, *i.e.*, these disulfide links cause the formation of a loop with the proline and glutamine residues protruding from the peptide backbone (24). Of special interest in this regard is the evidence indicating that this fragment of the A_1 chain of cholera toxin can enhance the adenylate cyclase activity of pigeon erythrocytes (25).

REFERENCES

1. Holmgren, J., Lönnroth, I., and Svennerholm, L. (1973) *Infect. Immun.* 8, 208-214.
2. King, C. A., and van Heyningen, W. E. (1973) *J. Infect. Dis.* 127, 639-647.
3. Cuatrecasas, P. (1973) *Biochemistry* 12, 3547-3558.
4. Cuatrecasas, P. (1973) *Biochemistry* 12, 3558-3566.
5. Sahyoun, N., and Cuatrecasas, P. (1975) *Proc. Nat. Acad. Sci. USA* 72, 3438-3442.
6. Mullin, B. R., Fishman, P. H., Lee, G., Aloj, S. M., Ledley, F. D., Winand, R. J., Kohn, L. D., and Brady, R. D. (1976) *Proc. Nat. Acad. Sci. USA* 73, in press.
7. Dayhoff, M. O., McLaughlin, P. J., Barker, W. C., and Hunt, L. T. (1975) *Naturwissenschaften* 62, 154-161.
8. Amir, S. M., Carraway, T. F., Jr., Kohn, L. D., and Winand, R. J. (1973) *J. Biol. Chem.* 248, 4092-4100.
9. Tate, R. L., Schwartz, H. I., Holmes, J. M., Kohn, L. D., and Winand, R. J. (1975) *J. Biol. Chem.* 250, 6509-6515.
10. Winand, R. J., and Kohn, L. D. (1972) *Proc. Nat. Acad. Sci. USA* 69, 1711-1715.
11. Bolonkin, D., Tate, R. L., Luber, J. H., Kohn, L. D., and Winand, R. J. (1975) *J. Biol. Chem.* 250, 6516-6521.
12. Mullin, B. R., Lee, G., Ledley, F. D., Winand, R. J., and Kohn, L. D. (1976) *Biochem. Biophys. Res. Commun.*, in press.
13. Kurosky, A., Markel, D., Touchstone, B., and Peterson, J. W. (1975) *Proc. Amer. Soc. Microbiol.* 75, 23.
14. Mendez, E., Lai, C. Y., and Wodnar-Filipowicz, A. (1975) *Biochem. Biophys. Res. Commun.* 67, 1435-1443.
15. Dayhoff, M. O., Hunt, L. T., McLaughlin, P. J., and Barker, W. C. (1972) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), Vol. 5, National Biomedical Research Foundation, Washington, D.C., pp. D1-D334.

16. Dayhoff, M. O., Hunt, L. T., McLaughlin, P. J., and Barker, W. C. (1973) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), Vol. 5, Suppl. 1, National Biomedical Research Foundation, Washington, D.C., pp. S9-S97.
17. Barker, W. C., and Dayhoff, M. O. (1972) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), Vol. 5, National Biomedical Research Foundation, Washington, D.C., pp. 101-110.
18. Dayhoff, M. O., Hunt, L. T., Barker, W. C., and Schwartz, R. S. (1976) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), Vol. 5, Suppl. 2,National Biomedical Research Foundation, Washington, D.C., in press.
19. Hunt, L. T., Barker, W. C., and Dayhoff, M. O. (1974) *Biochem. Biophys. Res. Commun.* 60, 1020-1028.
20. Dayhoff, M. O., Hunt, L. T., Barker, W. C., McLaughlin, P. J., and Shiu, M. R. (1973) in *Protein Segment Dictionary* (Dayhoff, M. O., ed.), National Biomedical Research Foundation, Washington, D.C.
21. Needleman, S. B., and Wunsch, C. D. (1970) *J. Mol. Biol.* 48, 443-453.
22. Wolff, J., Winand, R. J., and Kohn, L. D. (1974) *Proc. Nat. Acad. Sci. USA* 71, 3460-3464.
23. Liao, T.-H., and Pierce, J. G. (1970) *J. Biol. Chem.* 245, 3275-3281.
24. Cornell, J. S., and Pierce, J. G. (1974) *J. Biol. Chem.* 249, 4166-4174.
25. Gill, D. M., and King, C. A. (1976) *J. Biol. Chem.* 251, in press.