Complete Amino Acid Sequences of Bovine and Human Endozepines

HOMOLOGY WITH RAT DIAZEPAM BINDING INHIBITOR

(Received for publication, January 21, 1986)

Hans Marquardt, George J. Todaro, and Mohammed Shoyab
From the Department of Biochemistry, Oncogen, Seattle, Washington 98121

The complete amino acid sequences of bovine and human endozepines have been determined. The amino-terminal serine of both endozepines is acylated. Assignment of the first 7 residues was achieved through Edman degradation after acid-induced rearrangement and subsequent acid hydrolysis of the amino-terminal blocking group. Cleavage of endozepine by chemical and enzymatic techniques established all the fragments in an unambiguous sequence. Bovine and human endozepines are single-chain polypeptides of 86 residues, with calculated molecular weights of 9913, displaying 93% homology. A comparison between the sequences of bovine and human endozepines with the partial sequences of the functionally related diazepam binding inhibitors from rat brain reveals significant sequence homology. The reported results suggest that bovine and human endozepines as well as rat diazepam binding inhibitor belong to a new family of polypeptides which presumably take part in the modulation of \( \gamma \)-aminobutyric acid-ergic transmission.

Benzodiazepines are widely used to affect behavioral states ranging from anxiety to sedation. The pharmacological effect of these drugs is mediated through \( \gamma \)-aminobutyric acid-ergic transmission systems (1, 2). It is generally accepted that \( \gamma \)-aminobutyric acid acts by increasing chloride permeability in postsynaptic membranes (1, 2). Several polypeptides with similar properties have been purified from rat (1), human (2), and bovine brain, based on their capacity to competitively inhibit diazepam binding to its specific brain recognition sites on synaptosomes. In addition, these polypeptides were shown to elicit behavioral effects expected for a diazepam antagonist (1, 2).

Partial amino acid sequences have been published for rat diazepam binding inhibitor (1, 3). The complete amino acid sequences of bovine and human endozepines are described in this report. The data clearly show an extensive structural homology between bovine and human endozepines and rat diazepam binding inhibitor. The findings described here might permit direct examination of its proposed role as a precursor of small signal-transducing peptides (3).

EXPERIMENTAL PROCEDURES AND RESULTS AND DISCUSSION

The amino acid sequence of bovine endozepine (bEP), as presented in Fig. 1, was determined by automated Edman degradation of selected fragments using standard enzymatic and chemical cleavage techniques. The separation of complex peptide mixtures over a broad range of fragment sizes was achieved through a single rPHPLC chromatography step. The use of trifluoroacetic acid in both solvents appeared to maintain the solubility of even large fragments at high acetonitrile concentrations. In addition, the buffer system used is completely volatile, which enabled subsequent chemical characterization of separated peptides.

The amino-terminal serine residue of bEP is acylated. Acid-induced rearrangement and subsequent acid hydrolysis of the \( O \)-acyl derivative of the amino-terminal heptapeptide were used to establish an unambiguous assignment of the first 7 residues. The nature of the amino-terminal blocking group was not identified. The hydrophobic nature of the blocking group in EP was, however, recognized by rpHPLC. The observed retention time of endoproteinase Lys-C peptide 1-7 was slightly increased when compared with the retention time predicted by the sequence of the unmodified peptide (4), suggesting an \( N \)-acetyl blocking group contributing to the increased hydrophobicity.

The sequence of the amino-terminal heptapeptide and Edman degradation of selected fragments derived from bEP through cleavage with the endoproteinase Lys-C and Staphylococcus aureus V8 protease accounted for all the residues in the amino acid composition of the protein. Overlap information between glutamic acid residue 67 and aspartic acid residue 68 was provided by the sequence of chymotryptic peptide 62-73. Additional overlap information between glutamic acid residue 79 and leucine 80 was provided by the sequences of chymotryptic peptide 74-80 and cyanogen bromide peptide 71-86 which established all the fragments in an unambiguous sequence of 86 residues. The amino acid composition calculated from the complete sequence is in agreement with the compo-

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 M. Shoyab, L. E. Gentry, and G. J. Todaro, manuscript in preparation.

2 Portions of this paper (including "Experimental Procedures," part of "Results," Figs. S1-S4, and Tables S1 and SII) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

The abbreviations used are: bEP, bovine and human endozepines, respectively; EP, endozepine; DBI, diazepam binding inhibitor; rDN, rat diazepam binding inhibitor; rPHPLC, reversed-phase high performance liquid chromatography.
Fig. 1. Amino acid sequence of bovine endozepine and a schematic outline of the data supporting the sequence. Peptides obtained by cleavage with endoproteinase Lys-C (K), with S. aureus V8 protease (E), with chymotrypsin (Ch), and at methionyl residues with CNBr (CB) are indicated. Residues identified by Edman degradation are underlined. Data supporting the sequence are given in the Miniprint Section.

Fig. 2. Amino acid sequence of human endozepine and a schematic outline of the data supporting the sequence. Endoproteinase Lys-C peptides are indicated by K, and S. aureus V8 protease peptides are indicated by E. Residues identified by Edman degradation are underlined. Data supporting the sequence are given in the Miniprint Section.

Table I

| Amino acid | hEP* | hEP* |
|------------|------|------|
|            | residues/mol | residues/mol |
| Aspartic acid | 8.8 (7) | 7.1 (5) |
| Asparagine  | (2)   | (3)   |
| Glutamic acid | 12.3 (10) | 12.8 (11) |
| Glutamine   | (2)   | (2)   |
| Serine      | 2.8 (3) | 3.3 (3) |
| Glycine     | 5.1 (5) | 7.2 (6) |
| Histidine   | 1.8 (2) | 2.1 (2) |
| Arginine    | 1.4 (1) | 2.4 (2) |
| Threonine   | 3.8 (4) | 5.1 (5) |
| Alanine     | 8.6 (9) | 8.0 (8) |
| Proline     | 2.2 (2) | 2.3 (2) |
| Tyrosine    | 3.4 (4) | 4.1 (4) |
| Valine      | 3.0 (3) | 2.6 (3) |
| Methionine  | 2.7 (3) | 2.7 (3) |
| Half-cystine| 0 (0)  | 0 (0)  |
| Isoleucine  | 3.7 (4) | 3.4 (4) |
| Leucine     | 5.3 (5) | 4.7 (5) |
| Phenylalanine | 2.9 (3) | 2.7 (3) |
| Lysine      | 14.3 (15) | 12.0 (13) |
| Tryptophan  | ND (2) | ND (2) |
| Total       | 86    | 86    |
| Calculated M* | 9913 | 9913 |

* Values are based on two samples hydrolyzed for 18 h at 105 °C and calculated by assuming M* = 10,000. A value of zero indicates less than 0.2 residues. Values in parentheses indicate the number of residues per peptide, determined by amino acid sequence analysis.

Values per peptide.
ND, not determined.

sition determined by hydrolysis, as shown in Table I. Bovine EP is a single-chain polypeptide of 86 residues, with a calculated molecular weight of 9,913, consistent with the apparent molecular weight of 10,000 determined by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate and urea.1

The amino acid sequence of hEP is presented in Fig. 2. Human EP, like its bovine counterpart, is amino-terminally blocked, possibly through N-acetylation at serine. Treatment of endoproteinase Lys-C peptide 1–7 with hydrochloric acid partially deblocked the amino terminus and Edman degradation established the first 7 residues. No overlap information between lysine 32 and glutamine 33, glutamic acid 67 and aspartic acid 68, glutamic acid 79 and leucine 80 was provided by the experiments described. However, the amino acid composition of hEP, as presented in Table I, is in agreement with the proposed structure. In addition, the extensive sequence homology between hEP and hEP lends credence to the proposed structure. The calculated molecular weight of 9913 for hEP from the amino acid sequence is in agreement with the determined apparent molecular weight.1 Characterization of complementary DNA clones encoding both hEP and hEP confirmed the proposed amino acid sequences. In the nucleotide sequences of hEP and hEP of the cloned complementary DNAs, initiation of the open reading frame at methionine immediately precedes the amino-terminal serine and extends

* N. R. Webb, T. M. Rose, N. Malik, H. Marquardt, M. Shoyab, G. J. Todaro, and D. C. Lee, manuscript in preparation.
A comparison of the amino acid sequences of bEP and hEP is presented in Fig. 3. The data indicate that bEP and hEP differ from each other by only 6 conservative amino acid substitutions: 80 out of 86 residues, or 93%, are identical. Each of these amino acid substitutions is compatible with a single base change at the DNA level. The amino acid sequences of bEP and hEP are homologous to the recently published partial sequences of DBI isolated from rat brain (1, 3). Residues in the amino acid sequence of hEP which differ from bEP are underlined. Residues in the amino acid sequence of rDBI which differ from hEP are printed in boldface.

for 86 amino acids to the carboxyl-terminal isoleucine, followed by a stop codon.

The amino acid sequences of bEP and hEP are shown in Fig. 3. The data indicate that bEP and hEP differ from each other by only 6 conservative amino acid substitutions: 80 out of 86 residues, or 93%, are identical. Each of these amino acid substitutions is compatible with a single base change at the DNA level. The amino acid sequences of bEP and hEP are homologous to the recently published partial sequences of DBI isolated from rat brain (1, 3), as shown in Fig. 3. Forty out of 50 residues, or 80%, are identical when the sequences of bEP and rDBI are compared. A comparison of the amino acid sequence of hEP with the partial sequences of rDBI shows 38 identities out of 50 comparisons, or 76% sequence homology. The authors, however, have concluded (2) that rDBI contains 104 amino acid residues, a carboxyl-terminal tyrosine, and two identical copies of an octadecapeptide, comprising tryptic peptide 33–50.

Rat DBI has been shown to act as an inverse antagonist of benzodiazepines in vitro and in vivo (1). Benzodiazepines are thought to act by modulating γ-aminobutyric acid-ergic transmission. The completion of the primary structure of EP might permit direct examination of its proposed role as a precursor of small signal-transducing peptides (3). Isolation of complementary DNAs encoding the bovine and human forms of the proposed endogenous ligand should allow studies of its biosynthesis and the molecular mechanisms of its actions.

Acknowledgments—We thank Mario Lioubin for excellent technical assistance, Lowell Ericsson for valuable suggestions, and Bonnie Kirk for typing this manuscript.

REFERENCES
1. Guidotti, A., Forchetti, C. M., Corda, M. G., Konkel, D., Bennett, C. D., and Costa, E. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 3531–3535
2. Alfred, H., Costa, E., Ferrero, P., Fujimoto, M., Cozenz-Murphy, D., and Guidotti, A. (1985) Science 229, 179–182
3. Ferrero, P., Guidotti, A., Conti-Tronconi, B., and Costa, E. (1984) Neuropharmacology 23, 1359–1362
4. Walsh, K. A., and Sasagawa, T. (1984) Methods Enzymol. 106A, 225–230
5. IUPAC-IUB Commission on Biochemical Nomenclature (1968) J. Biol. Chem. 243, 3557–3559
6. Davoren, P. R. (1962) Biochim. Biophys. Acta 63, 150–153
7. Massotti, M., Guidotti, A., and Costa, E. (1981) J. Neurosci. 1, 409–418
8. Gens, E., and Wickop, B. (1969) J. Biol. Chem. 237, 1856–1860
9. Iwai, K., and Ando, T. (1967) Methods Enzymol. 11, 263–282
10. Cohen, S. A., Tarvin, T. L., and Bidlingmeyer, B. A. (1984) Am. Lab. (Fairfield Conn.) 16, 48–59
11. Hunkapiller, M. W., and Hood, L. E. (1983) Science 219, 650–659
12. Ozols, J., and Gerard, C. (1977) J. Biol. Chem. 252, 5986–5998
RESULTS

Enzymatic Cleavage of Endorepines and Peptide Isolation

Bovine and human EP were digested with the endoproteinase Lys-C. The enzyme cleaves primarily at the carboxyterminal side of lysine residues. The digests were acidified to pH 2.0 with trifluoroacetic acid and the peptides separated by rpHPLC. Representative chromatograms are shown in Fig. S1-A and B. Peptide-containing fractions were pooled. Aliquots were taken for sequence analysis and for amino acid analyses when necessary (Table S1 and SII). No sequences of K 1-7 of hEP or bEP were obtained. The complete sequence of the remaining listed peptides were determined. Peptides K 83-86 were the only peptides not containing a carboxyterminal lysine and were thus assumed to be the carboxyterminal peptides of hEP and bEP.

For cleavage at dicarboxylic acid residues, hEP and bEP were digested with V8 protease. The complete amide sequence of the endopeptidase Lys-C, yielding three major peptides resolvable by rpHPLC. Representative chromatograms are shown in Fig. S2-A and B. Peptides listed in Fig. S2-A and B were subjected to automated Edman degradation and were sequenced in their entirety. Peptides K 83-86 of hEP and bEP were the only peptides not containing a carboxyterminal glutamic or aspartic acid residue and were thus assumed to be the carboxyterminal peptides of hEP and bEP.

Bovine EP was digested with CDI chymotrypsin and the acidified digest separated by rpHPLC. Peptides listed in Fig. S3 were subjected to Edman degradation. Cleavage at carboxy terminal peptide  of phenylalanine, tryptophan, tyrosine, and leucine bonds was obtained.

Chemical Cleavage of Endorepines and Peptide Isolation

Gel permeation chromatography of the CNBr fragments of hEP on a Bio-Sil TSK-250 column and subsequent rpHPLC on a Bondapak C18 column resolved the peptides CB 1-24, CB 25-46, and CB 71-86. A representative chromatogram is shown in Fig. S1. CB 47-70 was not recovered from the Bio-Sil TSK-250 column. The presence of two tryptophan residues (CB 47-70) suggests that cleavage at these bonds with excess CNBr may have yielded smaller peptide fragments (12) which eluted on gel permeation chromatography in the total column volume together with UV-absorbing material derived from CNBr. The amine terminus of CB 1-24 was blocked, whereas peptides CB 25-46 and CB 71-86 had a free amino terminus and were subjected to Edman degradation. The sequences of CB 25-46 and CB 71-86 provided additional overlap information (Table S1).

The carboxyterminal CNBr peptide of hEP was further subfragmented with the endoproteinase Lys-C, yielding three major peptides resolvable by rpHPLC. 1350 pmol of the carboxyterminal blocked peptide K 1-7 were partially deblocked with 12 N HCl and the sequence of amid-treated K 1-7 determined. Automated Edman degradation was performed with 190 pmol of deblocked K 1-7 (based on the yield of identified Ala-3) (Table SII). 700 pmol of the carboxyterminal blocked peptide K 1-7 of hEP were partially deblocked with 12 N HCl in a similar manner and applied to the sequence without further purification. The complete sequence of the deblocked peptide K 1-7 of hEP (85 pmol, based on the yield of identified Ala-3) was obtained (Table SII).

Fig. S1. Reversed phase high performance liquid chromatography of endopeptidase Lys-C peptides of endorepine. Chromatography on a Bondapak C18 column (10 µm, 0.39 x 30 cm). (A) Elution pattern of 0.7 nmol of bovine endorepine digested with endopeptidase Lys-C. (B) Elution pattern of 1.4 nmol of human endorepine digested with endopeptidase Lys-C. The elution of peptides was achieved with a 2-6% linear gradient of acetonitrile in water to 60% acetonitrile containing 0.05% trifluoroacetic acid at a flow rate of 1 ml/min at room temperature. UV-absorbing material was monitored at 214 nm (---); the dashed lines (----) gives the concentration of acetonitrile. Peaks designated by a K refer to endopeptidase Lys-C peptides subjected to Edman degradation; numbers refer to the positions of that particular fragment in the complete amino acid sequence.

Fig. S2. Reversed phase high performance liquid chromatography of Staphylococcus V8 protease peptides of endorepine. Chromatography on a Bondapak C18 column. (A) Elution pattern of 0.6 nmol of bovine endorepine digested with V8 protease. (B) Elution pattern of 0.8 mol of human endorepine digested with V8 protease. The chromatography conditions are described in the legend to Fig. S1. Peaks designated by an E refer to V8 protease peptides subjected to Edman degradation.

Fig. S3. Reversed phase high performance liquid chromatography of chymotryptic peptides of bovine endorepine. Chromatography on a Bondapak C18 column. (A) Elution pattern of 0.2 nmol of bovine endorepine digested with CDI chymotrypsin. The chromatography conditions are described in the legend to Fig. S1. Peaks designated by a C refer to chymotryptic peptides subjected to Edman degradation.

Fig. S4. Reversed phase high performance liquid chromatography of cyanogen bromide peptides of bovine endorepine. Two nmol of bovine endorepine cleaved with CNBr were directly applied to a Bio-Sil TSK-250 column (0.75 x 60 cm), equilibrated with 0.15 trifluoroacetic acid containing 40% acetonitrile at a flow rate of 0.25 ml/min. The three longest CNBr fragments from bovine endorepine were recovered, prepared for injection, and chromatographed on a Bondapak C18 column. The chromatography conditions are described in the legend to Fig. S1. Peaks designated by a C refer to CNBr peptides subjected to Edman degradation.
### Table SI. Amino acid sequence data for bovine endorphine

| Position | Residue  | Peptide, (cycle), Yield (pol) |
|----------|----------|-------------------------------|
| 1        | Ser (17) | 60                            |
| 2        | Asp (3)  | 120                           |
| 3        | Glu (4)  | 140                           |
| 4        | Phe (5)  | 160                           |
| 5        | Ala (6)  | 200                           |
| 6        | Lys (7)  | 230                           |
| 7        | Ala (8)  | 260                           |
| 8        | Lys (9)  | 290                           |
| 9        | Val (10) | 320                           |
| 10       | Gly (11) | 350                           |
| 11       | Gly (12) | 380                           |
| 12       | Leu (13) | 410                           |
| 13       | Ser (14) | 440                           |
| 14       | Lys (15) | 470                           |
| 15       | Thr (16) | 500                           |
| 16       | Val (17) | 530                           |
| 17       | Lys (18) | 560                           |
| 18       | Ser (19) | 590                           |
| 19       | Glu (20) | 620                           |
| 20       | Ala (21) | 650                           |
| 21       | Lys (22) | 680                           |
| 22       | Thr (23) | 710                           |
| 23       | Leu (24) | 740                           |
| 24       | Val (25) | 770                           |
| 25       | Lys (26) | 800                           |
| 26       | Thr (27) | 830                           |
| 27       | Ala (28) | 860                           |
| 28       | Lys (29) | 890                           |
| 29       | Ser (30) | 920                           |
| 30       | Thr (31) | 950                           |
| 31       | Lys (32) | 980                           |
| 32       | Thr (33) | 1010                          |

### Table SII. Amino acid sequence data for human endorphine

| Position | Residue  | Peptide, (cycle), Yield (pol) |
|----------|----------|-------------------------------|
| 1        | Ser (17) | 60                            |
| 2        | Asp (3)  | 120                           |
| 3        | Glu (4)  | 140                           |
| 4        | Phe (5)  | 160                           |
| 5        | Ala (6)  | 200                           |
| 6        | Lys (7)  | 230                           |
| 7        | Ala (8)  | 260                           |
| 8        | Lys (9)  | 290                           |
| 9        | Val (10) | 320                           |
| 10       | Gly (11) | 350                           |
| 11       | Gly (12) | 380                           |
| 12       | Leu (13) | 410                           |
| 13       | Ser (14) | 440                           |
| 14       | Lys (15) | 470                           |
| 15       | Thr (16) | 500                           |
| 16       | Val (17) | 530                           |
| 17       | Lys (18) | 560                           |
| 18       | Ser (19) | 590                           |
| 19       | Glu (20) | 620                           |
| 20       | Ala (21) | 650                           |
| 21       | Lys (22) | 680                           |
| 22       | Thr (23) | 710                           |
| 23       | Leu (24) | 740                           |
| 24       | Val (25) | 770                           |
| 25       | Lys (26) | 800                           |
| 26       | Thr (27) | 830                           |
| 27       | Ala (28) | 860                           |
| 28       | Lys (29) | 890                           |
| 29       | Ser (30) | 920                           |
| 30       | Thr (31) | 950                           |
| 31       | Lys (32) | 980                           |
| 32       | Thr (33) | 1010                          |

### Additional Notes
- The tables provide amino acid sequences for both bovine and human endorphine peptides. Each position lists the amino acid followed by its cycle number and yield in milligrams (pol).