THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 262, No. 12, Issue of April 25, pp. 5499-5509, 1987
Printed in U.S.A.

The Primary Structure of Rabbit Liver Cytosolic Serine Hydroxymethyltransferase*

(Received for publication, November 10, 1986)

Filippo Martini‡, Sebastiana Angelaccio‡, Stefano Pascarella‡, Donatella Barra‡, Francesco Bossa‡, and Verne Schirch§¶

From the ‡Dipartimento di Scienze Biochimiche and Centro di Biologia Molecolare del Consiglio Nazionale delle Ricerche, Università La Sapienza, 00185 Roma, Italy and the §Department of Biochemistry, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298

The complete amino acid sequence of cytosolic serine hydroxymethyltransferase from rabbit liver was determined. The sequence was determined from analysis of peptides isolated from tryptic and cyanogen bromide cleavages of the enzyme. Special procedures were used to isolate and sequence the C-terminal and blocked N-terminal peptides. Each of the four identical subunits of the enzyme consists of 483 residues. The sequence could be easily aligned with the sequence of *Escherichia coli* serine hydroxymethyltransferase. The primary structural homology between the rabbit and *E. coli* enzymes is about 42%. The importance of the primary and predicted secondary structural homology between the two enzymes is discussed.

Serine hydroxymethyltransferase catalyzes the conversion of serine and tetrahydrofolate to glycine and 5,10-methylenetetrahydrofolate. This reaction serves as the primary source of one-carbon groups for biosynthetic reactions in the cell (1). The enzyme activity is widely distributed in nature, being found in prokaryotic, eukaryotic, and plant cells (2–6). The enzyme has been purified from several mammalian livers, plants and *Escherichia coli* (2–6). Recently the enzyme has been purified and crystallized from the obligate methylotroph, *Hypomicrobium methylovorum* (7). In addition to using the coenzyme tetrahydrofolate, the purified enzyme also contains pyridoxal-P. Extensive studies have been done on the mechanism of the enzyme with respect to the structure and rates of interconversion of intermediates (1). These studies suggest that the mechanism of the liver enzyme is very similar to the mechanism of the *E. coli* enzyme (2). The liver of mammals contains both cytosolic and mitochondrial forms of the enzyme (8). Several studies have been done on the plant enzyme but at this time few mechanistic or structural results permit comparison to the mammalian enzyme (5).

The *E. coli* gene coding for serine hydroxymethyltransferase has been cloned and the structural gene sequenced (9, 10). From the DNA sequence, an amino acid sequence for the *E. coli* enzyme has been proposed (10). This proposed sequence has been verified by determining the N-terminal, C-terminal, and active site sequences of the enzyme by classical protein sequencing methods (11). Cysteine-containing peptides and the pyridoxal-P-containing peptide of rabbit liver cytosolic and mitochondrial serine hydroxymethyltransferase have also been sequenced (12–14). These studies suggested that there is considerable sequence homology between the prokaryotic and eukaryotic forms of the enzyme (2).

In this paper we report the complete amino acid sequence of rabbit liver cytosolic serine hydroxymethyltransferase which has been determined by classical protein sequencing methods. This is the first reported sequence of the enzyme from a eukaryotic source. It confirms that there is considerable sequence homology with the *E. coli* enzyme. Serine hydroxymethyltransferase contains sites for binding the amino acid substrate and the two coenzymes. The sequence of a peptide involved in binding pyridoxal-P has been shown previously to be conserved between eukaryotic and prokaryotic forms of the enzyme (2). However, little is known about those amino acid sequences in the enzyme which are involved in binding of the substrate amino acid and tetrahydrofolate. Important catalytic residues are also expected to be conserved in these sites since the mechanisms of the enzymes from eukaryotic and prokaryotic sources seem to be very similar. We report the methods for determining the sequence and discuss the structural features which may be related to the function of this enzyme.

EXPERIMENTAL PROCEDURES AND RESULTS[1]

DISCUSSION

The complete amino acid sequence of rabbit liver cytosolic serine hydroxymethyltransferase is given in Fig. 1. The protein is made up of four identical subunits each containing 483 residues; the subunit molecular weight of the apoenzyme is calculated to be 52,894, which is in good agreement with that reported previously (15). The amino acid composition deduced from the sequence agrees with that obtained following acid hydrolysis (Table I). The sequence was deduced following the isolation and characterization of a complete set of tryptic peptides which were ordered by overlapping peptides obtained

* This work was supported in part by a grant from the Ministero Pubblica Istruzione (to F. B.) and by Grant GM 28143 from the National Institutes of Health (to V. S.). Part of this work will be submitted by S. P. in partial fulfillment of the requirements of the degree of Ph.D at the Università La Sapienza, Roma. 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.

¶ To whom correspondence should be addressed.

[1] Portions of this paper (including "Experimental Procedures," "Results," Tables II–XIV, and Figs. 3–7) are presented in miniprint at the end of this paper. The abbreviations used are: cmCys, *S*-carboxymethylcysteine; Hse, homoserine; Hsl, homoserine lactone. Miniprint is easily read with the end of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 86 M-3874, cite the authors, and include a check or money order for $14.00 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

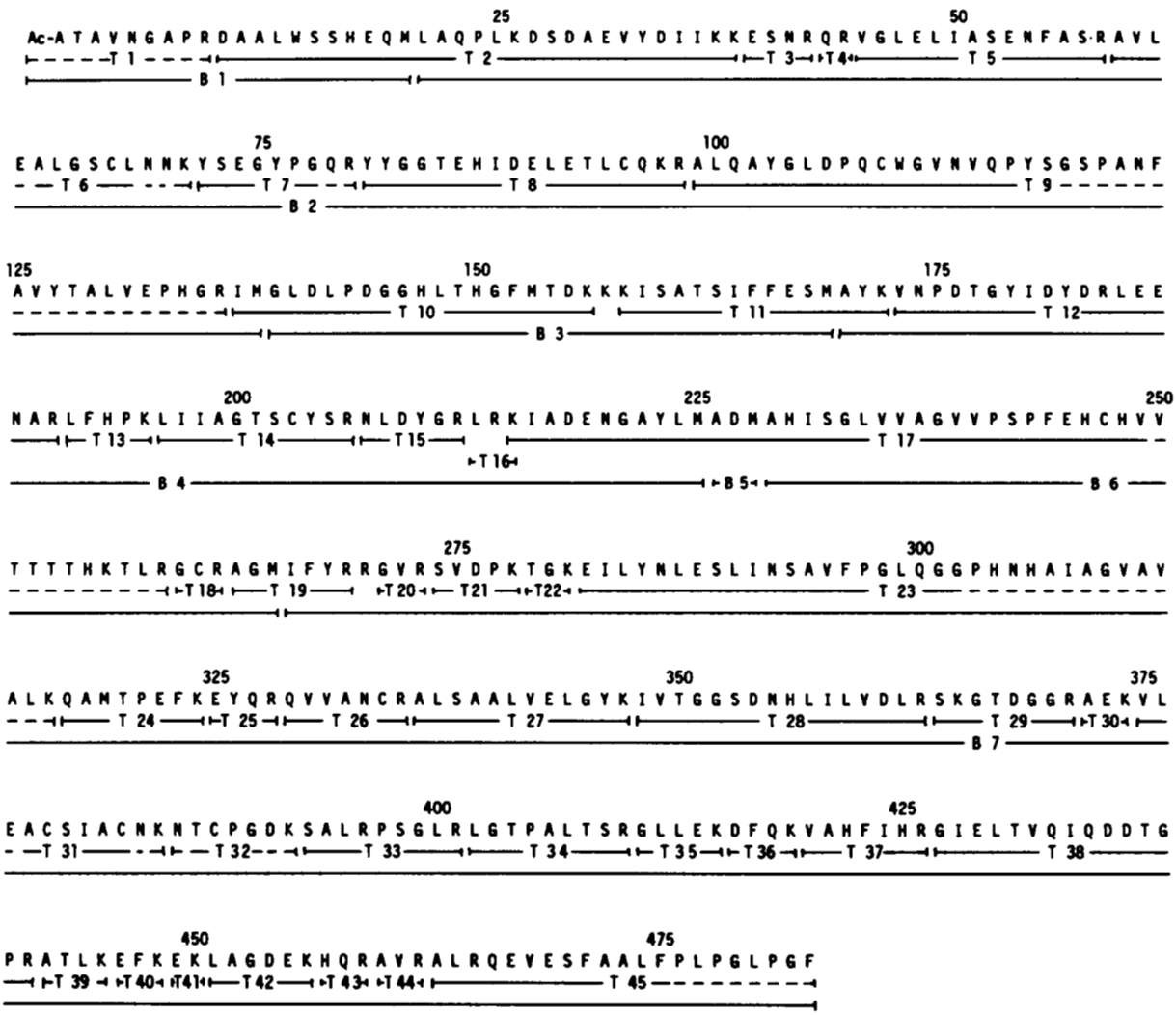


FIG. 1. **Complete amino acid sequence of cytosolic serine hydroxymethyltransferase from rabbit liver.** ——, extent of the various fragments used to construct the sequence; – – –, sequences inferred from amino acid compositions. *T*, tryptic peptides; *B*, cyanogen bromide peptides.

following cyanogen bromide cleavage.

The extensive use of high-performance liquid chromatographic procedures on macroporous reverse-phase columns greatly facilitated the work necessary for peptide purification as compared with our initial efforts based upon the use of ion-exchange columns and paper or thin layer separation techniques. Similarly, in the final part of the work, the availability of a gas-phase automatic sequencer was of great help mainly for confirmation of the structure in regions of the protein whose sequence was deduced only on the basis of relatively weak evidence from juxtaposed subfragments. Of great help in establishing the C-terminal sequence of the protein were the results obtained after application of a procedure aimed at the selective isolation of a covalently modified version of a C-terminal fragment. Moreover, the proposed C-terminal sequence agrees with the results of carboxypeptidase digestion of the intact protein which identified phenylalanine as the C terminus (18). The N-terminal residue of the protein is blocked, and the pertinent tryptic peptide was isolated after application of a selective procedure. Analysis of this peptide solved the N-terminal sequence of the protein, and the blocking group was established as acetyl.

The amino acid sequence of rabbit liver cytosolic serine hydroxymethyltransferase is compared in Fig. 2 with that of the protein coded by the *glyA* gene of *E. coli*, which was identified as the bacterial version of this enzyme (10). This identification was substantiated by our preliminary sequence data on both cytosolic and mitochondrial isoenzymes (18), by the isolation of the bacterial enzyme, by the study of some of its functional and structural properties (11), and now by the complete sequence data of the cytosolic isoenzyme as shown in Fig. 2. In order to optimize the homology, it was necessary to insert a number of gaps into the sequences. These gaps account for the larger size of the mammalian protein (483 residues) with respect to the bacterial one (417 residues). The 66 extra residues in the rabbit enzyme appear to be mainly clustered in a relatively restricted number of definite segments. The extent of homology among the two proteins is 42.0%. This is of the same order of that observed in another case of pyridoxal-P-dependent enzyme, *i.e.* aspartate aminotransferase, where the homology among the *E. coli* (19) and a corresponding mammalian isoenzyme, that from pig heart (20), is 38.8%. However, when the comparison of the sequence of serine hydroxymethyltransferase is restricted to the aligned segments, after exclusion from the calculation of the insertions included to maximize the alignment, the extent of homology among the eukaryotic and prokaryotic enzymes increases to 49.6%, whereas in the case of aspartate aminotransferase enzymes, which differ in length by only 16 residues, homology remains practically invaried (40.4%). This suggests

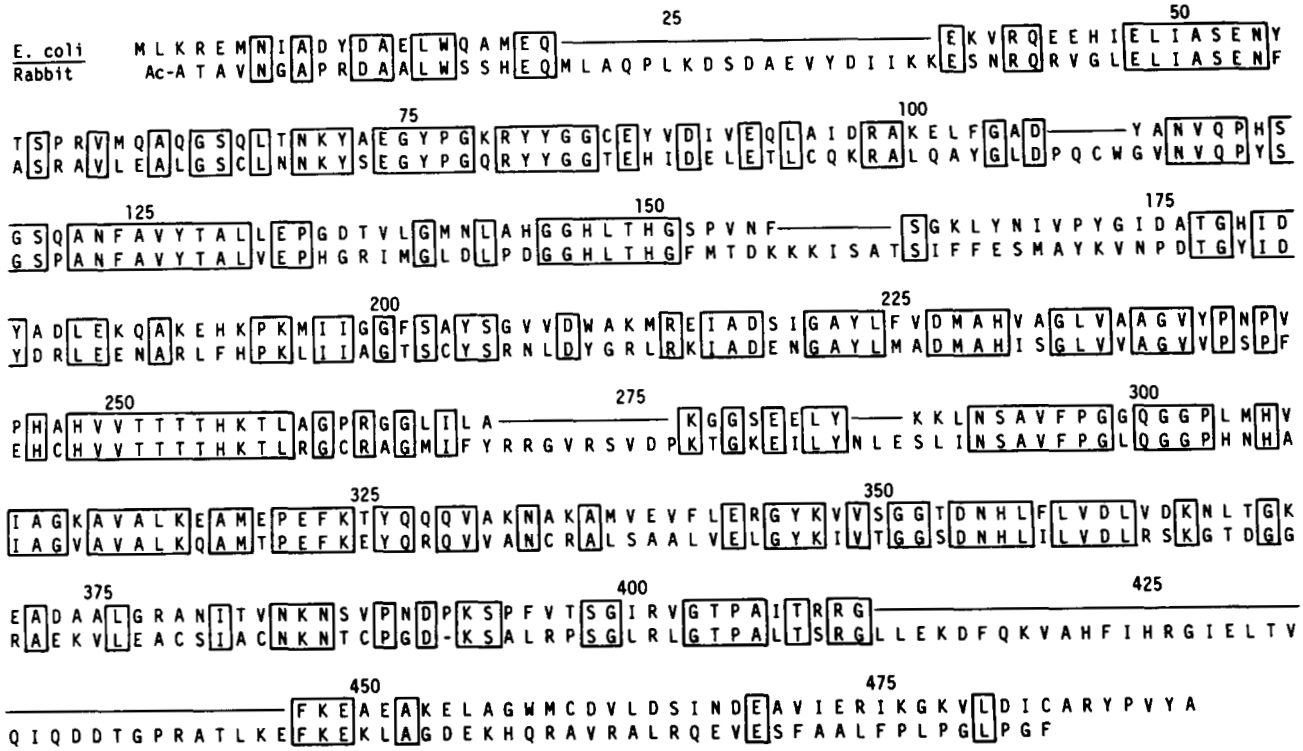


FIG. 2. **Comparison of the amino acid sequences of *E. coli* and rabbit liver cytosolic serine hydroxymethyltransferases.** Gaps (——) in the sequences have been introduced to maximize the homology. *Boxes* indicate positions at which residues are identical.

TABLE I

Amino acid composition of cytosolic serine hydroxymethyltransferase from rabbit liver

Acid hydrolyses were performed on unmodified serine hydroxymethyltransferase for 24, 48, and 72 h. The values of threonine and serine were obtained by extrapolation to zero time of hydrolysis. Values of valine and isoleucine were from 72-h hydrolysates.

Amino acid	Amino acid analysis	Sequence
	residues/subunit	
Aspartic acid	40.6	42[a]
Threonine	23.7	24
Serine	27.7	28
Glutamic acid	47.7	47[b]
Proline	24.3	23
Glycine	43.8	44
Alanine	46.6	48
Half-cystine	10.6[c]	10
Valine	28.9	30
Methionine	7.4	8
Isoleucine	21.8	22
Leucine	47.6	49
Tyrosine	17.2	18
Phenylalanine	15.7	16
Histidine	15.2	16
Lysine	25.7	27
Arginine	28.5	29
Tryptophan	1.8[d]	2

[a] 24 aspartic acid and 18 asparagine residues.
[b] 30 glutamic acid and 17 glutamine residues.
[c] Determined as cysteic acid after hydrolysis in the presence of dimethyl sulfoxide (16).
[d] Determined after hydrolysis with 4 N methanesulfonic acid (17).

more extensive requirements for conservation of catalytically and/or conformationally relevant residues in the case of serine hydroxymethyltransferase. Moreover, inspection of Fig. 2 shows that the distribution of conserved and substituted regions is clearly not uniform throughout the two sequences. In particular, homology is very high in central regions of the two proteins, such as that adjacent to lysine 256 which binds pyridoxal-P, and is virtually absent in the C-terminal portion. Of particular interest for considerations on the extent of predictable similarity of the three-dimensional folding of the two proteins is the conservation of residues such as glycyl and prolyl, which may play unique roles in determining specific chain packing and protein conformations. Out of the 44 glycyls and 23 prolyls present in the eukaryotic enzyme, 29 and 11, respectively, occupy the same position in the *E. coli* protein. Prediction of secondary structure by the method of Garnier *et al.* (21) shows (Fig. 6) that the two enzymes are 49.1% homologous by this criterion (58.0% after exclusion from the comparison of the inserted segments). Moreover, the calculated hydrophilicity profile of the inserted segments (Fig. 7) is compatible with their location on the surface of the protein molecule. This suggests that the inserted segments should cause no large deviance from a catalytically competent three-dimensional folding which should be very similar for both the rabbit and *E. coli* enzymes, as expected on the basis of our mechanistic studies (11).

Recognizing the similarity between the eukaryotic and prokaryotic forms of serine hydroxymethyltransferase is important in continuing our studies on the mechanism and function of this enzyme. We have recently changed an active site histidine to an asparagine in the *E. coli* enzyme (2). The importance of this histidine was suggested by the conservation of an 11-residue amino acid sequence between the rabbit cytosolic and mitochondrial isozymes and the *E. coli* enzyme. This histidine residue (position 255) was known to be at the active site because of it being adjacent to the lysyl residue which binds pyridoxal-P. Because it is relatively easy to change amino acid residues by site-directed mutagenesis, we will now be able to look at other sequences where there is strong homology as possible sites of important amino acid residues involved directly in the mechanism of this enzyme.

An example is the sequence Arg-Tyr-Tyr-Gly-Gly (position 80–84) which is present in all three enzyme forms. We are also interested in locating the polyglutamate binding domain of tetrahydropteroyl polyglutamate. We assume that a series of positive charges on the enzyme will interact with the negatively charged γ-carboxyl residues of the polyglutamate portion of the coenzyme. In the *E. coli* enzyme we have located 4 lysyl residues which occur in a region (187–195) predicted to be an α-helix. Model building studies show that the 4 lysyl residues are on the same side of the α-helix and are separated by about the same distance as the carboxyl residues of the tetrahydropteroyl triglutamate. Three of the four positive charges are conserved in the rabbit enzyme sequence. We plan to change these residues in the *E. coli* enzyme by site-directed mutagenesis to see whether they are involved in binding the glutamate residues of the coenzyme. We are currently determining the primary structure of the rabbit mitochondrial enzyme. As sequence information is accumulated, additional amino acid residues, which have been conserved in all three enzyme forms, will be tested for function by site-directed mutagenesis on the *E. coli* enzyme.

Knowing the primary structure of the cytosolic enzyme has also aided our studies in trying to isolate the gene which codes for this enzyme from a cDNA library. We have made a series of DNA probes to the most probable nucleotide sequence coding for positions 381–394 and 253–258. Our preliminary results suggest that we have several clones which hybridize with these probes. We are currently sequencing these cDNA clones to see whether they indeed code for cytosolic serine hydroxymethyltransferase.

REFERENCES

1. Schirch, L. (1984) in *Folates and Pterins: Chemistry and Biochemistry of Folates* (Blakley, R. L., and Benkovic, S. J., eds.) pp. 399–431, Wiley-Interscience, New York
2. Hopkins, S., and Schirch, V. (1986) *J. Biol. Chem.* **261,** 3363–3369
3. Ulevitch, R. J., and Kallen, R. G. (1977) *Biochemistry* **16,** 5342–5350
4. Jones, C. W., III, and Priest, D. G. (1976) *Arch. Biochem. Biophys.* **174,** 305–311
5. Rao, D. N., and Rao, A. N. (1982) *Plant Physiol.* **69,** 11–16
6. Palekar, A. G., Tate, S. S., and Meister, A. (1973) *J. Biol. Chem.* **248,** 1158–1167
7. Miyazaki, S S., Toki, S., Izumi, Y., and Yamada, H. (1986) *Biochem. Biophys. Res. Commun.* **139,** 71–78
8. Fujioka, M. (1969) *Biochim. Biophys. Acta* **185,** 338–349
9. Stauffer, G. V., Plamann, M. D., and Stauffer, L. T. (1981) *Gene (Amst.)* **14,** 63–72
10. Plamann, M., Stauffer, L., Urbanowski, M., and Stauffer, G. (1983) *Nucleic Acids Res.* **11,** 2065–2075
11. Schirch, V., Hopkins, S., Villar, E., and Angelaccio, S. (1985) *J. Bacteriol.* **163,** 1–7
12. Gavilanes, F., Peterson, D., Bullis, B., and Schirch, L. (1983) *J. Biol. Chem.* **258,** 13155–13159
13. Gavilanes, F., Peterson, D., and Schirch, L. (1982) *J. Biol. Chem.* **257,** 11431–11436
14. Schirch, L., Slagel, S., Barra, D., Martini, F., and Bossa, F. (1980) *J. Biol. Chem.* **255,** 2986–2989
15. Martinez Carrion, M., Critz, W., and Quashnock, J. (1972) *Biochemistry* **11,** 1613–1615
16. Spencer, R. L., and Wold, F. (1969) *Anal. Biochem.* **32,** 185–190
17. Simpons, R. J., Neuberger, M. R., and Liu, T.-Y. (1976) *J. Biol. Chem.* **251,** 1936–1940
18. Barra, D., Martini, F., Angelaccio, S., Bossa, F., Gavilanes, F., Peterson, D., Bullis, B., and Schirch, L. (1983) *Biochem. Biophys. Res. Commun.* **116,** 1007–1012
19. Kondo, K., Wakabayashi, S., Yagi, T., and Kagamiyama, H. (1984) *Biochem. Biophys. Res. Commun.* **122,** 62–67
20. Barra, D., Bossa, F., Doonan, S., Fahmy, H. M. A., Hughes, G. J., Martini, F., Petruzzelli, R., and Wittmann-Liebold, B. (1980) *Eur. J. Biochem.* **108,** 405–414
21. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120
22. Schirch, L., and Peterson, D. (1980) *J. Biol. Chem.* **255,** 7801–7806
23. Barra, D., Schininà, M. E., Simmaco, M., Bannister, J. V., Bannister, W. H., Rotilio, G., and Bossa, F. (1984) *J. Biol. Chem.* **259,** 12595–12601
24. Bossa, F., Barra, D., Martini, F., Schirch, L. V., and Fasella, P. (1976) *Eur. J. Biochem.* **70,** 397–401
25. Giglio, J. R. (1977) *Anal. Biochem.* **82,** 262–264
26. Simmaco, M., Barra, D., and Bossa, F. (1985) *J. Chromatogr.* **349,** 99–103
27. Yoshida, A. (1972) *Anal. Biochem.* **49,** 320–325
28. Hargrave, P. A., and Wold, F. (1973) *Int. J. Peptide Protein Res.* **5,** 85–89
29. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
30. Tsunasawa, S., and Narita, K. (1982) *J. Biochem. (Tokyo)* **92,** 607–613
31. Bornstein, P., and Balian, G. (1977) *Methods Enzymol.* **47,** 132–145
32. Martini, F., Bossa, F., and Barra, D. (1985) *Peptides* **6,** 103–105
33. Offord, R. E. (1966) *Nature* **211,** 591–593

THE PRIMARY STRUCTURE OF RABBIT LIVER CYTOSOLIC SERINE HYDROXYMETHYLTRANSFERASE

BY

FILIPPO MARTINI, SEBASTIANA ANGELACCIO, STEFANO PASCARELLA, DONATELLA BARRA, FRANCESCO BOSSA AND VERNE SCHIRCH

EXPERIMENTAL PROCEDURES

Materials . The enzyme was isolated from rabbit liver according to Schirch and Peterson (22). Trypsin (code TRTPCK), chymotrypsin (code CDI), carboxypeptidase A (COAPMS) and B (COBPMS) were from Worthington Biochemical Co.; carboxypeptidase Y and pyroglutamate aminopeptidase from Boehringer GmbH; Staphylococcus aureus V-8 protease from Miles; thermolysin from Merck. Iodo (2-^{14}C)acetate was from Radiochemical Centre, Amersham, and guanidine-HCl (from Carlo Erba) was recrystallized from methanol. Cyanogen bromide, glycinamide and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide were from Fluka.

S-carboxymethylation of the protein . Alkylation with iodo (2-^{14}C) acetate, was performed as recently described (23).

Proteolytic cleavages . A sample (500 mg) of enzyme reduced with sodium borohydride (24) and then S-carboxymethylated was suspended in 25 ml of 0.1 M ammonium bicarbonate and incubated at 37° C for 2.5 h after addition of 10 mg of trypsin. The cleavage reaction was stopped by addition of glacial acetic acid to 10% final concentration. Peptides, soluble under these conditions, were separated from the insoluble ones by centrifugation and the two fractions separately lyophilized. A second sample of 300 mg of S-carboxymethylated protein was dissolved in 7 ml of 70% (v/v) formic acid and incubated with 600 mg of CNBr for 20 h at room temperature in the dark.

Cleavage of the largest peptide fractions was performed with trypsin, chymotrysin or thermolysin at an E/S ratio 1/30 for 2 to 5 h at 37°C in 0.1 M ammonium bicarbonate. Subdigestions with S. aureus protease were performed under identical conditions except the E/S ratio was 1/20 and the presence of 2 mM EDTA included in the digestion mixture. When the digestion mixture remained insoluble, acetonitrile was added to 10-20% (v/v) final concentration to improve substrate solubilization.

Peptide purification . The tryptic peptides, soluble in 10% acetic acid, were first fractionated by gel-filtration on a Sephadex G-25 fine column (2.5 x 120 cm) in 10% acetic acid (Fig. 3). These peptides were then purified by ion-exchange and thin layer chromatography or electrophoresis as previously described (23). An aliquot from the various fractions collected after Sephadex G-25 chromatography was set apart and purified using a Beckman model 332 hplc, on macroporous reverse phase columns (Aquapore RP-300, 4.6 or 7.0 x 250 mm, 10 μm, Brownlee Labs) eluted with gradients of 0 to 70% acetonitrile in 0.2% (v/v) trifluoroacetic acid, at a flow rate of 1.2 or 3.0 ml/min. Elution of the peptides was monitored on a Beckman 165 Spectrophotometer at 220 and 280 or 325 nm. Products of secondary fragmentation procedures were purified by high performance liquid chromatography (hplc) under similar conditions.

CNBr peptides were first fractionated on a column (2.5 x 140 cm) of Sephadex G-50 superfine in 10% acetic acid (Fig. 4). The various fractions were subdigested with trypsin, chymotrypsin, thermolysin or with S. aureus protease either directly or after further purification by hplc as described above.

The subdigestion mixtures were also generally purified by a two-step procedure, utilizing pre-fractionation by gel-filtration (Sephadex G-25 superfine, 2.3 x 120 cm; 10% acetic acid) and final purification by reverse-phase hplc.

Analytical techniques . Amino acid analysis and sequence determination were performed as described in reference 23. Dansyl-Trp was identified after 5 h-hydrolysis at 110°C with 4 N methanesulfonic acid containing 0.2% 3-(2-aminoethyl)indole, other conditions being identical to those described in reference 25. More recently, the amidation states of Glu and Asp were assigned by direct identification by hplc of the phenylthiohydantoin (PTH) derivatives released during dansyl-Edman sequence analysis. The same procedure was used for the identification of the PTH derivative of carboxymethylcysteine (26). In

some cases peptides were sequenced automatically by using a model 470 A gas-phase sequencer from Applied Biosystems equipped with a model 120 A PTH analyzer for the on-line detection of PTH-amino acids.

Isolation of the N-terminal blocked peptide . The classical procedure of Yoshida (27) was applied with some minor modifications. Carboxymethylated protein (35 mg) was digested with 1 mg of trypsin for 4 h at 37°C in 0.1 M ammonium bicarbonate and then with 0.4 mg of carboxypeptidase B for 3 h under the same conditions. The mixture of peptides after repeated lyophilization was loaded onto a column (2 x 25 cm) of SP-Sephadex G-25 in the H^+ form and the column eluted with 0.1 M acetic acid. The eluted fraction was lyophilized and subjected to final purification by hplc under the conditions described in the section Peptide purification.

Isolation of the C-terminal peptide . A modified version of the procedure reported by Hargrave and Wold (28) was followed. Carboxymethylated enzyme (25 mg) was dissolved in 2.5 ml of 8 M urea solution; 1 M glycinamide and 0.1 M 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide. The reaction mixture was adjusted to and mantained at pH 4.75 for 3 h at room temperature. The protein was then extensively dialyzed against 1 mM acetic acid and lyophilized. The modified protein was suspended in 2.0 ml of 0.1 M ammonium bicarbonate and first digested for 4h at 37°C with 2.5 mg of trypsin and then with 0.4 mg of carboxypeptidase B overnight.

After repeated lyophilization, the trypsin-carboxypeptidase B digest was chromatographed on a column (2 x 25 cm) of QAE-Sephadex A-25 equilibrated and eluted with 1 mM NaOH. The unretained fraction was concentrated by lyophilization and subjected to a final purification step by hplc under conditions described in the section Peptide purification.

Peptide nomenclature . The peptides were numbered retrospectively according to their location in the sequence, starting from the N-terminus. Tryptic peptides were designated with a T, CNBr peptides with a B. All fragments obtained from subdigestions of T or B peptides were designated by a letter indicating the proteolytic method (t = trypsin; C = chymotrypsin, L = thermolysin; S = S. aureus protease) and numbered according to their position in the sequence, starting from the N-terminus.

Structure predictions . Secondary structure of the proteins were predicted according to the method of Garnier et al. (21). The decision constant (DC_S) for alpha-helix was set at -75, the DC_S for beta-sheet to -87.5 and the others to zero. The hydropathic profiles were calculated according to Kyte and Doolittle (29). The plots utilize a span setting of 7: the original hydropathy indexes were multiplied by -1 to show the hydrophilic values as positive and the hydrophobic values as negative. In both cases calculations were performed by means of simple programs written in Applesoft Basic on an Apple IIe.

RESULTS

N-terminal sequence of the protein . Final hplc purification of the acidic peptides obtained after application of the procedure of Yoshida aimed at the isolation of N-terminal blocked peptides is shown in Fig. 5. Both peaks 1 and 2 correspond to peptide material with an unreactive N-terminal residue. Further analyses of the peptide contained in peak 1 identified the following structure: Glp-Ala-Met-Thr-Pro-Glu. This peptide clearly originated from a tryptic fragment corresponding to T24, with the N-terminal glutamine cyclized to pyrrolidone carboxylic acid and the C-terminal Phe and Lys residues removed after the carboxypeptidase B treatment. Amino acid analysis of peak 2 gave the following composition (mol per mol of peptide): Asp(1.1), Thr(0.8), Pro(1.0), Gly(1.0), Ala(2.8) and Val(1.1). This peptide was subdigested with thermolysin for 2h at 37°C in 0.1M ammonium bicarbonate (E/S ratio = 1/50) and the fragments purified by hplc. Fragment L1 was unreactive to dansyl-Cl and had the following amino acid composition: Ala (2.0), Thr (0.8). Digestion of this material with carboxypeptidase A released after 30 min Ala (0.6) and after 3 h Ala (1.0) and Thr (0.8) and an N-terminal blocked residue which cochromatographed with a standard of authentic acetyl-alanine when analyzed by hplc on a Spheri-5 column (4.6 x 250 mm, 5 um, Brownlee Labs), eluted with a linear gradient of acetonitrile in 0.2% aqueous trifluoroacetic acid, according to a slightly modified version of the method of Tsunasawa (30). The sequence of fragment L2, as determined by the dansyl-Edman procedure, was: Val-Asn-Gly-Ala-Pro. The whole of these results indicated the following structure for the N-terminus of the protein: acetyl-Ala-Thr-Ala-Val-Asn-Gly-Ala-Pro.

The material corresponding to peak 2a had the same analytical features as the peptide in peak 2, but was present in much less quantity. The chromatographic difference could be explained by the formation of a β-aspartyl peptide bond from the Asn-Gly at positions 5-6 (31).

C-terminal sequence of the protein . Final hplc purification of the peptide fraction obtained after the procedure of Hargrave and Wold aimed at the isolation of the C-terminal fragment of the protein allowed the isolation of two peptides. Analysis of one of these revealed the sequence: Ser-Ala-Leu-Arg-Pro-Ser-Gly which identifies it with the tryptic peptide T33, clearly originating from an internal region of the polypeptide chain. The second peptide was completely analyzed by dansyl-Edman degradation and had the following structure: Ala-Ala-Leu-Phe-Pro-Leu-Pro-Gly-Leu-Pro-Gly-Phe. After the 12th cycle of Edman degradation, the material remaining in the reaction tube was reacted with dansyl-Cl and analyzed by thin layer chromatography without previous hydrolysis with 6 N HCl. A spot with the same chromatographic properties of dansyl-glycinamide was clearly evident, thus indicating that the peptide originated from the C-terminus of the polypeptide chain. It should be noted that this peptide actually originated from a chymotryptic-like cleavage at a Phe-Ala (residues 471-472) bond, possibly due to the relatively high E/S (1/10) ratio used during the tryptic digestion. The tryptic peptide corresponding to the segment from Gln 466 to Phe 483 plus glycinamide was not found.

Tryptic peptides . Analytical data on the tryptic peptides are reported in Tables II and III. Arginine and lysine were also obtained as free amino acids from the tryptic digest. In most cases each peptide became available through two different purification procedures (see the section Peptide purification). In these cases only the data relative to the peptide which gave the more successful analyses were reported. Peptides T24 and T26 with a cyclyzed Gln at the N-terminus, were digested with pyroglutamate aminopeptidase and purified as reported in reference 32. The deblocked peptides were further analyzed as indicated in Table III. Amino acid compositional data for the fragments obtained after subdigestion of some tryptic peptides with various proteolytic enzymes are reported in Table IV. Detailed analytical data for peptides T2, T11, T17 and T23, which were sequenced by automated Edman degradation, are reported in Table V.

Cyanogen bromide peptides . Analytical data on the cyanogen bromide peptides are reported in Tables VI and VII. Amino acid compositional data for the fragments obtained after subdigestion of some cyanogen bromide peptides with various proteolytic enzymes are reported in Tables VIII-XII. Detailed analytical data for peptide B3, B4, B6, B7S9 and B7S11, which were sequenced by automated Edman degradation, are reported in Tables XIII and XIV.

Peptide B1 was subdigested with chymotrypsin and fragments C1, C1a and C2 corresponding to residues 1-14, 6-14 and 15-20, respectively, were purified and analyzed (Tables VII and VIII). Fragment C1 was digested with carboxypeptidase Y, which released after 1 h Trp (0.9) (mol/mol of peptide), Leu (0.8) and Ala (0.6) and after 3 h Trp (1.0), Leu (1.0), Ala (1.8) and Asp (0.7). Fragment C1 was further subdigested with thermolysin and subfragments C1L1 (1-3), C1L2 (3-12), C1L2a (4-12) were purified and analyzed (Tables VII and VIII). Fragment C1L1 was digested for 3 h with carboxypeptidase Y, which released Ala (1.0) and Thr (0.7). These results, together with those obtained from the selective procedure outlined above, allowed reconstruction of the structure of the N-terminal region of the protein.

Peptide B2 was subdigested with trypsin and the fragments t1-t11, corresponding to residues 21-38, 38-42, 43-45, 46-58, 59-71, 72-80, 81-98, 99-103, 104-134, 135-136 and 137-138, were purified and analyzed (Table IX, where only the amino acid compositions of tryptic peptides not found in the tryptic digest of the whole protein are reported). A second aliquot of B2 was subdigested with S.aureus protease and fragments S1-S11, corresponding to residues 21-31, 32-39, 40-48, 49-56, 57-62, 63-74, 75-90, 91-92, 93-104, 105-118 and 119-138, were purified and analyzed (Table IX). The sequence information obtained from the various fragments of B2 is summarized in Table VII and allowed reconstruction of its complete structure.

Peptide B3 was sequenced by automated Edman degradation up to residue 161 (Table XIII). An aliquot was subdigested with trypsin and two fragments corresponding to residues 139-157 and 158-169 were isolated and analyzed (Table X). The sequence information, reported in Table VII, allowed reconstruction of the complete structure of peptide B3.

Peptide B4 was sequenced by automated Edman degradation up to residue 221 (Table XIV). An aliquot was subdigested with thermolysin and fragments L1-L7, corresponding to residues 170-179, 180-191, 192-196, 197-207, 208-215, 216-223 and 224-225 were isolated and analyzed (Table X). The sequence information reported in Table VII allowed reconstruction of the structure of peptide B4.

Peptide B5 was directly analyzed by dansyl-Edman degradation (Table VII).

Peptide B6 was sequenced by automated Edman degradation up to residue 256 (Table XIII). An aliquot was subdigested with chymotrypsin and fragments C1-C5, corresponding to residues 229-234, 235-244, 245-255, 256-258 and 259-265 were isolated and analyzed (Table XI). A second aliquot was digested with S. aureus protease and two fragments S1 and S2, corresponding to residues 229-245 and 246-265, were isolated and analyzed (Table XI). The sequence information reported in Table VII allowed reconstruction of the complete structure of peptide B6.

Peptide B7 was subdigested with S. aureus protease and fragments S1-S16, corresponding to residues 266-282, 283-288, 289-322, 313-322, 323-325, 326-343, 344-359, 360-373, 374-377, 378-403, 404-414, 415-429, 430-446, 447-449, 450-455, 456-467, 468-471, 470-483 were isolated and analyzed (Tables XII and XIII). A second aliquot was subdigested with chymotrypsin. In Table XII are reported the analytical data only of those fragments, corresponding to residues 274-285, 286-290, 305-315, 318-326, 338-346, 364-376, 377-384, 403-407, 409-417, 424-430, 445-447, 448-458, 465-471 and 472-483, which were useful for reconstructing the sequence of peptide B7. The sequence information obtained from the various subfragments of B7 is summarized in Table VII and allowed reconstruction of its complete sequence.

Structure predictions . Prediction of the secondary structure and the hydropathic profiles for E. coli and rabbit liver serine hydroxymethyltransferase are shown in Fig. 6 and 7, respectively. The two sequences were aligned according to Fig. 2. The predicted secondary structure homology between the two proteins is 49.1%, which is higher than their sequence homology (42.0%). The predicted content in alpha-helix and extended conformation is, respectively, 48.2% - 40.0% for the E. coli enzyme and 46.2 - 34.5 for the mammalian enzyme.

TABLE II

Amino Acid Compositions of Tryptic Peptides

The composition from sequence analysis of each peptide is indicated by the numbers in parentheses

Peptide	T 1	T 2	T 3	T 4	T 5	T 6	T 7	T 8	T 9	T 10	T 11	T 12
Residue nos.	1-9	10-38	39-42	43-44	45-58	59-71	72-80	81-98	99-136	137-146	158-172	173-190
cmCys						0.9(1)		0.6(1)	0.9(1)			
Asp	1.1(1)	3.7(4)	1.0(1)		0.9(1)	1.9(2)		1.0(1)	3.0(3)	3.0(3)		4.5(5)
Thr	0.8(1)							1.5(2)	1.1(1)	2.1(2)	0.9(1)	0.9(1)
Ser		2.4(3)	0.7(1)		1.5(2)	0.8(1)	0.9(1)		2.0(2)		2.5(3)	
Glu		4.0(4)	0.9(1)	0.8(1)	1.8(2)	1.0(1)	2.0(2)	3.6(4)	3.9(4)		1.0(1)	2.0(2)
Pro	1.0(1)	0.9(1)					0.8(1)		3.2(4)	1.1(1)		0.9(1)
Gly	1.0(1)				1.0(1)	0.9(1)	2.0(2)	2.0(2)	3.6(4)	4.0(4)		1.2(1)
Ala	2.7(3)	3.8(4)			1.7(2)	1.8(2)			4.4(5)		2.0(2)	1.0(1)
Val	0.9(1)	1.0(1)			0.8(1)	0.8(1)			3.2(4)			0.9(1)
Met		1.0(1)								1.2(2)	0.8(1)	
Ile		1.4(2)			0.9(1)			0.9(1)		1.0(1)	1.6(2)	1.0(1)
Leu		2.8(3)			1.9(2)	3.0(3)		2.0(2)	3.0(3)	3.2(3)		1.1(1)
Tyr		1.0(1)					2.0(2)	1.7(2)	2.3(3)		0.9(1)	1.5(2)
Phe					0.9(1)				1.1(1)	0.8(1)	1.8(2)	
His		0.9(1)						0.9(1)	0.5(1)	1.9(2)		
Lys		2.7(3)				0.9(1)		1.0(1)		1.0(1)	2.0(2)	
Arg	0.9(1)		0.9(1)	1.0(1)	1.0(1)		1.1(1)	1.1(1)	0.6(1)			1.8(2)
Trp		+(1)							+(1)			
Purification steps	G,H	G,I,H	G,I,E	G,I,C	G,I	G,I	G,I,C,H	G,I	G,I,H	G,I	G,I	G,I,C
Total yield %	12.3	6.1	32.8	11.4	12.4	23.8	6.7	21.6	6.0	34.0	25.8	12.4
Net Charge	ND	ND	0	+1	-1	-1	ND	ND	ND	ND	ND	ND

Peptide	T 13	T 14	T 15	T 16	T 16a	T 17	T 18	T 19	T 20	T 21	T 22	T 23
Residue nos.	191-195	196-206	207-212	213-215	213-214	216-259	260-262	263-269	271-273	274-278	279-281	282-316
cmCys		0.4(1)				0.8(1)	0.4(1)					
Asp			1.8(2)			2.8(3)				1.0(1)		3.0(3)
Thr		1.0(1)				4.2(5)					1.0(1)	
Ser		1.7(2)				2.0(2)				1.0(1)		1.9(2)
Glu						2.1(2)						3.1(3)
Pro	1.0(1)					1.4(2)				0.7(1)		2.0(2)
Gly		1.1(1)	1.0(1)			3.0(3)	1.0(1)	1.0(1)	0.9(1)		1.0(1)	3.9(4)
Ala		1.0(1)				4.9(5)		0.7(1)				4.7(5)
Val						3.2(6)			1.0(1)	0.9(1)		2.9(3)
Met						1.8(2)		0.9(1)				
Ile		1.6(2)				1.6(2)		0.7(1)				2.5(3)
Leu	1.0(1)	0.8(1)	0.9(1)	1.0(1)	1.0(1)	2.9(3)						4.5(5)
Tyr		0.7(1)	1.0(1)			0.9(1)		1.0(1)				1.0(1)
Phe	1.0(1)					1.0(1)		1.0(1)				1.0(1)
His	0.9(1)					3.4(4)						1.8(2)
Lys	1.0(1)			0.9(1)		1.1(1)				1.0(1)	0.9(1)	1.0(1)
Arg		1.1(1)	0.9(1)	1.0(1)	1.0(1)	0.7(1)	1.1(1)	1.1(1)	1.1(1)			
Trp												
Purification steps	G,I,C	G,I,E	G,I,E,H	G,H	G	G,I,H	G,I,E	G,I	G,I,C	G,I,H	G,I,C	G,I,H
Total yield %	37.8	4.3	38.9	0.1	48.9	1.9	4.0	31.1	5.2	12.2	8.9	12.5
Net Charge	ND	0	0	ND	ND	ND	0	ND	ND	ND	ND	ND

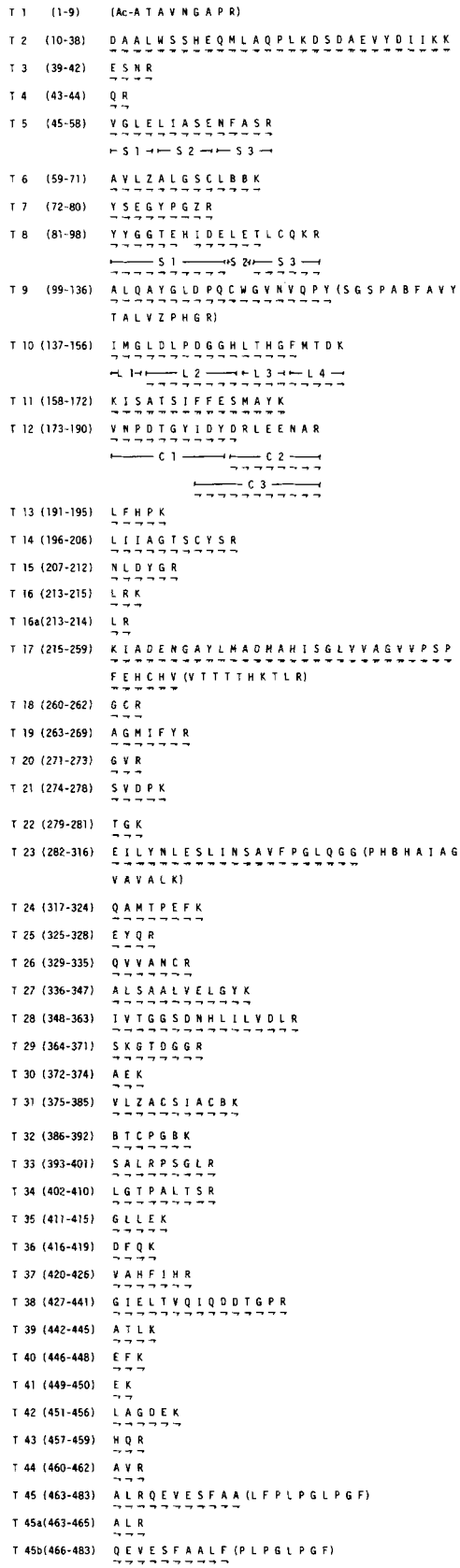
Peptide	T 24	T 25	T 26	T 27	T 28	T 29	T 30	T 31	T 32	T 33	T 34	T 35
residues nos.	317-324	325-328	329-335	336-347	348-363	364-371	372-374	375-385	386-392	393-401	402-410	411-415
CmCys			0.9(1)					1.3(2)	1.0(1)			
Asp			1.0(1)		3.0(3)	1.0(1)		1.0(1)	2.0(2)			
Thr	0.9(1)					1.0(1)	0.9(1)			0.9(1)		1.7(2)
Ser				1.0(1)	1.0(1)	0.9(1)		0.8(1)		1.6(2)	1.0(1)	
Glu	2.1(2)	2.0(2)	1.0(1)	1.0(1)			1.0(1)	1.0(1)				1.0(1)
Pro	1.0(1)								0.9(1)	1.0(1)	1.0(1)	
Gly				0.9(1)	2.0(2)	2.8(3)			1.1(1)	0.9(1)	1.0(1)	0.9(1)
Ala	1.0(1)		0.9(1)	2.8(3)			1.0(1)	2.0(2)		1.0(1)	1.0(1)	
Val			1.4(2)	0.8(1)	1.6(2)			0.9(1)				
Met												
Ile					1.6(2)			1.0(1)				
Leu				2.8(3)	2.7(3)			1.1(1)		2.1(2)	1.9(2)	1.8(2)
Tyr		0.9(1)		0.8(1)								
Phe	0.9(1)											
His					1.0(1)							
Lys	0.8(1)			0.9(1)		1.1(1)	0.9(1)	0.9(1)	1.1(1)			1.1(1)
Arg		0.9(1)	0.8(1)		1.0(1)	0.7(1)				2.0(2)	0.9(1)	
Trp												
Purification steps	G,I,E	G,I,E	G,H	G,I,H	G,H	G,I	G,I	G,I	G,I	G,I,C	G,I	G,I,C
Total yield %	17.8	13.6	8.2	1.3	10.0	18.0	25.3	32.1	60.3	20.3	50.0	45.6
Net Charge	0	0	ND	ND	ND	+1	0	-2	-1	ND	+1	0

Peptide	T 36	T 37	T 38	T 39	T 40	T 41	T 42	T 43	T 44	T 45	T 45a	T 45b
residues nos.	416-419	420-426	427-441	442-445	446-448	449-450	451-456	457-459	460-462	463-483	463-465	466-483
CmCys												
Asp	1.0(1)		1.7(2)				1.0(1)					
Thr			1.5(2)	0.9(1)								
Ser										0.9(1)		0.8(1)
Glu	0.8(1)		2.8(3)		1.0(1)	1.0(1)	1.0(1)	1.0(1)		3.0(3)		2.8(3)
Pro			0.7(1)							2.7(3)		2.7(3)
Gly			2.0(2)				1.0(1)			2.0(2)		2.0(2)
Ala		1.0(1)		1.0(1)			0.9(1)		1.0(1)	2.8(3)	1.0(1)	1.9(2)
Val		0.9(1)	0.8(1)						0.9(1)	0.9(1)		0.7(1)
Met												
Ile		0.7(1)	1.4(2)									
Leu			0.9(1)	1.1(1)			0.7(1)			3.6(4)	1.0(1)	2.6(3)
Tyr												
Phe	0.8(1)	0.8(1)			1.0(1)					2.6(3)		2.7(3)
His		1.5(2)						0.7(1)				
Lys	1.0(1)			1.0(1)	1.0(1)	0.9(1)	1.0(1)					
Arg		1.0(1)	0.8(1)					1.0(1)	0.9(1)	0.9(1)	0.9(1)	
Trp												
Purification steps	G,I	G,I	G,H	G,I,C	G,I	G,I,C	G,I,E	G,I	G,I,C	G,I	G,H	G,H
Total yield %	90.6	28.9	5.7	43.3	15.9	50.0	4.4	13.3	14.8	55.0	27.3	6.7
Net Charge	0	ND	ND	ND	0	0	-1	+2	ND	ND	ND	ND

Presence of tryptophan was indicated by Ehrlich reaction and/or by absorbance at 280 nm. The purification steps required for peptide purification are indicated as follows: G = gel filtration; I = ion-exchange chromatography on Chromobead P resin; E = electrophoresis at pH 6.5; H = reverse-phase hplc. The net charge was determined after electrophoresis at pH 6.5 according to Offord (33).

TABLE III

Summary of Sequence Studies on Tryptic Peptides



The residues above the arrow were identified by dansyl-Edman degradation (→) or by automated Edman degradation on a gas-phase sequencer (⇒). Subfragments obtained after digestion with S. aureus protease (S), thermolysin (L) or chymotrypsin (C) are indicated by solid lines.

TABLE IV

Amino Acid Composition of Fragments Obtained after Subdigestion of Tryptic Peptides

The composition from sequence analysis of each peptide is indicated by the numbers in parentheses

Peptide	T5 S1	T5 S2	T5 S3	T8 S1	T8 S2	T8 S3	T10 L1	T10 L2	T10 L3	T10 L4	T12 C1	T12 C2	T12 C3
Residue nos.	45-48	49-53	54-58	81-90	91-92	93-98	137-139	140-147	148-151	152-156	173-182	183-190	180-190
cmCys						0.8(1)							
Asp			1.0(1)	1.0(1)				2.0(2)		1.0(1)	2.9(3)	2.0(2)	3.0(3)
Thr				0.8(1)		0.9(1)			0.9(1)	0.9(1)	0.8(1)		
Ser		0.9(1)	0.8(1)										
Glu	1.0(1)	1.0(1)		1.9(2)	1.0(1)	1.0(1)						2.0(2)	1.9(2)
Pro								0.9(1)			0.8(1)		
Gly	1.1(1)			2.1(2)			1.0(1)	2.0(2)	1.0(1)		1.0(1)		
Ala		1.0(1)	1.1(1)									1.0(1)	1.0(1)
Val	0.9(1)										0.9(1)		
Met							0.9(1)			0.9(1)			
Ile		0.8(1)		0.8(1)			0.9(1)				0.8(1)		0.8(1)
Leu	1.0(1)	0.9(1)			0.9(1)	1.0(1)		1.9(2)	1.0(1)			1.0(1)	0.9(1)
Tyr				1.9(2)							1.7(2)		0.8(1)
Phe			0.9(1)							0.9(1)			
His				0.8(1)				0.8(1)	0.9(1)				
Lys						0.9(1)				1.0(1)			
Arg			0.9(1)			1.0(1)						1.8(2)	1.8(2)
N-terminal residue	Val	Leu	Asx	Tyr	Leu	Thr	Ile	Leu	Leu	Phe	Val	Asx	Ile

TABLE V

Automated Edman Degradation of Tryptic Peptides

Cycle No.	T 2 (2.5 nmol)		T 11 (1.5 nmol)		T 17 (1.5 nmol)		T 23 (1.5 nmol)	
	PTH-aa	Yield (pmol)	PTH-aa	Yield (pmol)	PTH-aa	Yield (pmol)	PTH-aa	Yield (pmol)
1	Asp	1443	Lys	446	Lys	696	Glu	618
2	Ala	1263	Ile	300	Ile	700	Ile	564
3	Ala	1327	Ser	293	Ala	515	Leu	562
4	Leu	971	Ala	418	Asp	410	Tyr	485
5	Trp	798	Thr	390	Glu	440	Asn	420
6	Ser	732	Ser	156	Asn	228	Leu	376
7	Ser	894	Ile	185	Gly	253	Glu	342
8	His	346	Phe	178	Ala	230	Ser	155
9	Glu	551	Phe	205	Tyr	156	Leu	305
10	Gln	729	Glu	145	Leu	175	Ile	227
11	Met	867	Ser	110	Met	105	Asn	199
12	Leu	787	Met	158	Ala	103	Ser	53
13	Ala	425	Ala	130	Asp	105	Ala	86
14	Gln	550	Tyr	70	Met	62	Val	39
15	Pro	447	Lys	113	Ala	52	Phe	47
16	Leu	410			His	NQ[a]	Pro	46
17	Lys	369			Ile	33	Gly	53
18	Asp	189			Ser	26	Leu	39
19	Ser	215			Gly	50	Gln	62
20	Asp	177			Leu	30	Gly	56
21	Ala	119			Val	13	Gly	73
22	Glu	132			Val	25		
23	Val	147			Ala	32		
24	Tyr	111			Gly	34		
25	Asp	79			Val	9		
26	Ile	62			Val	15		
27	Ile	67			Pro	11		
28	Lys	29			Ser	16		
29	Lys	33			Pro	13		
30					Phe	12		
31					Glu	10		
32					His	NQ[a]		
33					cmCys	4		
34					His	NQ[a]		
35					Val	12		
Repetitive yield %	97 (Leu 8-Leu 12)		91 (Ile 2-Ile 7)		82 (Ile 2-Ile 17)		93 (Ile 2-Ile 10)	
Position	10-38		158-172		215-259		282-316	

[a] NQ: observed but not quantitated.

TABLE VI

Amino Acid Composition of CNBr Peptides

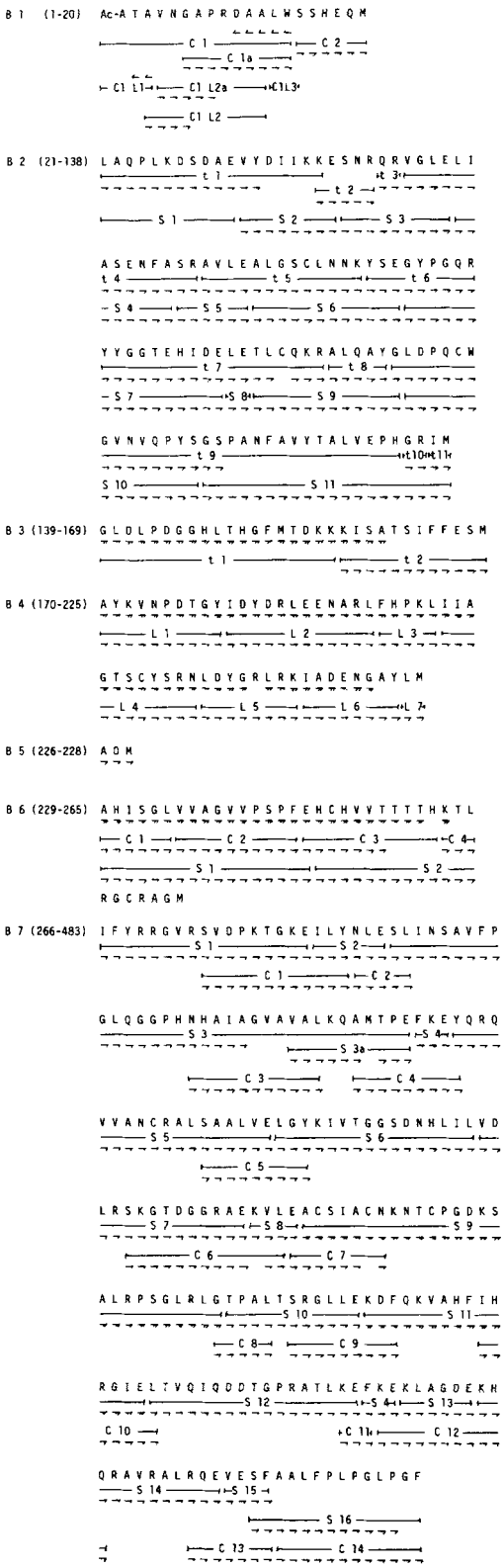
The composition from sequence analysis of each peptide is indicated by the numbers in parentheses

Peptide	B 1	B 2	B 3	B 4	B 5	B 6	B 7
residues nos.	1-20	21-138	139-169	170-225	226-228	229-265	266-483
cmCys		3.0(3)		0.9(1)		2.0(2)	4.1(4)
Asp	2.0(2)	10.9(11)	3.0(3)	8.8(9)	1.0(1)		15.8(16)
Thr	1.0(1)	3.1(3)	2.9(3)	1.9(2)		4.3(5)	9.7(10)
Ser	1.7(2)	8.1(8)	2.5(3)	1.7(2)		1.8(2)	10.5(11)
Glu	2.1(2)	17.2(17)	1.1(1)	3.1(3)		1.1(1)	22.8(23)
Pro	1.1(1)	5.6(6)	1.1(1)	2.0(2)		1.8(2)	10.7(11)
Gly	1.0(1)	10.3(10)	3.8(4)	3.8(4)		3.8(4)	21.0(21)
Ala	5.0(5)	11.0(11)	1.0(1)	5.0(5)	1.0(1)	3.0(3)	21.8(22)
Val	1.0(1)	6.5(7)		1.0(1)		4.3(6)	14.7(15)
Ile		4.2(5)	2.0(2)	3.2(4)		0.9(1)	9.2(10)
Leu	1.0(1)	11.3(12)	3.0(3)	5.7(6)		1.8(2)	25.0(25)
Tyr		6.2(8)		5.8(6)			3.9(4)
Phe		1.7(2)	3.1(3)	1.1(1)		1.0(1)	8.6(9)
His	1.1(1)	1.8(2)	2.0(2)	1.0(1)		3.7(4)	3.3(6)
Lys		5.4(5)	2.9(3)	2.7(3)		1.0(1)	15.2(15)
Arg	0.9(1)	6.2(6)		5.1(5)		2.0(2)	14.8(15)
Trp	+(1)	+(1)					
Hse/Hsl	+(1)	+(1)	+(1)	+(1)	+(1)	+(1)	+(1)
Total yield%	17.2	42.6	19.8	31.1	29.6	39.7	29.3
Purification steps	G,H	G	G,H	G,H	G,H	G,H	G

Presence of tryptophan was indicated by Ehrlich reaction and/or by absorbance at 280 nm. The purification steps required for peptide purification are indicated as follows: G = gel filtration; H = reverse-phase hplc.

TABLE VII

Summary of Sequence Studies on CNBr Peptides



The residues above the arrow were identified by dansyl-Edman degradation (→) or by automated Edman degradation on a gas-phase sequencer (⇝). The sequences determined by carboxypeptidase Y digestion are indicated by arrows (←) above the corresponding residues. Subfragments obtained after digestion with chymotrypsin (C), thermolysin (L), S. aureus protease (S) or trypsin (t) are indicated by solid lines.

TABLE VIII

Amino Acid Composition of Fragments Obtained after Subdigestion of B 1

The composition from sequence analysis of each peptide is indicated by the numbers in parentheses

Peptide	C 1	C 1a	C 2	C1 L1	C1 L2	C1 L2a	C1 L3
residue nos.	1-14	6-14	15-20	1-3	3-12	4-12	13-14
Asp	1.9(2)	1.0(1)			2.0(2)	2.1(2)	
Thr	0.9(1)			1.0(1)			
Ser			1.8(2)				
Glu			2.0(2)				
Pro	0.9(1)	0.7(1)			0.8(1)	1.1(1)	
Gly	1.0(1)	0.8(1)			1.2(1)	1.1(1)	
Ala	5.1(5)	3.0(3)		2.0(2)	3.7(4)	3.0(3)	
Val	1.0(1)				1.0(1)	1.0(1)	
Ile							
Leu	1.0(1)	0.9(1)					1.0(1)
His			0.8(1)				
Arg	1.1(1)	0.6(1)			0.8(1)	1.0(1)	
Trp	+ (1)	+ (1)					+ (1)
Hse/Hsl			+ (1)				
N-terminal residue	-	Gly	Ser	-	Ala	Val	Leu

Presence of tryptophan was indicated by absorbance at 280 nm.

TABLE IX

Amino Acid Composition of Fragments Obtained after Subdigestion of B 2

The composition from sequence analysis of each peptide is indicated by the numbers in parentheses

Peptide	t 1	t 2	t 8	t 9	t 10	t 11	S 1	S 2	S 3
residue nos.	21-38	38-42	99-103	104-134	135-136	137-138	21-31	32-39	40-48
cmCys				0.9(1)					
Asp	3.0(3)	0.8(1)		2.9(3)			2.0(2)	1.0(1)	1.0(1)
Thr				0.9(1)					
Ser	1.1(1)	0.7(1)		1.7(2)			0.9(1)		0.9(1)
Glu	2.0(2)	1.0(1)	1.1(1)	3.0(3)			2.1(2)	1.1(1)	2.2(2)
Pro	1.1(1)			3.8(4)			1.0(1)		
Gly				3.0(3)	1.0(1)				1.0(1)
Ala	2.0(2)		2.0(2)	2.9(3)			1.9(2)		
Val	0.9(1)			3.6(4)				0.9(1)	0.9(1)
Ile	1.3(2)					1.0(1)		1.6(2)	
Leu	2.1(2)		1.0(1)	2.0(2)			2.0(2)		1.0(1)
Tyr	0.8(1)		1.0(1)	1.8(2)				0.9(1)	
Phe				0.9(1)					
His				0.9(1)					
Lys	3.1(3)	0.9(1)					1.0(1)	1.9(2)	
Arg		1.0(1)			0.9(1)				1.8(2)
Trp				+(1)					
Hse/Hsl						+(1)			
N-terminal residue	Leu	Lys	Ala	Gly	Gly	Ile	Leu	Val	Ser

Peptide	S 4	S 5	S 6	S 7	S 8	S 9	S 10	S 11
residue nos.	49-56	57-62	63-74	75-90	91-92	93-104	105-118	119-138
cmCys			0.8(1)			0.9(1)	1.0(1)	
Asp	1.0(1)		2.0(2)	1.0(1)			2.0(2)	1.1(1)
Thr				0.8(1)		0.8(1)		0.8(1)
Ser	1.0(1)	0.9(1)	1.7(2)				0.9(1)	0.9(1)
Glu	1.1(1)	1.1(1)	1.0(1)	2.9(3)	1.0(1)	2.0(2)	2.1(2)	1.1(1)
Pro				1.0(1)			1.9(2)	1.6(2)
Gly			1.1(1)	3.9(4)		1.1(1)	1.0(1)	1.8(2)
Ala	1.9(2)	1.0(1)	1.0(1)			2.0(2)		3.0(3)
Val		1.0(1)					1.9(2)	1.7(2)
Ile	1.0(1)			0.7(1)				0.9(1)
Leu	1.1(1)	1.0(1)	1.7(2)		0.7(1)	1.8(2)	0.9(1)	1.0(1)
Tyr			0.8(1)	2.5(3)		1.0(1)	1.0(1)	0.8(1)
Phe	1.1(1)							1.0(1)
His				0.9(1)				0.8(1)
Lys			0.9(1)			1.0(1)		
Arg		1.0(1)		0.9(1)		0.9(1)		0.7(1)
Trp							+(1)	
Hse/Hsl								+(1)
N-terminal residue	Leu	Ser	Ala	Gly	Leu	Thr	Leu	Gly

Presence of tryptophan was indicated by absorbance at 280 nm.

TABLE X

Amino Acid Composition of Fragments Obtained after Subdigestion of B 3 and B 4

The composition from sequence analysis of each peptide is indicated by the numbers in parentheses

	B 3		B 4						
Peptide	t 1	t 2	L 1	L 2	L 3	L 4	L 5	L 6	L 7
residue nos.	139-157	158-169	170-179	180-191	192-196	197-207	208-215	216-223	224-225
cmCys						0.9(1)			
Asp	2.8(3)		2.0(2)	2.9(3)		1.0(1)	1.0(1)	2.0(2)	
Thr	1.8(2)	0.9(1)	1.0(1)			0.9(1)			
Ser		2.6(3)				1.9(2)			
Glu		1.0(1)		2.0(2)				1.0(1)	
Pro	1.0(1)		0.8(1)		1.0(1)				
Gly	4.0(4)		1.1(1)			1.1(1)	1.0(1)	1.1(1)	
Ala		1.0(1)	1.0(1)	0.9(1)		1.0(1)		1.9(2)	
Val			1.0(1)						
Ile		1.8(2)		0.9(1)		1.4(2)		0.9(1)	
Leu	2.7(3)			1.9(2)	1.1(1)		1.9(2)	0.9(1)	1.0(1)
Tyr			1.7(2)	0.9(1)		1.0(1)	0.9(1)	1.0(1)	
Phe	0.8(1)	1.8(2)			1.0(1)				
His	1.7(2)				1.0(1)				
Lys	1.7(2)	1.0(1)	1.0(1)		1.0(1)		1.0(1)		
Arg			2.0(2)		1.0(1)	1.6(2)			
Hse/Hsl	+(1)	+(1)							+(1)
N-terminal residue	Gly	Lys	Ala	Ile	Phe	Ile	Leu	Ile	Leu

TABLE XI

Amino Acid Composition of Fragments Obtained after Subdigestion of B 6

The composition from sequence analysis of each peptide is indicated by the numbers in parentheses

Peptide	C 1	C 2	C 3	C 4	C 5	S 1	S 2
residue nos.	229-234	235-244	245-255	256-258	259-265	229-245	246-265
cmCys			0.9(1)		1.0(1)		1.9(2)
Thr			3.7(4)	0.9(1)			4.8(5)
Ser	0.9(1)	0.9(1)				1.9(2)	
Glu			1.0(1)			1.1(1)	
Pro		1.9(2)				2.0(2)	
Gly	1.1(1)	1.0(1)			1.9(2)	2.1(2)	2.0(2)
Ala	1.0(1)	1.0(1)			1.0(1)	2.0(2)	1.0(1)
Val		2.5(4)	1.4(2)			3.2(4)	1.6(2)
Ile	1.0(1)					1.0(1)	
Leu	1.0(1)			1.0(1)		1.0(1)	1.0(1)
Phe		0.9(1)				0.9(1)	
His	0.9(1)		3.0(3)			0.9(1)	2.9(3)
Lys				0.9(1)			0.7(1)
Arg					1.9(2)		1.9(2)
Hse/Hsl					+ (1)		+(1)
N-terminal residue	Ala	Val	Glx	Lys	Arg	Ala	His

TABLE XII

Amino Acid Composition of Fragments Obtained after Subdigestion of B 7

The composition from sequence analysis of each peptide is indicated by the numbers in parentheses

Peptide	S 1	S 2	S 3	S 3a	S 4	S 5	S 6	S 7	S 8	S 9	S 10
residue nos.	266-282	283-288	289-322	313-322	323-325 447-449	326-343	344-359	360-373	374-377	378-403	404-414
cmCys						0.9(1)				2.7(3)	
Asp	1.0(1)	1.0(1)	2.1(2)			1.1(1)	2.0(2)	2.0(2)		3.1(3)	
Thr	0.7(1)		0.9(1)	1.0(1)			0.9(1)	1.0(1)		1.0(1)	1.6(2)
Ser	1.1(1)		1.7(2)			0.9(1)	0.9(1)	0.9(1)		2.5(3)	0.8(1)
Glu	1.1(1)	1.1(1)	2.8(3)	1.8(2)	1.0(1)	2.6(3)		1.1(1)	1.0(1)		1.1(1)
Pro	1.0(1)		2.4(3)	0.6(1)						1.6(2)	0.8(1)
Gly	2.0(2)		3.8(4)				3.0(3)	3.0(3)		3.0(3)	1.0(1)
Ala			6.0(6)	2.0(2)		4.0(4)		1.2(1)		3.1(3)	1.0(1)
Val	1.6(2)		2.4(3)	0.7(1)		2.2(3)	0.6(1)	1.0(1)	1.0(1)		
Ile	0.7(1)	0.9(1)	1.6(2)				1.6(2)			0.9(1)	
Leu		1.7(2)	2.6(3)	0.8(1)		1.5(2)	2.8(3)	0.9(1)	0.9(1)	2.5(3)	2.9(3)
Tyr	0.7(1)	0.8(1)				1.0(1)	0.9(1)				
Phe	0.7(1)		0.8(1)		0.8(1)						
His			1.5(2)				0.9(1)				
Lys	1.9(2)		0.9(1)	1.1(1)	0.8(1)		0.9(1)	1.0(1)	0.9(1)	1.7(2)	
Arg	2.5(3)					1.9(2)		1.8(2)		1.7(2)	0.9(1)
Hse/Hsl			+(1)	+(1)							
N-terminal residue	Ile	Ile	Ser	Val	Phe	Tyr	Leu	Val	Lys	Ala	Thr

Peptide	S 11	S 12	S 13	S 14	S 15	S 16	C 1	C 2	C 3	C 4
residue nos.	415-429	430-446	450-455	456-467	468-471	470-483	274-285	286-290	305-315	318-326
cmCys										
Asp	1.0(1)	1.9(2)	1.0(1)				1.0(1)	0.9(1)	1.0(1)	
Thr		2.6(3)								0.9(1)
Ser					0.9(1)	1.0(1)	0.7(1)	0.9(1)		
Glu	2.0(2)	2.7(3)	1.1(1)	2.8(3)	1.0(1)		1.0(1)	1.0(1)		2.0(2)
Pro		0.7(1)				2.5(3)	1.0(1)			0.9(1)
Gly	0.9(1)	1.0(1)	1.0(1)			2.0(2)	1.1(1)		1.1(1)	
Ala	1.2(1)	1.0(1)	1.1(1)	2.0(2)		2.2(2)			4.0(4)	1.1(1)
Val	1.0(1)	0.8(1)		1.0(1)	0.9(1)		0.9(1)		1.8(2)	
Ile	1.7(2)	0.8(1)					0.9(1)		1.0(1)	
Leu		1.6(2)	0.9(1)	1.0(1)		3.0(3)	1.1(1)	2.0(2)	1.0(1)	
Tyr							0.8(1)			0.8(1)
Phe	1.6(2)				0.8(1)	2.9(3)				0.8(1)
His	1.7(2)			0.9(1)					0.9(1)	
Lys	1.9(2)	1.0(1)	0.8(1)	0.9(1)			1.6(2)			0.9(1)
Arg	0.9(1)	0.8(1)		2.9(3)						
Hse/Hsl										+(1)
N-terminal residue	Lys	Leu	Lys	Lys	Val	Ser	Ser	Asx	Asx	Ala

Peptide	C 5	C 6	C 7	C 8	C 9	C 10	C 11	C 12	C 13	C 14
residue nos.	338-346	364-376	377-384	403-407	409-417	424-430	445-447	448-458	465-471	472-483
cmCys			1.8(2)							
Asp		1.0(1)	0.9(1)		1.0(1)			1.0(1)		
Thr		0.8(1)		1.0(1)						
Ser	0.8(1)	0.8(1)	1.1(1)		0.8(1)				0.8(1)	
Glu	1.0(1)	1.0(1)	1.0(1)		1.0(1)	1.0(1)	1.0(1)	2.8(3)	3.0(3)	
Pro				0.9(1)						3.0(3)
Gly	1.0(1)	2.7(3)		1.0(1)	1.0(1)	1.0(1)		1.0(1)		2.0(2)
Ala	1.8(2)	0.9(1)	2.0(2)	1.0(1)				0.9(1)		1.8(2)
Val	0.8(1)	0.7(1)							0.9(1)	
Ile			0.7(1)			1.8(2)				
Leu	1.7(2)	0.7(1)		1.0(1)	1.7(2)	1.0(1)		0.9(1)		2.8(3)
Tyr	0.8(1)									
Phe					0.8(1)		0.9(1)		0.9(1)	1.8(2)
His						0.8(1)		0.8(1)		
Lys		1.7(2)			0.9(1)		1.0(1)	2.7(3)		
Arg		0.9(1)			1.0(1)	0.9(1)			1.1(1)	
Hse/Hsl										
N-terminal residue	Ser	Ser	Glx	Gly	Ser	Ile	Lys	Lys	Arg	Ala

TABLE XIII

Automated Edman Degradation of Cyanogen Bromide Peptides

Cycle No.	B 3 (1.3 nmol)		B 6 (1.5 nmol)		B7 S9 (1.5 nmol)		B7 S11 (1.2 nmol)	
	PTH-aa	Yield (pmol)	PTH-aa	Yield (pmol)	PTH-aa	Yield (pmol)	PTH-aa	Yield (pmol)
1	Gly	454	Ala	611	Ala	493	Lys	343
2	Leu	508	His	213	cmCys	210	Asp	188
3	Asp	439	Ile	504	Ser	285	Phe	303
4	Leu	521	Ser	273	Ile	195	Gln	218
5	Pro	449	Gly	330	Ala	345	Lys	203
6	Asp	378	Leu	232	cmCys	170	Val	183
7	Gly	301	Val	256	Asn	238	Ala	268
8	Gly	407	Val	376	Lys	158	His	88
9	His	58	Ala	325	Asn	178	Phe	200
10	Leu	266	Gly	245	Thr	193	Ile	143
11	Thr	40	Val	253	cmCys	90	His	63
12	His	43	Val	359	Pro	103	Arg	95
13	Gly	113	Pro	121	Gly	113	Gly	155
14	Phe	69	Ser	69	Asp	83	Ile	122
15	Hse	23	Pro	99	Lys	88	Glu	40
16	Thr	26	Phe	64	Ser	25		
17	Asp	45	Glu	49	Ala	98		
18	Lys	41	His	NQ[a]	Leu	65		
19	Lys	61	cmCys	12	Arg	43		
20	Lys	71	His	25	Pro	55		
21	Ile	31	Val	43	Ser	35		
22	Ser	24	Val	81	Gly	45		
23	Ala	18	Thr	6	Leu	50		
24			Thr	8	Arg	19		
25			Thr	11	Leu	43		
26			Thr	13	Gly	20		
27			X[a]					
28			Lys	NQ[a]				
Repetitive yield %	89 (Leu 4-Leu 10)		92 (Ala 1-Ala 9)		91 (Ala 1-Ala 5)		93 (Phe 3-Phe 9)	
Position	139-169		229-265		378-403		415-429	

[a]X: not identified; NQ: observed but not quantitated.

TABLE XIV

Automated Edman Degradation of Peptide B 4

(1.5 nmol)

cycle	PTH-aa	yield (pmoles)	cycle	PTH-aa	yield (pmoles)
1	Ala	883	27	Leu	184
2	Tyr	720	28	Ile	122
3	Lys	800	29	Ile	204
4	Val	752	30	Ala	107
5	Asn	753	31	Gly	74
6	Pro	727	32	X[a]	
7	Asp	538	33	Ser	21
8	Thr	88	34	cmCys	10
9	Gly	501	35	Tyr	39
10	Tyr	422	36	Ser	12
11	Ile	448	37	Arg	30
12	Asp	502	38	Asn	27
13	Tyr	347	39	Leu	26
14	Asp	395	40	Asp	24
15	Arg	289	41	Tyr	22
16	Leu	237	42	Gly	27
17	Glu	198	43	X[a]	
18	Glu	307	44	Leu	24
19	Asn	182	45	Arg	22
20	Ala	168	46	Lys	NQ[a]
21	Arg	163	47	Ile	17
22	Leu	176	48	Ala	18
23	Phe	192	49	Asp	15
24	His	98	50	Glu	12
25	Pro	168	51	Asn	9
26	Lys	68	52	Gly	5

Repetitive yield: 94% (Tyr 2 - Tyr 10)
Position 170-225

[a]X: not identified; NQ: observed but not quantitated.

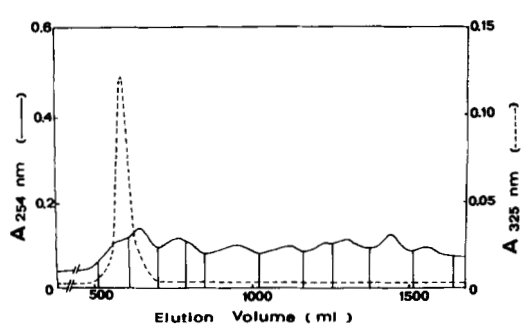


Fig. 3. Elution profile of the tryptic digest on Sephadex G-25. Flow-rate was 20 ml/h. Fractions were pooled as indicated by vertical lines.

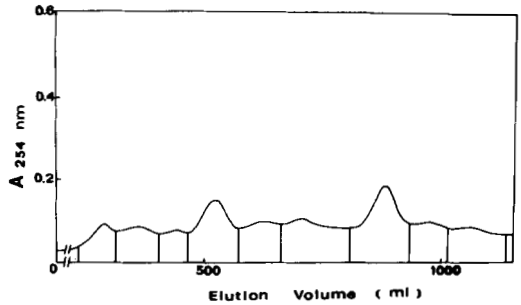


Fig. 4. Elution profile of the CNBr digest on Sephadex G-50. Flow-rate was 12.5 ml/h. Fractions were pooled as indicated by vertical lines.

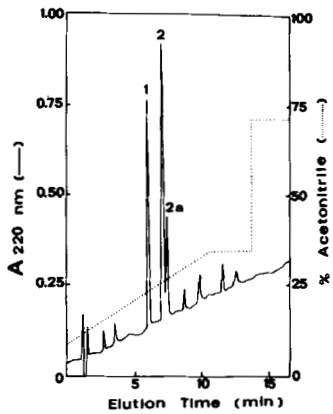


Fig. 5. Reverse-phase hplc of the acidic peptides obtained from a trypsin-carboxypeptidase B digest of the protein. Conditions of analysis are reported in the text. The number above the peaks refer to the peptides whose analytical data are given in the text.

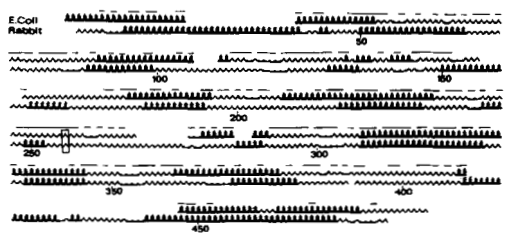


Fig. 6. Comparison of the predicted secondary structures of E. coli and rabbit serine hydroxymethyltransferases (SHMT) performed according to Garnier et al. (21). Gaps were introduced according to the alignment shown in Fig. 2. Horizontal lines above the predictions indicate segments with identical sequence in the two proteins; the box indicates the active-site lysine. : alpha-helix; : extended conformation; : turn; : random coil.

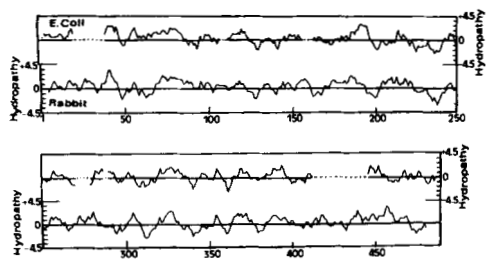


Fig. 7. Comparison of the hydropathic profiles of E. coli and rabbit serine hydroxymethyltransferases (SHMT) calculated according to Kyte and Doolittle (29). Gaps (- - -) were introduced according to the alignment shown in Fig. 2. Consecutive hydropathy values are plotted at the mid-point of the seven-residues segment as it advances from N- to C-terminus.