# AMINO ACID SEQUENCE OF HUMAN ERYTHROCYTE CARBONIC ANHYDRASE B

B. Andersson, P.O. Nyman, and L. Strid
Department of Biochemistry, University of Göteborg and Chalmers Institute of Technology, Fack, S-402 20 Göteborg 5, Sweden

**Received June 28,1972**

SUMMARY. The amino acid sequence of human carbonic anhydrase B, the major form of the enzyme in human red cells, has been investigated using a strategy based upon restriction of tryptic hydrolysis to arginyl peptide bonds by reversible blocking of lysine side chains by amidination. A proposal is made for the complete sequence of the protein which consists of 260 amino acid residues in a single polypeptide chain devoid of disulfide bridges. From X-ray diffraction studies on a homologous form of the enzyme three histidine residues in the central portion of the polypeptide chain have been suggested as ligands to the zinc ion in the active site. The results obtained here provide support for this suggestion, at least for two of these ligands.

The enzyme carbonic anhydrase catalyzes the reversible hydration of $CO_2$. The most studied forms of the enzyme are those from mammalian sources, in particular human red cells (1). Two distinct forms of the enzyme, usually called B and C, occur in human red cells. Both have molecular weights around 30 000 and contain a firmly bound zinc ion but differ in catalytic properties and amino acid sequence. Enzyme B is most abundant, occurring in about five times the amount of form C. Form C is a better catalyst in the $CO_2$ reaction with an efficiency among the highest encountered in enzymology. Both forms are under investigation by X-ray diffraction and amino acid sequence determination. Recently, a three-dimensional electron-density distribution to 2 Å resolution was reported for enzyme C (2), while the studies of the B form have not yet progressed to that stage (3). So far, the complete amino acid sequence has not been reported for any carbonic anhydrase. Due to its accessibility the human B enzyme has been the most popular form in primary structure studies and has been under investigation in several laboratories (4-9). In this paper, its complete sequence is presented.

## MATERIALS AND METHODS

Human carbonic anhydrase B was prepared as described elsewhere (10,11).

The lysine side chains were amidinated using methyl acetimidate and the product was hydrolyzed with trypsin as previously described (12). The amidin groups were removed according to Ludwig and Byrne (13) by treatment with ammonia: acetic acid at pH 11.3. The amidinated tryptic fragment containing the single sulfhydryl group of the protein was reduced with mercaptoethanol and treated with maleic anhydride which was found to block irreversibly the cysteine side chain. Hydrolysis of human carbonic anhydrase B with chymotrypsin and thermolysin and the digestions of peptides with these and other enzymes such as papain and subtilisin (Nagarse) were performed with buffers and enzyme: substrate ratios commonly used in primary-structure investigations (14).

Gel filtrations on Sephadexes were used for peptide separations, in particular for the large fragments. The runs were carried out in 1 M acetic acid often containing 2 % dodecylamine. $R_f$-values for peptides were often markedly changed by the presence of this detergent. High-voltage paper electrophoreses were run according to Smillie and Hartley (15). Amide groups were determined from electrophoretic mobility or by amino acid analyses of carboxypeptidase or aminopeptidase digests.

Sequential degradations of peptides from the amino-terminal end were carried out by dansyl-Edman, Edman, and Sanger procedures, and by digestions with leucine aminopeptidase or aminopeptidase M. Carboxyl-terminal analyses were made by digestions with carboxypeptidases A or B, or the more recently introduced carboxypeptidase C (16).

## RESULTS AND DISCUSSION

The results obtained are summarized in Figs. 1 and 2. Fig. 1 shows the amino acid sequence proposed for human carbonic anhydrase B and tries to indicate the essential features of the strategy used in the sequence determination.

Tryptic hydrolysis of protein having the lysine side chains masked by amidin groups leads to a restriction of the tryptic hydrolysis to arginyl

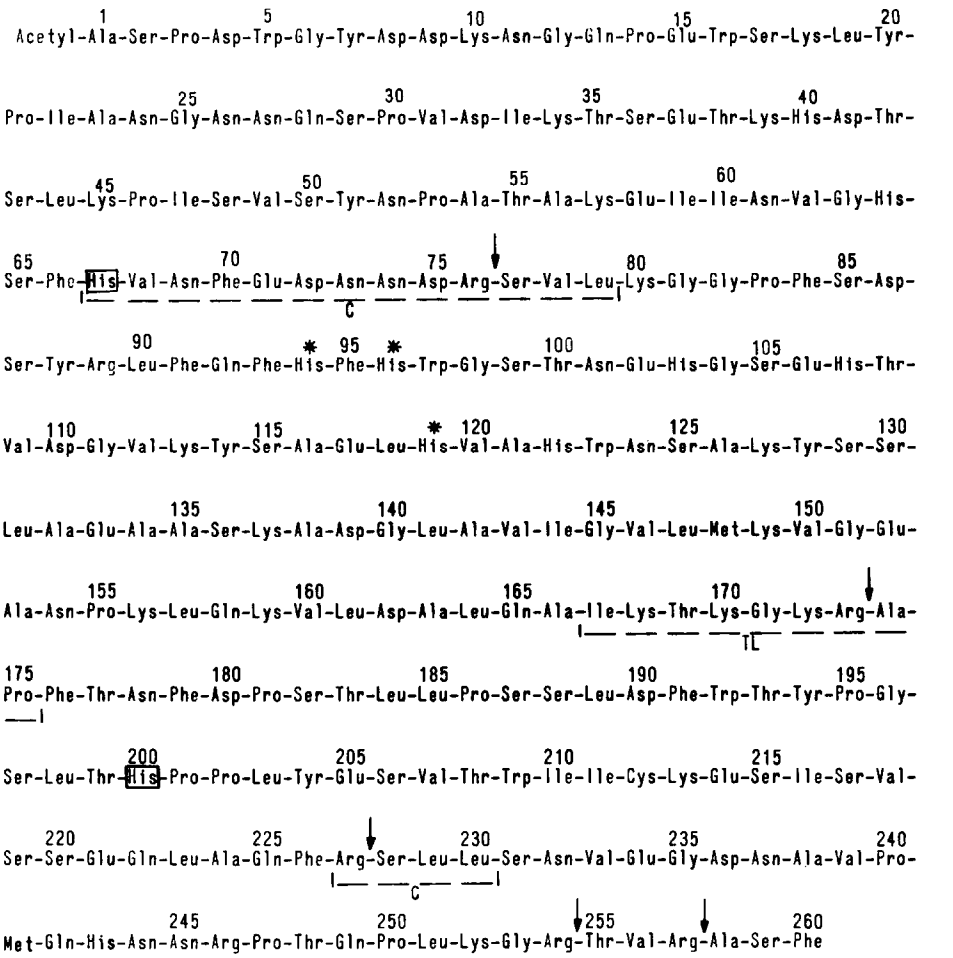


Fig. 1. Amino acid sequence proposed for human erythrocyte carbonic anhydrase B. The sequence has been numbered starting from the acetylated amino-terminal end of the polypeptide chain. Cleavages of arginyl peptide bonds by tryptic hydrolysis of the protein after blocking of lysine side chains by amidination have been indicated by arrows. The fragments obtained were arranged in order using the arginine-containing bridge peptides, shown in the diagram, which were isolated from thermolytic (TL) and chymotryptic (C) digests of the protein. The region from position 228 to 260 has been investigated earlier (10, 12). The sequencing of the three large tryptic fragments from amidinated enzyme covering the stretch from position 1 to 227 is illustrated in Fig. 2. Residues 67 and 200 are histidines modifiable with active-site directed reagents (9,21, 22). The three histidine residues at positions 94, 96, and 119 are candidates for the ligands binding the zinc ion in the active site as suggested from high-resolution X-ray studies on a homologous form of the enzyme (2). Tyrosine residues modifiable with tetranitromethane are located at positions 20, 88, and 114 (8).

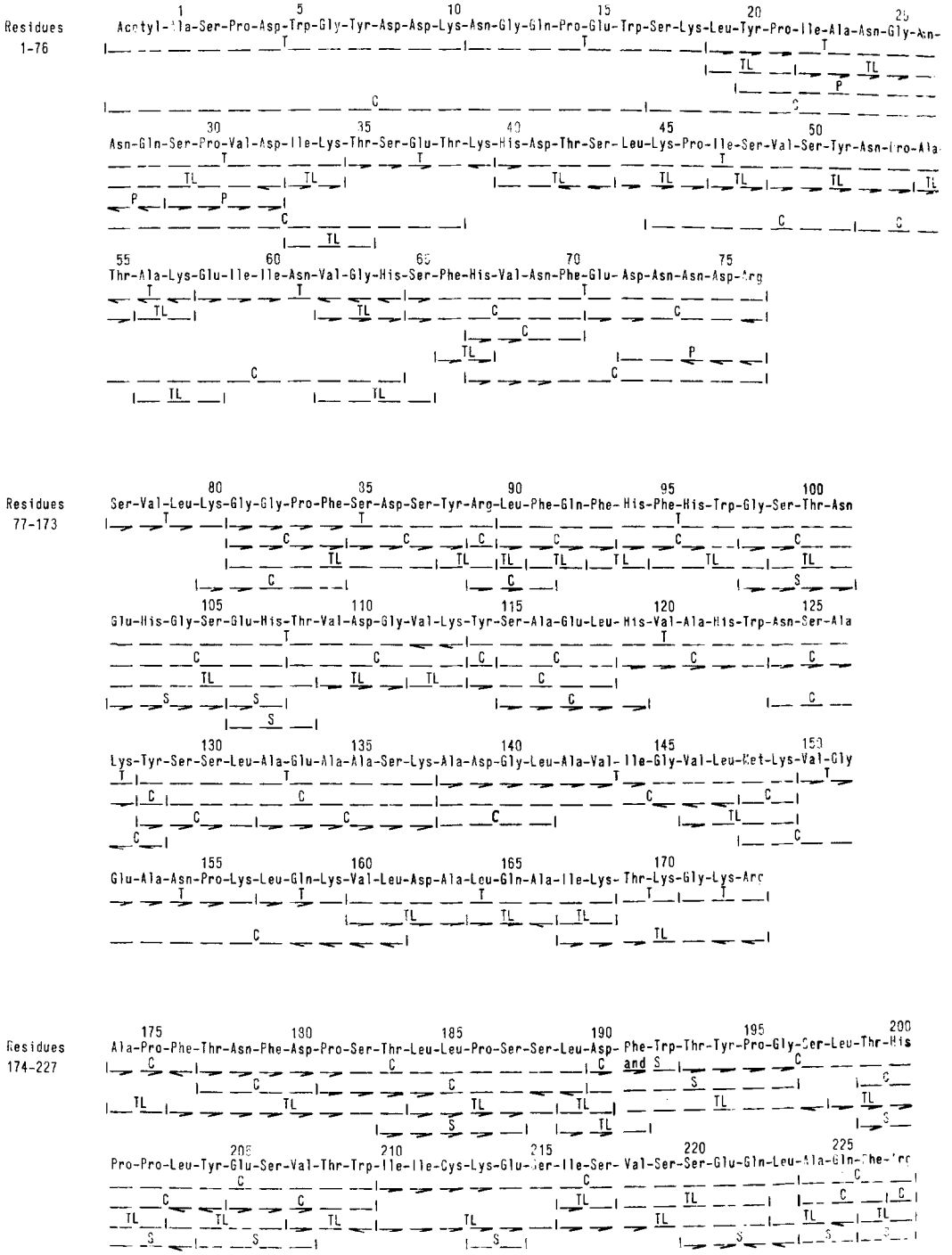


Fig. 2. See legend on following page.

peptide bonds as indicated by arrows in Fig. 1. The protein molecule contains in total seven arginine residues. Three of these are located close together in the carboxyl-terminal region of the protein. The sequence of this portion of the polypeptide chain, namely from residue number 228 to 260 in Fig. 1, has been investigated earlier (10,12,6). Of the four arginine residues located in the remaining portion of the protein, three were found to be readily hydrolyzed with trypsin, namely the bonds at positions 76, 173, and 227 as indicated in Fig. 1. The cleavages of these bonds produce three large fragments containing respectively 76, 97, and 54 amino acid residues. The sequence determination of these fragments is outlined in Fig. 2. By the help of "overlap" peptides bridging them together, the fragments were then arranged in order to give the complete sequence for the protein shown in Fig. 1. The positions in the sequence of these "overlap" peptides are indicated in Fig. 1.

The strategy described above is in principle simple and straight-forward. The investigation clearly demonstrates the usefulness of the amidination reaction for the reversible masking of the lysine side chains in order to limit tryptic digestion. The approach involves the purification of large peptides from the tryptic digest of amidination protein and a considerable part of the total effort was spent on that.

The sequence proposed in Fig. 1 corresponds to the following amino acid composition: $Lys_{18}His_{11}Arg_{7}Trp_{6}Asp_{14}Asn_{17}Thr_{14}Ser_{30}Glu_{13}Gln_{9}Pro_{17}Gly_{16}Ala_{19}$

Fig. 2. Diagram showing the amino acid sequence studies on three large tryptic fragments from amidinated human carbonic anhydrase B. The investigations were carried out by the isolation and characterization of smaller peptides from digests obtained with trypsin (T), chymotrypsin (C), thermolysin (TL), papain (P), or subtilisin (S) of amidinated or deamidinated pieces. Broken lines indicate quantitative amino acid analysis. ⟵ ⟶ denote sequences determined by amino-terminal analysis using Edman or dansyl-Edman degradation, dinitrophenylation or aminopeptidase digestion. ⟵ ⟵ indicate sequences investigated by digestions with carboxypeptidases A, B, or C. The amino-terminal sequence from position 1 to 18 has previously been determined by Laurent et al. (23) and was not reinvestigated here. Compositions of tryptic and chymotryptic peptides from that region obtained here support the results of Laurent et al.

$Val_{17}Met_2Ile_{10}Leu_{20}Tyr_8Phe_{11}Cys_1$. It is identical with the composition determined previously for the enzyme (17), with the exception of Glu+Gln and Pro which differ by 2 and 1 residues respectively. The formal molecular weight, including the residue weights of 260 amino acid residues + 1 water molecule + 1 acetyl group + 1 zinc ion, is around 28 850 and the protein molecule should contain 353 nitrogen atoms corresponding to a nitrogen content of 17.14 %.

The results obtained here can be compared with the three-dimensional model for human carbonic anhydrase C derived from X-ray diffraction data to 2 Å resolution. In fact, the attempts to construct that model (2) have been made with the aid of much of the sequence information presented here.

The active site of enzyme C has been shown to consist of a crevice in the globular protein molecule. The zinc ion which is believed to partake in the catalytic reaction is located at the bottom of this cleft, bound to the protein molecule by three ligands in an approximately tetrahedral coordination. These ligands have in the X-ray model been suggested to be histidine side chains occupying positions 94, 96, and 118 in the carbonic anhydrase sequence[x]. Inspection of the B enzyme sequence in Fig. 1 shows that histidine residues occupy positions 94 and 96. This finding would support the idea that residues 94 and 96 are ligands to the metal ion. Position 118, proposed by the X-ray work for the third metal ligand, is in the B enzyme occupied by a leucine residue (see Fig. 1). This leucine residue is, however, adjacent to a histidine and a glutamic acid residue at positions 119 and 117 respectively. His-119 would be a candidate for the third metal ligand, but the possibility that Glu-117 instead partakes in the zinc binding has not yet been excluded. The latter alternative would mean a metal binding similar to that in carboxy-

x) The numbering of residues in enzyme C used by the crystallographers differs from that used here. Residues 64, 67, 94, 96, 118, and 129 in the present paper correspond to positions 63, 66, 93, 95, 117, and 128 in the discussion of Liljas *et al.* (2). A common numbering would be desirable and we would like to suggest that provisionally the B enzyme numbering also be used for other carbonic anhydrases.

peptidase where one glutamic acid and two histidine side chains act as ligands (18). In carbonic anhydrase as well as in carboxypeptidase two of the ligands are located close together in the sequence suggesting a role for the metal ion in the folding of the polypeptide chain into the native conformation (cf. ref. 19). Electron spin resonance studies on human carbonic anhydrase B having the $Zn^{2+}$ ion replaced by a $^{63}Cu^{2+}$ ion give results consistent with the presence of two nitrogen ligands (20). Spectroscopic studies of the metal binding in the human carbonic anhydrases where the $Zn^{2+}$ ion has been replaced by a $Co^{2+}$ ion point towards a similarity in metal coordination (1), which would imply that the nature of the metal ligands may well be identical in the two proteins.

By chemical modification two histidine residues in the active site of enzyme B can be specifically labelled (1). Primary structure studies on modified enzyme (21,22) show that N-chloroacetylchlorothiazide reacts with His-67 and bromoacetate and other halo acids alkylate His-200. According to the stretching of the polypeptide chain proposed from the X-ray studies on human enzyme C, the regions around positions 67 and 200 should both be elements of the active-site region. Our finding supports the validity of the proposed X-ray model (2) and would furthermore indicate that the conformations of human enzymes B and C may be closely similar. The X-ray studies suggest two additional histidine residues in the region of the active site. They are believed to be associated by hydrogen bonding to a cluster of ordered water molecules of unknown functional significance in the active-site crevice. One of the residues, His-64, is also present in the B enzyme sequence as shown in Fig. 1. However, the other one, ascribed to position 129, is absent. For the B enzyme it would be tempting to believe that the adjacent residue, Tyr-128, plays the corresponding role. Determination of the amino acid sequence of human enzyme C, under way in this laboratory, will identify the side chains in this and other regions thus providing a basis for more conclusive comparisons between the sequences and X-ray structures of the isoenzymes.

ACKNOWLEDGEMENTS. We are indebted to numerous persons for support and technical help. The investigation has been aided by the Swedish Medical and Natural Science Research Councils, the Swedish Board of Technical Development, and the Institute of General Medical Science, U.S. Public Health Service (GM 12280-06).

REFERENCES

1. Lindskog, S., Henderson, L.E., Kannan, K,K,, Liljas, A., Nyman, P.O., and Strandberg, B., in The Enzymes, 3rd ed., Vol. 5 (P.D. Boyer, Ed.), Academic Press, New York and London 1971, p. 587.
2. Liljas, A., Kannan, K.K., Bergstén, P.C., Waara, I., Fridborg, K., Strandberg, B., Carlbom, V., Järup, L., Lövgren, S., and Petef, M., Nature New Biology 235, 131 (1972).
3. Kannan, K.K., Fridborg, K., Bergstén, P.C., Liljas, A., Lövgren, S., Petef, M., Strandberg, B., Waara, I., Adler, L., Falkbring, S.O., Göthe, P.O., and Nyman, P.O., J. Mol. Biol. 63, 601 (1972).
4. Henderson, L.E., Henriksson, D., Nyman, P.O., and Strid, L., in Oxygen Affinity of Hemoglobin and Red Cell Acid-Base Status (M. Rørth and P. Astrup, Eds.), Alfred Benzon Symposium IV (1971), Munksgaard, Copenhagen, and Academic Press, New York 1972, p. 341.
5. Sciaky, M., Limozin, N., Giraud, N., Marriq, C., Charrel, M., and Laurent, G., Biochim. Biophys. Acta 221, 665 (1970).
6. Funakoshi, S. and Deutsch, H.F., J. Biol. Chem. 245, 4913 (1970).
7. Tashian, R.E. and Stroup, S.R., Biochem. Biophys. Res. Commun. 41, 1457 (1970).
8. Dorner, F., J. Biol. Chem. 246, 5896 (1971).
9. Bradbury, S.L., J. Biol. Chem. 244, 2010 (1969).
10. Nyman, P.O., Strid, L., and Westermark, G., European J. Biochem. 6, 172 (1968).
11. Henderson, L.E. and Henriksson, D., in manuscript.
12. Andersson, B., Göthe, P.O., Nilsson, T., Nyman, P.O., and Strid, L., European J. Biochem. 6, 190 (1968).
13. Ludwig, M.L. and Byrne, R., J. Am. Chem. Soc. 84, 4160 (1962).
14. Smyth, D.G., in Methods in Enzymology, Vol. 11 (C.H.W. Hirs, Ed.) Academic Press, New York and London 1967, p. 214.
15. Smillie, L.B. and Hartley, B.S., Biochem. J. 101, 232 (1966).
16. Tschesche, H. and Kupfer, S., European J. Biochem. 26, 33 (1972).
17. Nyman, P.O. and Lindskog, S., Biochim. Biophys. Acta 85, 141 (1964).
18. Hartsuck, J.A. and Lipscomb, W.N., in The Enzymes, 3rd ed., Vol. 3 (P.D. Boyer, Ed.), Academic Press. New York and London 1971, p. 1.
19. Yazgan, A. and Henkens, R.W., Biochemistry 11, 1314 (1972).
20. Taylor, J.S. and Coleman, J.E., J. Biol. Chem. 246, 7058 (1971); cf. Proc. Nat. Acad. Sci. US 69, 859 (1972).
21. Andersson, B., Nyman, P.O., and Strid, L., in $CO_2$: Chemical, Biochemical, and Physiological Aspects (R.E. Forster, J.T. Edsall, A.B. Otis, and F.J.W. Roughton, Eds.), National Aeronautics and Space Administration, NASA SP-188, Washington, D.C. 1969, p. 109.
22. Göthe, P.O. and Nyman, P.O., FEBS Letters 21, 159 (1972).
23. Laurent, G., Marriq, C., Garcon, D., Luccioni, F., and Derrien, Y., Bull. Soc. Chim. Biol. 49, 1035 (1967).