Biochimica et Biophysica Acta, 371 (1974) 534–548

BBA 36880

AMINO ACID SEQUENCE OF SHEEP CARBONIC ANHYDRASE C

ROBERT J. TANIS, ROBERT E. FERRELL and RICHARD E. TASHIAN

Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Mich. 48104 (U.S.A.)

(Received June 7th, 1974)

SUMMARY

The sequence of amino acid residues comprising the major form of sheep red cell carbonic anhydrase C has been determined. The primary sequences of peptides derived from cyanogen bromide cleavage and tryptic digestion were obtained primarily through the use of the Edman degradation procedure. The ordering of these peptides in the sheep molecule is based on the high degree of homology between the sheep enzyme and the previously sequenced human and bovine carbonic anhydrase C molecules. Based on comparisons with the three-dimensional structure of human carbonic anhydrase C, the function of certain residues which appear to be involved either in the maintenance of structure or in the active site of the sheep enzyme is discussed.

INTRODUCTION

Previous work from this laboratory has shown that sheep erythrocytes contain only a single major form of carbonic anhydrase (carbonate hydro-lyase, EC 4.2.1.1), whereas erythrocytes from most mammalian species contain two distinct isozymes of carbonic anhydrase (designated CA B or CA I and CA C or CA II) known to be the products of independent genetic loci [1, 2]. Furthermore, on the basis of kinetic properties and the amino acid sequence of the C-terminal portion of the molecule, it was postulated that the sheep enzyme was homologous to the high activity C form observed to occur in other mammals [1].

The complete primary sequences of the human B and C forms have been reported [3–6]. Further, a considerable portion of the sequence of bovine CA C has also been determined [2, 7, 8]. Complimenting these sequence efforts has been the X-ray analysis of the crystal structure of the human C and B forms which have now been resolved to the 2.0- and 2.3-Å levels, respectively [9, 10, 20]. Because of the availability of this type of comparative data from the human enzymes, it was of interest to determine the amino acid sequence of the sheep red cell enzyme in order to ascertain what features it might share in common with the human carbonic anhydrase C both structurally and functionally.

Abbreviations: F_3Ac, trifluoroacetic acid; MalNEt, *N*-ethylmorpholine.

Part of our previous report on the sheep enzyme dealt with the complete amino acid sequence of the penultimate and C-terminal CNBr fragments, and provided the first primary sequence data for the sheep enzyme. That study indicated that CNBr cleavage at each of the three methionyl residues offered a promising method of attacking the entire molecule. Continuation of these studies, reported here, involved the preparation and purification of the two N-terminal CNBr fragments as well as the determination of their primary sequences. Together, these four CNBr fragments satisfactorily account for the amino acid composition of sheep carbonic anhydrase C.

MATERIALS AND METHODS

Enzyme preparation

Sheep red cell carbonic anhydrase was purified according to a previously reported procedure [1]. After selective removal of hemoglobin, the enzyme was isolated using DEAE-cellulose and DEAE-Sephadex column chromatography procedures. Electrophoresis on starch gel was employed to confirm that the final preparation was essentially homogeneous. The previously reported $A_{280\ nm}$ value of 16.1 was used to determine the protein concentration of the final preparation.

Reaction with CNBr

As previously reported [1], the sheep enzyme contains three methionyl residues all of which react quantitatively with CNBr. Using the amino acid analyzer to measure methionine destruction, we observed that the CNBr reaction in 70% trifluoroacetic acid (F_3Ac) was complete within 12 h and therefore this time was routinely employed. Upon dilution of the reaction mixture, to 5% F_3Ac, one insoluble and three soluble fragments were obtained.

Isolation of CNBr fragments

The following designations have been applied to the CNBr fragments: N-terminal fragment, F_1 (Residues 1–58); the large insoluble fragment, F_2 (59–221); penultimate fragment, F_3 (222–239), and C-terminal fragment, F_4 (240–259).

The three 5% F_3Ac-soluble CNBr fragments (F_1, F_3 and F_4) were separated from the single insoluble fragment by centrifugation, and the remaining F_3Ac was removed from the former by lyophilization. To separate the F_3Ac-soluble fragments, the lyophilized material was taken up in 0.17 M pyridine–acetate buffer (pH 4.7), and applied directly to a 1.5 cm × 30 cm column of SE- or SP-Sephadex equilibrated with the 0.17 M buffer. A 150-ml linear gradient (0.17–2.0 M pyridine–acetate) was used to elute the column, and 2-ml fractions were collected. Individual fractions were pooled on the basis of ninhydrin-positive material. Routinely, three major peaks were obtained corresponding to the three soluble CNBr fragments. To confirm the purity of these fragments, high-voltage paper electrophoresis (Whatman 3 MM) was carried out in pyridine–acetate–water (100:4:900, by vol.) (pH 6.5); the N-terminal fragment F_1, identified with both ninhydrin and the Pauli stain for histidine, appeared pure by this isolation method and no further purification was undertaken.

Because the F_1 fragment was blocked at the N-terminus, it was necessary to cleave this peptide into fragments which could then be sequenced directly. One approach was the limited digestion with trypsin (2 h, 37 °C, 1% TPCK-treated

trypsin) followed by separation of the tryptic peptides on a 1.5 cm × 30 cm SP-Sephadex column using a 280 ml 0.17–4 M pyridine–acetate gradient.

Two procedures were used during the course of this work to further purify the 5% F_3Ac insoluble CNBr fragment F_2. The first technique involved resuspending the insoluble CNBr fragment in 25% acetic acid; the residual amounts of the soluble fragments were then removed by passing the suspension through a 0.9 cm × 90 cm Sephadex G-100 column equilibrated with 25% acetic acid. A single major peak emerged near the void volume of such a column, and this material corresponded to the large insoluble CNBr fragment. During the latter part of this work it was found that the residual amounts of soluble peptide could be removed by washing the 5% F_3Ac pellet 3 times with distilled water. The washed pellet was soluble in 5% N-ethyl morpholine (MalNEt), and the solution could be adjusted to pH 8.5 with 6 M HCl. The latter procedure was found to be the more useful because it was rapid and the MalNEt buffer (pH 8.5) was, after dilution, compatible with subsequence strategies involving the use of lysine blockage and/or treatment with trypsin.

Sequence determination

The Beckman Model 890 automatic sequencer was used to obtain the sequences reported below. Details of the procedures employed have been previously reported [1]. During the course of this work, the Braunitzer reagent [11] (4-sulfophenylisothiocyanate) became commercially available (Pierce Chemical Co.) and was employed during the sequencing of peptides containing a C-terminal lysine. Routinely, a suitable amount of peptide was dissolved in 0.2 ml dimethylallylamine to which was added 0.1 ml of a 5% solution (in water) of Braunitzer reagent. After flushing with N_2, the mixture was incubated at 50 °C for 1 h and then sequenced without further treatment. An unreacted portion of the peptide was used to determine the N-terminal residue.

In all cases, duplicate or triplicate sequencer runs were performed on the peptides reported below. Phenylthiohydantoin derivative yields, as determined with the gas chromatograph, ranged from a high of 75% to a low of 2–3%. Arginine yields were not quantitated during this work.

Tryptic peptide separation of the F_2 fragment

To facilitate sequence determination of the large insoluble CNBr fragment, F_2, it was found useful to treat the material with 1% TPCK-trypsin (2 h, 37 °C) and isolate the resulting tryptic peptides. A 0.6 cm × 100 cm column of chromobeads (Technicon type P), maintained at 50 °C, was employed for such peptide separations. Routinely, a double linear gradient was used; the first gradient ranged from 0.05 M pyridine–acetate buffer (pH 2.4) to 0.5 M pyridine–acetate buffer (pH 3.1), while the second gradient extended from 0.5 to 2.0 M pyridine–acetate buffer (pH 5.0). Only those peptides soluble in the starting buffer were applied to such a peptide column. Approx. 85% of the column effluent was routed to a fraction collector; the remainder was mixed with ninhydrin, and the colorimetric readings were recorded using techniques similar to those reported by Jones [12]. After pooling the fractions, the pyridine–acetate buffer was removed by lyophilization. Where necessary, the peptides were desulted using a Sephadex G-15 column equilibrated with 1.0 M or 25% acetic acid.

Masking of lysine residues

Another technique, used to prepare tryptic peptides from the F_2 fragment, involved the reacting of lysine side chains with citraconic anhydride followed by tryptic cleavage of arginine bonds essentially as described by Gibbons and Perham [13], except that a 2-fold excess of citraconic anhydride was used. The blocking groups were quantitatively removed, at the appropriate time, by subjecting these peptides to low pH conditions (approx. pH 2.0).

Amino acid composition

The methods reported earlier [1] were used except that the peptides were routinely hydrolyzed for 18–24 h with redistilled 6 M HCl. No phenol was added to protect tyrosine residues, and no corrections were applied for slowly-released amino acids, or for amino acid destruction occurring during the hydrolysis.

Carboxypeptidase digestion

Carboxypeptidase A and B were obtained from the Worthington Biochemical Co.; both were treated with diisopropyl phosphofluoridate to eliminate trace tryptic and chymotryptic activity. The carboxypeptidase B was used directly, and the carboxypeptidase A was prepared according to the procedure described by Ambler [14]. The time span of a typical experiment, carried out at 37 °C, ranged from 5 min to 3 h with 7 intermediate times. The released amino acid residues were, when necessary, quantitated using a Beckman PA-35 physiological column in order to resolve the amide amino acids. All carboxypeptidase results are based on duplicate or triplicate digestion experiments.

RESULTS

Characterization of N-terminal CNBr fragment

In order to isolate the N-terminal CNBr fragment, F_1, 3 μmoles of whole enzyme were subjected to the CNBr cleavage reaction. After removal of the insoluble material and F_3Ac, the resulting fragments were purified on a SE-Sephadex column (see Materials and Methods). Three ninhydrin-positive peaks were observed, the first and third were the penultimate (F_3) and C-terminal (F_4) CNBr fragments, respectively; the primary sequences of which have been previously reported [1]. The intermediate peak proved to be the N-terminal fragment (F_1). Subsequent examination of this fragment with high-voltage paper electrophoresis indicated that no detectable impurities were present.

The amino acid composition of the N-terminal fragment is given in Table I; this fragment is comprised of 58 residues, and is presumably acetylated at the N-terminus on the basis of its failure to undergo Edman degradation in the automatic sequencer. Because of the blocked N-terminus, it was useful to subsequently digest this fragment with trypsin. Following incubation with trypsin, the tryptic peptides were separated using SE-Sephadex column chromatography. Four major ninhydrin-positive peaks were observed with the following order of elution: Tp2; TP3; Tp4; Tpl. High-voltage paper electrophoresis was performed to determine their purity; none of the peptides appeared to require further purification. The amino acid composition of each of these peptides is given in Table I, and together they account for

TABLE I

AMINO ACID COMPOSITION OF THE N-TERMINAL CNBr FRAGMENT, F_1, AND THE COMPONENT TRYPTIC PEPTIDES

The residue values reported represent the observed and nearest integers, respectively.

Amino acid	Entire F_1 fragment	Tp 1	Tp 2	Tp 3	Tp 4
Asp	6.1 (6)	1.0 (1)	3.9 (4)	1.2 (1)	
Thr	0.9 (1)		0.9 (1)		
Ser	2.6 (3)	0.8 (1)	0.9 (1)	1.0 (1)	
Glu	6.4 (6)	2.2 (2)	2.0 (2)	2.3 (2)	
Pro	6.2 (6)	1.3 (1)	2.4 (2)	3.4 (3)	
Gly	4.2 (4)	3.1 (3)	0.9 (1)		
Ala	6.0 (6)		1.1 (1)	5.2 (5)	
Val	3.3 (3)		1.1 (1)	1.8 (2)	
Met*	(1)				(1)
Ile	2.0 (2)		1.8 (2)		
Leu	4.5 (4)			3.7 (4)	
Tyr	1.9 (2)	0.9 (1)		1.0 (1)	
Phe	1.2 (1)		0.9 (1)		
Lys	3.1 (3)	1.0 (1)	1.1 (1)	1.0 (1)	
His	4.9 (5)	5.1 (5)			
Arg	3.3 (3)		0.9 (1)	0.9 (1)	0.9 (1)
Trp**	2.4 (2)	(2)			
Total	58	17	18	21	2

* Scored as homoserine.

** The values are based on extinction at 287.5 and 278.0 nm [18].

the composition of the N-terminal CNBr fragment. Tpl is the blocked N-terminal tryptic peptide of the fragment.

Further treatment of Tpl was undertaken to prepare it for sequencing. The material was treated for 35 min with 1% chymotrypsin, and the resulting peptide mixture was analyzed using high voltage paper electrophoresis. Peptides were located

TABLE II

AMINO ACID COMPOSITION OF THE CHYMOTRYPTIC PEPTIDES COMPRISING TRYPTIC PEPTIDE, Tp 3

Amino acid	Tp 3	Cp 1	Cp 2	Cp 3
Asp	1		0.9 (1)	
Ser	1	0.9 (1)		
Glu	2		2.0 (2)	
Pro	1		1.1 (1)	
Gly	3		2.8 (3)	
Tyr	1		0.8 (1)	
Lys	1			1.2 (1)
His	5	2.1 (2)	1.7 (2)	0.8 (1)
Trp	2	1*	1*	

* Based on positive reaction with Erlich's stain; confer text for details.

on the paper (Whatman 3 MM) by staining guide strips. Three peptides were observed, only two of which were ninhydrin-positive. The third, Cpl, was observed when the paper was sprayed with Ehrlich's stain (*p*-dimethylaminobenzaldehyde) which is specific for tryptophan, or with the Pauli stain for histidine; it was therefore assumed to represent the acetylated N-terminal chymotryptic peptide. The amino acid compositions of the three chymotryptic peptides, isolated from paper, are given in Table II.

To obtain the sequence of Tpl, each of the chymotryptic peptides was isolated using a 0.6 cm × 15 cm SP-Sephadex column eluted with an 80 ml 0.1–2 M pyridine–acetate buffer gradient. Two ultraviolet-positive peaks were observed, one corresponding to Cpl and the other to Cp2; the observed elution profile is presented in Fig. 1. When Cpl was treated with carboxypeptidase A, the sequence His–His–Trp

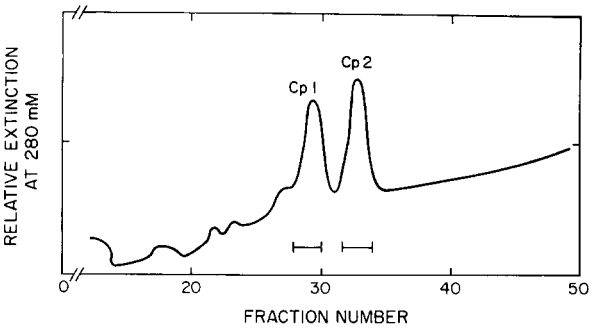


Fig. 1. Isolation of Cpl and Cp2 using a SP-Sephadex column developed with a 0.17–2 M gradient of pyridine–acetate. Fraction volumes collected were 2 ml; the horizontal bars represent pooled fractions.

was observed, leading us to conclude that the complete sequence of Cpl is Acetyl-Ser–His–His–Trp.

Cp2 was sequenced directly using the C-terminal peptide program (Beckman 092370); the results are presented in Table III. It should be pointed out that by employing limited chymotryptic digestion, minimal cleavage appears to occur on the carboxyl-side of the tyrosine at Residue Position 6.

The sequence of Cp3 is presumed to be His–Lys on the basis of trypsin specificity.

Sequence determinations of Tp2 and Tp3 were undertaken using peptides isolated by SE-Sephadex chromatography. The results are given in Table III, and in each case account completely for the known peptide amino acid composition. Tp4 is presumed to be Arg–Met based on the known specificity of the CNBr reaction.

Characterization of the CNBr fragment, F_2

The amino acid composition of the F_2 material, (Table IV), after fractionation on a column of Sephadex G-100 equilibrated to 25% acetic acid, suggested that this CNBr fragment represented the remainder of the sheep carbonic anhydrase molecule. This fragment was composed of 160 amino acid residues, and a number of strategies were employed to obtain the primary sequence.

TABLE III

AUTOMATIC SEQUENCER RESULTS OF TRYPTIC, CHYMOTRYPTIC AND CNBr PEPTIDES

Peptide	
Cp2	Gly–Tyr–Gly–Glu–His–Asn–Gly–Pro–Glu–His–Trp
Tp2	Asp–Phe–Pro–Ile–Ala–Asp–Gly–Glu–Arg–Gln–Ser–Pro–Val–Asp–Ile–Asp–Thr–Lys*
Tp3	Ala–Val–Val–Pro–Asp–Pro–Ala–Leu–Lys–Pro–Leu–Ala–Leu–Leu–Tyr–Glu–Gln–Ala–Ala–Ser–Arg*
F_2'	Val–Asn–Asn–Gly–His–Ser–Phe–Asn–Val–Glu–Phe–Asp–Asp–Ser–Gln–Asp–Lys–Ala–Val–Leu–Lys–Asp–Gly–Pro–Leu–
Tp6	Ala–Val–Leu–Lys*
Tp7	Asp–Gly–Pro–Leu–Thr–Gly–Thr–Tyr–Arg*
Tp9	Tyr–Ala–Ala–Glu–Leu–His–Leu–Val–His–Trp–Asn–Thr–Lys*
Tp8	Leu–Val–Gln–Phe–His–Phe–His–Trp–Gly–Ser–Ser–Asp–Asp–Gln–Gly–Ser–Glu–His–
Tp10	Tyr–Gly–Asp–Phe–Gly–Thr–Ala–Ala–Gln–Gln–Pro–Asp–Gly–Leu–Ala–Val–Val–Gly–Val–Phe–Leu–Lys*
Tp11	Val–Gly–Asp–Ala–Asn–Pro–Ala–Leu–Gln–Lys*
Tp12	Val–Leu–Asp–Val–Leu–Asp–Ser–Ile–Lys*
Tp15	Ser–Ala–Asp–Phe–Pro–Asn–Phe–Asn–Pro–Ser–Ser–Leu–Leu–Lys*
Tp16***	Ala–Leu–Asn–Tyr–Trp–Thr–Tyr–Pro–Gly–Ser–Leu–Thr–Asn–Pro–Pro–Leu–Leu–Glu–Ser–Val–Thr–Trp–Val–Val–Leu–Lys–Glu–Pro–X**–X**–Val–

* Not actually sequenced; residue presumed on basis of amino acid composition.
** Residue not identifiable using the automatic sequencer.
*** Includes residues beyond the C-terminus of Tp13; confer Results for details.

TABLE IV

AMINO ACID COMPOSITION OF THE F_2 CNBr FRAGMENT

The residue values reported represent the observed and nearest integers, respectively.

Amino acid	Observed	Nearest integer
Asp	20.5	21
Thr	9.5	10
Ser	13.7	14
Glu	13.8	14
Pro	10.4	10
Gly	12.3	12
Ala	10.3	10
Val	14.8	15
Met*	—	1
Ile	1.4	1
Leu	17.1	17
Tyr	4.8	5
Phe	7.6	8
Lys	12.3	12
His	6.1	6
Arg	2.8	3
Trp**	—	3
Total		162

* Scored as homoserine.
** Based on the automatic sequencer data of various peptides comprising this CNBr fragment.

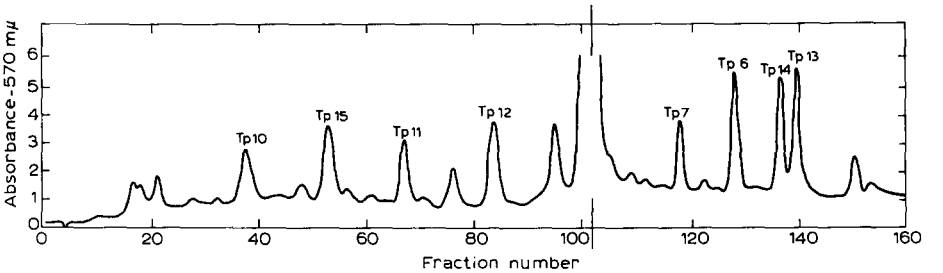


Fig. 2. Typical elution profile of sheep carbonic anhydrase C tryptic peptides separated on a column of Technicon Type P resin developed with a pyridine acetate gradient. Confer text for procedural details.

To determine the amino acid sequence of the N-terminal portion of the F_2 fragment the material was sequenced directly using the automatic sequencer; this permitted the determination of the sequence of the first 25 residues. These data are presented in Table III as sequence F_2'.

To obtain additional sequence information, it was necessary to cleave this CNBr fragment with trypsin. Tryptic peptides, soluble in 0.05 M pyridine–acetate buffer (pH 2.4), were separated by ion-exchange chromatography using pyridine–acetate gradients. Up to 2 μmoles of CNBr fragments were routinely treated in this fashion. Fig. 2 shows a typical elution profile of the soluble tryptic peptides obtained by this method. After suitable pooling, the samples were lyophylized and the amino acid compositions determined. The composition data are given in Table V; and

TABLE V

AMINO ACID COMPOSITION OF THE SOLUBLE TRYPTIC PEPTIDES FRACTIONATED ON THE PEPTIDE COLUMN

The residue values represent observed and nearest integers, respectively.

Amino acid	Tp6	Tp7	Tp10	Tp11	Tp12	Tp13	Tp14	Tp15
Asp		1.3 (1)	2.2 (2)	2.0 (2)	2.4 (2)			2.8 (3)
Thr		1.8 (2)	1.0 (1)			0.9 (1)		
Ser					0.9 (1)			2.6 (3)
Glu			2.2 (2)	1.0 (1)				
Pro		1.1 (1)	1.3 (1)	1.1 (1)				1.6 (2)
Gly		2.0 (2)	3.9 (4)	0.9 (1)			1.0 (1)	
Ala	1.0 (1)		3.2 (3)	1.8 (2)				0.9 (1)
Val	1.2 (1)		2.8 (3)*	1.0 (1)	2.0 (2)			
Ile					0.9 (1)			
Leu	1.0 (1)	1.4 (1)	2.1 (2)	0.9 (1)	1.9 (2)			2.0 (2)
Tyr		1.0 (1)	1.0 (1)					
Phe			1.5 (2)					1.7 (2)
Lys	0.9 (1)		1.0 (1)	1.0 (1)	1.0 (1)	1.3 (1)	1.1 (1)	1.0 (1)
Arg		1.0 (1)						
Residue total	4	9	22	10	9	2	2	14

* Required 72 h acid hydrolysis to obtain quantitative valine release.

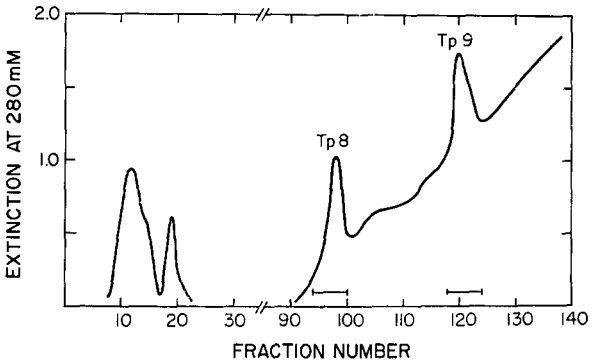


Fig. 3. Isolation of Tp8 and Tp9 using a SE-Sephadex column developed initially with 100 ml of 0.17 M pyridine–acetate followed by a 0.17–2 M gradient of pyridine–acetate (125 ml/125 ml). Fraction volumes collected were 2 ml; the horizontal bars represent pooled fractions.

together, these eight tryptic peptides account for 72 amino acid residues. Each peptide except Tp13 and Tp14, was individually sequenced with the automatic sequencer, and the results are presented in Table III (Tp6, Tp7, Tp10, Tp11, Tp12, Tp13, Tp14 and Tp15). Because of large extractive losses during the automatic sequencing of Tp10, a mixed carboxypeptidase A and B digest was subsequently carried out to confirm the sequence of the four C-terminal residues of this peptide. The carboxypeptidase data confirmed the sequence obtained with the automatic sequencer.

The sequence of Tp13 (Thr–Lys) and Tp14 (Gly–Lys) was determined on the basis of trypsin specificity.

TABLE VI

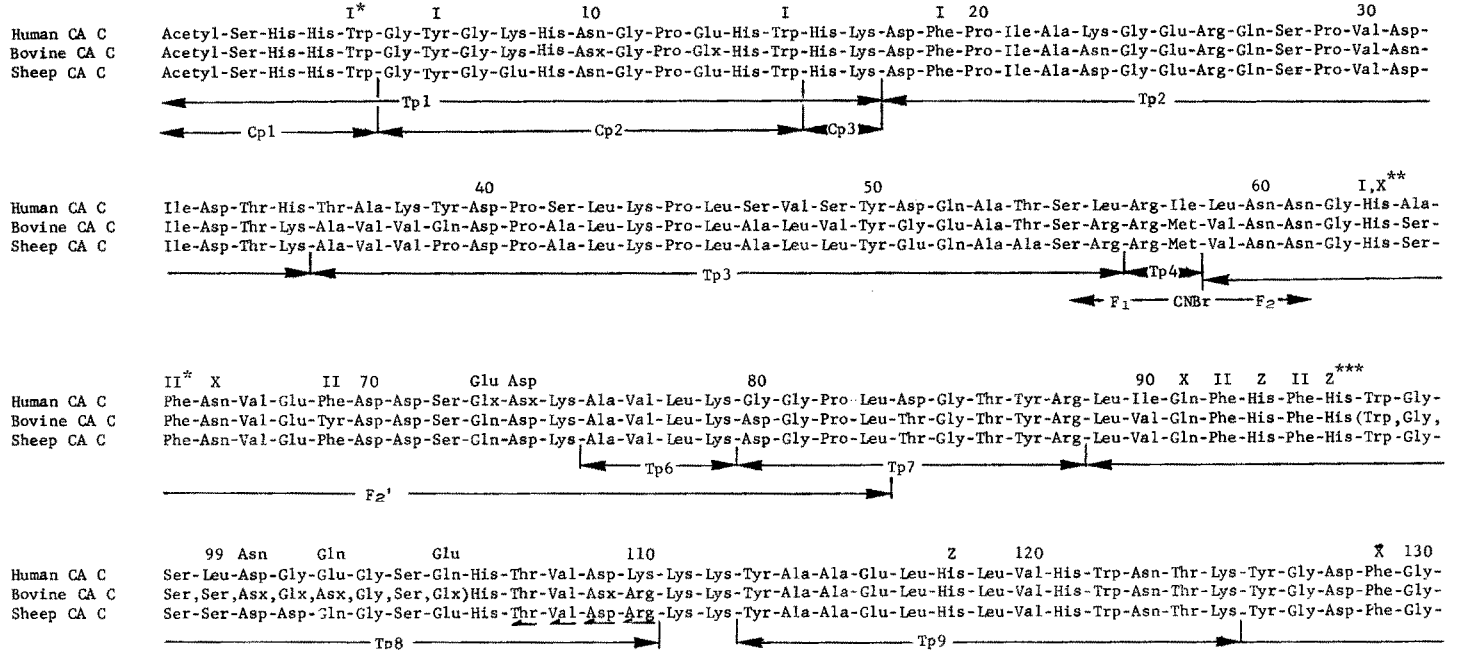
AMINO ACID COMPOSITION OF TRYPTIC PEPTIDES Tp8 AND Tp9 ISOLATED FROM A SP-SEPHADEX COLUMN DEVELOPED WITH A PYRIDINE-ACETATE GRADIENT
The residue values represent observed and nearest integers, respectively.

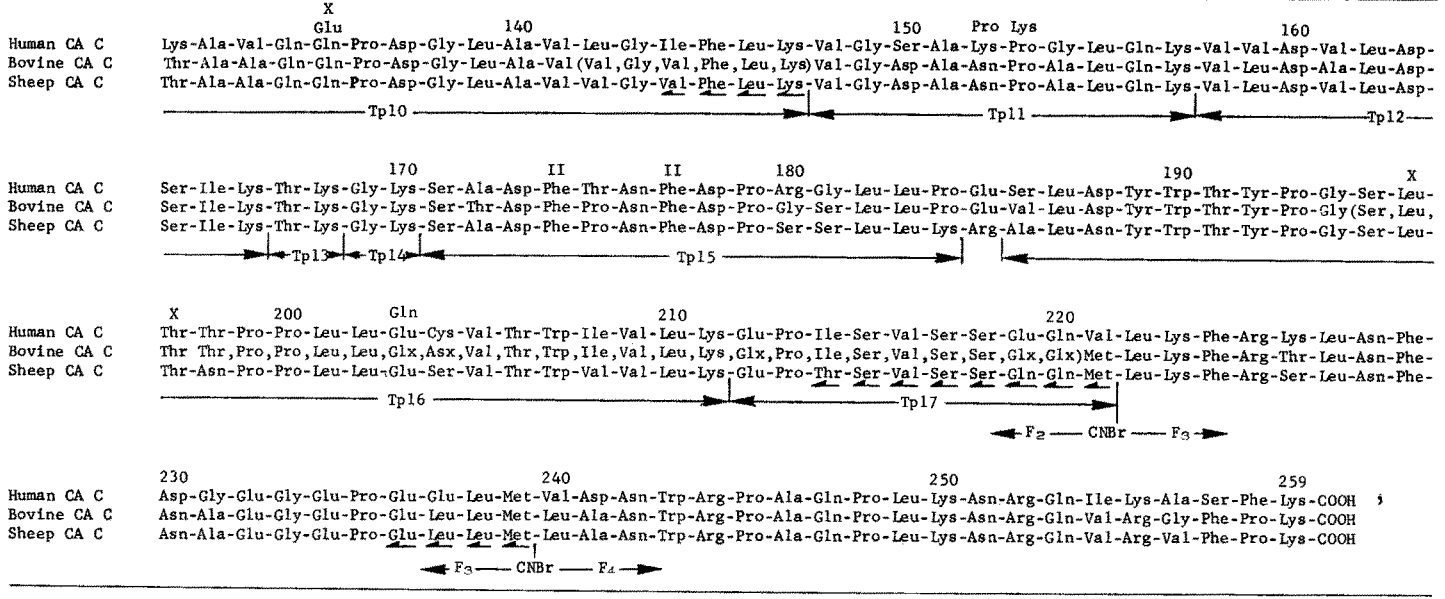
Amino acid	Tp8	Tp9
Asp	2.7 (3)	1.1 (1)
Thr	0.9 (1)	1.0 (1)
Ser	2.6 (3)	
Glu	2.7 (3)	1.1 (1)
Gly	1.9 (2)	
Ala		1.9 (2)
Val	1.8 (2)	1.0 (1)
Leu	1.0 (1)	2.0 (2)
Tyr		0.9 (1)
Phe	2.0 (2)	
Lys		1.1 (1)
His	2.5 (3)	1.8 (2)
Arg	1.1 (1)	
Trp*	(1)	(1)
Residue total	22	13

* Determined by automatic sequencer analysis.

TABLE VII

Amino acid sequence of sheep carbonic anhydrase C. The sequence for human carbonic anhydrase C is from Henderson et al. [6] and the bovine sequence from Foveau et al. [7], and Nyman et al. [8]. The residues shown above the human C sequence at positions 73, 74, 100, 102, 105, 136, 152, 153, and 203 are those residues from the human C sequence of Lin and Deutsch [4] which differ from the sequence of Henderson et al. [6]. Residues determined or confirmed with carboxypeptidase are symbolized by ←.





* Residues which are thought to comprise aromatic clusters I and II (see text) have been designated I and II, respectively.

** Active site residues discussed in the text are marked with an X.

*** Residues known to form ligands with the zinc atom at the active site are marked with a Z.

Tryptic Peptides Tp8 and TP9 were isolated in high yields using material derived from a tryptic digestion of the CNBr fragment F_2 without acidification. After enzymatic digestion at pH 8.4 in 0.05 M MalNEt, the sample was carefully concentrated and equilibrated with 0.17 M pyridine–acetate buffer prior to fractionation on a 1.5 cm × 30 cm SP-Sephadex column equilibrated with 0.17 M pyridine–acetate buffer (pH 5.0). This column was initially developed with 100 ml of 0.17 M pyridine–acetate buffer (pH 5.0), followed by a 250 ml 0.17–2.0 M gradient. Using this gradient, two clearly resolved, ultraviolet positive peaks, were observed; the first corresponded to Tp8 while the second proved to be Tp9. Fig. 3 shows a typical elution profile and Table VI presents the amino acid composition data of these two peptides; subsequent sequence determinations accounted completely for these composition results.

Using the automatic sequencer it was possible to determine the sequence of the first 18 residues of Tp8. However, some difficulty was experienced with the first two histidine residues (sequencer Steps 5 and 7) generating out of phase sequences. To complete the sequence of Tp8 (Residues 19–22), a mixed carboxypeptidase A and B digest was employed; carboxypeptidase B treatment for 20 min followed by carboxypeptidase A digestion for various time intervals up to 2 h. The results of such a digestion provided the following sequence for the C-terminal portion of Tp8: Thr(107)–Val–Asp–Arg.

The complete sequence of Tp9 was determined after treating the peptide with the Braunitzer reagent. Contrary to our experience with Tp8 no difficulty was experienced with the two histidines contained in this peptide.

The final amino acid sequence for both Tp8 and Tp9 is presented in Table III. These two peptides are linked through a pair of lysine residues as shown in Table VII. Evidence for this linkage was acquired by treating the CNBr F_2 material with citraconic anhydride followed by tryptic digestion. The resulting peptides were acidified with acetic acid, and the insoluble material isolated by centrifugation. This insoluble material was used directly for automatic sequence analysis. During the course of analyzing the first five steps, three sequences were clearly observed, two of which could be accounted for by previous work (Tp8 and Tp16). The third sequence proved to be Lys–Lys–Tyr–Ala–Ala and on the basis of homology this material was assumed to represent Residues 111–115 (see Table VII).

Further sequence information in the F_2 region was obtained by treating native enzyme with citraconic anhydride followed by digestion at arginine bonds with trypsin for up to 6 h. Adjustment of the pH of this peptide mixture, to between pH 5 and 6, with 1.0 M acetic acid resulted in a precipitate which was subsequently washed with water and partially resuspended in 0.5 M acetic acid. The material which was not soluble in 0.5 M acetate was isolated by centrifugation and designated Tp16. Amino acid analysis of Tp16 suggested a low level of contamination; however, attempts at further fractionation were unsuccessful. On the other hand, when this material was analyzed in the automatic sequencer only one major sequence was observed; two others, occurring at low levels, could be accounted for and were no longer observed following sequencer Step 6. The amino acid sequence of Tp16 is presented as all 26 residues of Tp16 plus three residues of Tp17 (Table III); Steps 29 and 30 could not be positively identified using the automatic sequencer due to extractive losses.

The amino acid sequence of the Tp17 region was completed by carrying out carboxypeptidase A digestion of the entire F_2 CNBr fragment over a time span of

5 min to 3.5 h. This fragment was soluble in the MalNEt buffer titrated to pH 8.5. The low activity of the serine residues modulated the rate of amino acid release, and the placement of the valine at Position 216, using the automatic sequencer, greatly assisted in our interpretation of the sequence. The proline at Position 213 prevented additional carboxypeptidase A digestion. The results of such a carboxypeptidase digestion provided the following sequence: Thr(214)–Ser–Val–Ser–Ser–Gln–Gln–Met. The complete sequence of Tp17 is shown in Table VII.

DISCUSSION

Our present results provide the amino acid sequence of several additional peptides comprising sheep erythrocyte carbonic anhydrase C, and along with previously reported work [1], these peptides account for the complete primary sequence of the enzyme. Because very little overlapping sequence data were obtained, the final alignment of the sheep peptides was constructed, for the most part, on the basis of inferred homology with the known sequences of the human and bovine CA C enzymes [2, 6–8]. Because reasonable homology was observed at all times, this approach could be employed quite effectively. Table VII shows the previously reported sequences of the human and bovine enzymes along with the sequences of the sheep enzyme. The tryptic, chymotryptic and CNBr data for the sheep molecule are indicated. For the purpose of convenience, we will follow the numbering of the human carbonic anhydrase C sequence [6] in aligning the homologous sites of the C enzymes (Table VII).

CNBr fragment, F_1, represents the N-terminus of the molecule, and is comprised of four tryptic peptides (Tp1, Tp2, Tp3, and Tp4). Two of these, Tp2 and Tp3, were sequenced directly and demonstrated a high degree of homology to the corresponding human and bovine sequences. The dipeptide, Tp4, comprised of methionine and arginine, is obviously the C-terminus of F_1. The blocked N-terminus tryptic peptide, Tp1, required further cleavage with chymotrypsin to yield material which could be analyzed in the automatic sequencer or treated with carboxypeptidase A. The Arg–Gln bond (Positions 26 and 27 of Tp2) was observed to be resistant to short-term tryptic cleavage, and digestion times of 6 h or longer were required to quantitatively hydrolyze this bond.

Following isolation, the CNBr fragment F_2 could be sequenced directly to obtain the primary sequence of the first 25 residues. Furthermore, Tp6 and part of Tp7 overlap and confirm the C-terminal portion of the sequence determined in this fashion. Much of the remaining portions of this fragment were sequenced as tryptic peptides either with or without prior treatment to block tryptic cleavage at lysine residues. Some of these peptides were observed to be quite insoluble and therefore difficult to isolate in a purified state.

The arginine at Position 185 was placed on the observation that, when the lysine residues in the whole enzyme are blocked with citraconic anhydride and treated with trypsin, the Peptide Tp16 is quantitatively released. This peptide is adjacent to the carboxyl side of Arg-185.

Because of the obvious homology between human, bovine and sheep carbonic anhydrase C, it was of interest to compare in detail the sheep sequence with the amino acid sequences determined for these other sources as well as to compare residues

whose positions and possible functions have been based on the three-dimensional structure of human carbonic anhydrase C [9, 10, 15, 16].

The homologies of the C enzymes are apparent when the human C sequence is compared to those of bovine and sheep (Table VII). The following percentages of identical amino acid homologies were found: human C/sheep C, 78%; human C/bovine C, 77%; sheep C/bovine C, 92%. In contrast, the degree of identical homology between the various C enzymes and the human low activity B isozyme, carbonic anhydrase B, are: human B/sheep C, 58%, and human B/human C, 60%. The implications concerning the differential rates of evolution of the human, sheep, and bovine carbonic anhydrases have been treated elsewhere [17, 19], and will not be detailed here.

The three-dimensional studies of the active site region of human carbonic anhydrase C indicate that the histidine residues at Positions 94, 96 and 118 are involved in binding the zinc to the active site [9, 10, 15, 16]. The sheep and bovine sequences also show histidine residues at these positions.

Other residues which have been located in the active site region of the human carbonic anhydrase C molecule by X-ray diffraction studies [9, 10, 15, 16] such as His-63, Asn-66 (formerly designated Gln-66), Gln-91, Phe-129 (formerly His-128), Leu-196, Thr-197, and Glu-116 (formerly Glx-115) were found to be at corresponding positions in the sheep enzyme without exception.

On the basis of the X-ray studies of human carbonic anhydrase C [9, 16], two pronounced aromatic clusters have been described: one which is located near the N-terminus (Cluster I) and the other (Cluster II) in the hydrophobic core of the molecule. It is of interest that the residues found in these clusters have remained essentially the same in the human, bovine and sheep enzymes. The five residues comprising Cluster I of human carbonic anhydrase C (Trp-4, Tyr-6, Trp-15, Phe-19 and His-63) are the same in all three carbonic anhydrases. The six phenylalanyl residues comprising Cluster II at Positions 65, 69, 92, 94, 174, and 177 occupy homologous positions in the human and sheep and bovine with the exception of Phe-69 which is a tyrosine residue in the bovine enzyme.

ACKNOWLEDGEMENT

We wish to thank Dr Anders Liljas and Dr K. K. Kannan (University of Uppsala) for the many helpful discussions we had with them concerning the three-dimensional structure of human carbonic anhydrase C. We are also grateful to Dr Louis Henderson (University of Göteborg) for providing us with much useful information concerning the primary sequence of human carbonic anhydrase C. This work was supported by research grant GM-15419 from the U.S. Public Health Service. One of us (R.E.F.) was supported as a postdoctoral trainee under U.S. Public Health Service Training Grant T01 GM0007.

REFERENCES

1 Tanis, R. J. and Tashian, R. E. (1971) Biochemistry 10, 4852–4858
2 Tashian, R. E., Tanis, R. J. and Ferrell, R. E. (1972) in Oxygen Affinity of Hemoglobin and Red Cell Acid Base Status (Rørth, H. and Astrup, P., eds) pp. 353–362, Alfred Benzon Symposium IV, Academic Press, New York

3 Andersson, B., Nyman, P. O. and Strid, L. (1972) Biochem. Biophys. Res. Commun. 48, 670–677
4 Lin, K.-T. D. and Deutsch, H. F. (1974) J. Biol. Chem. 249, 2329–2337
5 Lin, K.-T. D. and Deutsch, H. F. (1973) J. Biol. Chem. 238, 1881–1884
6 Henderson, L. E., Henriksson, D. and Nyman, P. O. (1973) Biochem. Biophys. Res. Commun. 52, 1388–1394
7 Foveau, D., Sciaky, M. and Laurent, G. (1974) C. R. Acad. Sci. Paris Ser. D 278, 959–962
8 Nyman, P. O., Strid, L. and Westermark, G. (1968) Eur. J. Biochem. 6, 172–189
9 Kannan, K. K., Liljas, A., Waara, I., Bergstén, P. C., Lövgren, S., Strandberg, B., Bengtsson, U., Carlbom, U., Fridborg, K., Järup, L. and Petef, M. (1971) Cold Spring Harbor Symp. Quant. Biol. 36, 221–231
10 Liljas, A., Kannan, K. K., Bergstén, P. C., Waara, I., Fridborg, K., Strandberg, B., Carlbom, U., Järup, L., Lövgren, S. and Petef, M. (1972) Nature New Biol. 235, 131–137
11 Braunitzer, G., Schrank, B. and Ruhfus, A. (1970) Hoppe-Seyler's Z. Physiol. Chem. 351, 1589–1590
12 Jones, R. T. (1970) Methods Biochem. Anal. 18, 205–258
13 Gibbons, I. and Perham, R. N. (1970) Biochem. J. 116, 843–849
14 Ambler, R. P. (1972) in Methods in Enzymology (Hirs, C. H. W. and Timasheff, S. N., eds) Vol. 25 (Part B), pp. 262–272, Academic Press, New York
15 Waara, I., Lövgren, S., Liljas, A., Kannan, K. K. and Bergstén, P. O. (1972) in Hemoglobin and Red Cell Structure and Function (Brewer, G. J., ed.) pp. 169–187, Plenum Press, New York
16 Bergstén, P. C., Waara, I., Lövgren, S., Liljas, A., Kannan, K. K. and Bengtson, U. (1972) in Oxygen Affinity of Hemoglobin and Red Cell Acid Base Status (Rørth, H. and Astrup, P., eds) pp. 363–383, Alfred Benzon Symposium IV, Academic Press, New York
17 Tashian, R. E., Tanis, R. J., Ferrell, R. E., Stroup, S. K. and Goodman, M. (1972) J. Human Evolution 1, 545–552
18 Beaven, G. H. and Holiday, E. R. (1952) Advan. Protein Chem. 7, 319–386
19 Tashian, R. E., Goodman, M., Tanis, R. J., Ferrell, R. E. and Osborne, W. R. A. in Proc. 3rd Int. Isozyme Conf. (Markert, C. L., ed.) Academic Press, New York, in press
20 Notstrand, B., Waara, I. and Kannan, K. K. in Proc. 3rd Int. Isozyme Conf. (Markert, C. L., ed.) Academic Press, New York, in press