Use of Dipeptidyl aminopeptidase I (Cathepsin C) for Sequence analysis of peptides.

IKRAM ULLAH KHAN*, AHMAD SAEED**
AND MASOOD R.BAIG***

*Department of Pharmacy and **Department of Chemistry,
Gomal University, Dera Ismail Khan, Pakistan.

(Received 10th December, 1984, Revised Received 24th March,1985)

Summary:A possible method for the determination of amino acid sequence of polypeptides involves their degradation into dipeptide fragements with dipeptidyl amino peptidase I (Cathepsin C). The action of the enzyme was characterized on different model paptides of known structure.

Some tryptic paptides of rabbit muscle aldolase were isolated by gel filtration on Sephadex G-25 column and purified by electrophoretic and chromatographic methods. These peptides and synthetic peptide $ACTH_{1-32}$ were subjected to dipeptidyl aminopeptidase I digestion. The yield of the dipeptides were reasonable good and the structure of the dipeptides was determined by dansyl-Edman method, thin layer ion exchange chromatography and high voltage electrophoresis.

The action of dipeptidyl aminopeptidase I (Cathepsin C) was completely inhibited by internal prolyl residues.

Introduction

The proteolytic enzymes which catalyze the hydrolysis of proteins, are present in the lysosomes of cells ranging from protozoa to the tissue of higher animals. All organelles contain proteolytic activity in various amounts, the lysosomes in particular are rich in proteinases and play an important role in cell physiology.

Cathepsin C (dipeptidyl aminopeptidase I) has been extensively purified by Planta and Gruber [1]; Metrione et al [2].

Highly purified cathepsin C from rat liver and bovine spleen degrades a wide variety of peptides. More than 200 dipeptides derived from polypeptides digested with dipeptidyl aminopeptidase I have been identified by mass spectrometry [3]. As the enzyme can liberate dipeptides and since a free α-amino or α-imino group on the substrate is required, the name dipeptidyl amino peptidase I has been used by McDonald et al [4]. This enzyme has been used as a "sequencing reagent" as a result of its broad specificity in catalyzing the consective removal of dipeptide moieties from the unsubstituted NH_2-termini of polypeptide chains. Several methods have been applied for the analysis of dipeptides e.g. ion exchange chromatography, paper chromatography and gas chromatography or combined gas chromatography mass spectrometry [3,5].

In the present paper we examined the possible use of cathepsin C (di-

peptidyl aminopeptidase I) for sequence analysis of peptide. We separated the dipeptide products by two dimensional electrophoresis and chromatography from rabbit muscle aldolase and $ACTH_{1-32}$. The amino acid composition of the peptides was determined by one dimensional thin layer ion exchange chromatography and by high voltage paper electrophoresis.

Experimental

Materials

Cathepsin C was purchased from Sigma (London) Chemical Co., Sephadex G-25 was supplied by Pharmacia Fine Chemicals AB. Dowex 50x8 resin of Analytical Reagent grade.

Methods

(i) Carboxymethylation and tryptic digestion of aldolase

The rabbit muscle aldolase was isolated according to the method of Taylor et al [6]. The aldolase was dissolved in 0.1M Tris buffer pH 8.5 containing 8M Urea, to a final concentration of 7 mg/ml. The SH group (8 moles/subunit of aldolase) were carboxymethylated by adding 2.5 moles of bromoacetate per mole SH group. The mixture was incubated at room temperature for 1h., then the solution was dialysed against 0.001M HCl. To the dialysed solution trypsin was added trypsin: protein weight ratio of 1:50. The pH was adjusted to 8.5 with solid NH_4HCO_3 and digestion was carried out at 37°C for 3 hours.

(ii) Gel filtration

The tryptic digest was applied to a Sephadex G-25 column (2x180 cm) equilibrated with 0.1M NH_4HCO_3, pH 8.0 containing 1% n-butanol. The chromatography was performed at room temperature with flow rate of 12 ml/h. 3 ml fractions were collected. The fractions were detected by measuring absorbance at 225 nm.

(iii) High Voltage Electrophoresis (HVE)

In order to resolve the tryptic peptides, high voltage electrophoresis was used.

(a) pH 6.5 HVE:

The material to be fractionated was applied as a thin line to the middle of the filter paper. The paper was wetted with buffer pH 6.5 pyridine-acetic acid - water (90:4:900 by volume). The excess of buffer was blotted with a filter paper and the loaded paper was put in high voltage electrophoresis tank. A potential of 4000 volts was applied. Using a 20 cm wide Whatman No.3 mm filter paper, a current of 40 - 50 mA was produced. The paper was removed after 1h run and dried in 45°C air drier. Control strips on the two sides were cut off, stained with ninhydrin cadmium acetate solution and dried.

(b) pH 1.9 HVE

With the guidance of control, a strip approximately 1 cm wide was cut from the unstained part of the paper. The strip was sewn onto another sheet of filter paper at the anodic end for second electrophoresis at pH 1.9. The electrophoresis was performed in acetic acid-formic acid-water (8:2:90 by volume) buffer system. As a marker εDNP-lysine was used and potential of 4500 volts was applied.

(iv) Ascending Chromatography

To obtain a good resolution of neutral peptides after two dimensional

electrophorogram, the paper strip which contained neutral peptides, was cut out. Then it was sewn to another sheet of filter paper and subjected to ascending chromatography in pyridine-isoamyl alcohol-water (35:35:30 by volume), in the direction of right angle to that of pH 1.9 electrophoresis.

(v) Digestion of peptides with Cathepsin C

The typical reaction mixture contained a peptide dissolved in 0.2 ml of buffer solution (0.8% pyridine, 16mM HCl, 15mM 2-mercaptoethanol, pH 5.0). The reaction was initiated by adding 30-50ug of bovine spleen cathepsin C. After 18h of incubation, the dipeptides were separated by one or two dimensional electrophoretic and chromatographic methods. The products were eluted from paper by 0.1M NH_4HCO_3 solution and the constituent amino acids were indentified by thin layer ion exchange chromatography and pH 1.9 HVE.

(vi) N-terminal determination

The peptide solution was dried in a vacuum desicator. The dry sample was dissolved in 10 ul of 0.2M $NaHCO_3$ and dried again in vacuo. The dry sample was dissolved with 10 ul deionized water then 10 ul dansyl chloride reagent (2 mg/ml) was added. The tube was sealed with parafilm and incubated at 37°C for 30 minutes. The reaction mixture was dried and 20 ul 6N HCl was added. The sample was hydrolysed at 105°C for 18h. The hydrolysed sample was dired in vacuo and dansyl amino acid was identified accoridng to the methods reported by Hartley [7], Sajgo [8], Edman and Begg [9].

(vii) Thin layer ion exchange chromatography

The method for one dimentional separation of amino acids by thin layer ion exchange chromatography on plates coated with Dowex 50x8 type resin was described by De-venyl et al [10]. The amino acids were identified by comparison with their standards. 0.4M citrate buffer pH 3.3 was used for the purpose.

Results and Discussion

The amino acid sequence of rabbit muscle aldolase is already known [11]. In order to obtain model-peptides for studying the effect of cathepsin, we isolated some peptides from tryptic digest of carboxymethylated rabbit muscle aldolase. The peptides were separated by Gel chromatography. Nine peaks were observed and pooled to have ATS_1 to ATS_9 peptide fractions respectively (Fig.1). The peptides of ATS_1 fraction were present in relatively low yield due to their large size. On the other hand, the size of peptides in fractions ATS_5-ATS_9 were small and not suitable for further

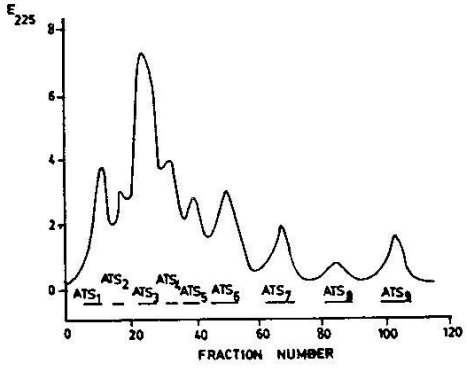

Fig.1: Separation of tryptic peptides of rabbit muscle aldolase by Gel filtration. 100 mg of tryptic digest was applied to the Sephadex G-25. The column (2x80 cm) was equilibrated with 0.1M NH_4HCo, pH 8.0 and operated at room temperature at flow rate of 12 ml_3/h. 3 ml fractions were collected. Fractions pooled are shown by bars and designated by ATS_1, ATS_3, ATS_4, ATS_5, ATS_6, ATS_7, ATS_8, and ATS_9 peptide fractions.

Table-1: List of peptides used for cathepsin C digestion

Peptide Fractions	Model peptide	N-terminal	Position in the sequence
ATS_2	ATS_2B_{22}	Pro	1-13
ATS_3	ATS_3 N	Asp	135-141
	ATS_3 B_{11}	Ala	306-313
ATS_4	ATS_4 N_{12}	Glu	320-323
	ATS_3 N_{13}	Val	106-108
	ATS_4 B_1	Val	200-206

cathepsin C digestion. While the peptides in fractions ATS_2-ATS_4 were subjected to cathepsin digestion. Therefore the peptides from these three fractions were further purified by two dimensional electrophoresis and chromatography (Table 1). Their N-terminal residues were also determined by dansyl chloride method. The site of these model-peptides within the sequence of aldolase was located on the basis of the known sequence (Fig.2).

The model-peptides were treated with cathepsin C. After incubation period, the dipeptides were isolated by HVE and chromatography. The N-terminal as well as second residue of these peptides were determined by Edman-dansyl method. The amino acid composition was also determined by thin layer ionexchange chromatography and pH 1.9 HVE. The results are summarized in Table 2. A set of peptide products obtained from model peptides by cathepsin C digestion were compared with the sequence described and the dipeptides were fitted in the sequence as follows.

1-ATS_2 B_{22} $\underset{ed_2}{\text{Pro-His-}}\underset{ed_1}{\text{Ser-His-Pro-Ala-}}$
$\underline{\text{Ley-Thr-Pro-Glu-Gln-Lys-Lys}}$

2-ATS_3 B_{11} $\underset{N_1b}{\text{Ala-Leu-}}\underset{N_1a}{\text{Gln-Ala-}}\underset{B_2}{\text{Ser-Ala-Leu-}}\underset{B_5}{\text{Lys}}$

3-ATS_3N $\underset{cdA_2}{\text{Asn-Gly-}}\underset{cdA_1}{\text{Ala-Asp-}}\underset{cdN_1}{\text{Phe-}}\underset{cdB}{\text{Ala-Lys}}$

4-ATS_4N_{12} $\underset{cd_1}{\text{Glu-Asn-}}\underset{cd_2}{\text{Leu-Lys}}$

5-ATS_4 N_{13} $\underset{cd_1}{\text{Val-Asp-}}\underset{cd_2}{\text{Lys}}$

6-ATS_4 B_1 $\underset{cd_2}{\text{Val-Leu-}}\underset{cd_2}{\text{Ala-Ala-}}\underset{cd_4}{\text{Val-Tyr-}}\underset{cd_4}{\text{Lys}}$

We checked the effect of intermediate prolyl residues on the digestibility of peptides by cathepsin C. This was

Fig.2: The amino acid sequence of rabbit muscle aldolase subunit.

```
          5                  10                  15                  20                  25
Pro-His-Ser-His-Pro-Ala-Leu-Thr-Pro-Glu-Gln-Lys-Lys-Glu-Leu-Asp-Ser-Ile-Ala-His-Arg-Ile-Val-Ala-Pro-

          30                  35                  30                  45                  50
Gly-Lys-Gly-Ile-Leu-Ala-Ala-Asp-Glu-Ser-Thr-Gly-Ser-Ile-Ala-Lys-Lys-Leu-Gln-Ser-Ile-Gly-Glx-Thr-Asx-

          55                  60                  65                  70                  75
Thr-Glx-Glx-Asx-Arg-Arg-Phe-Tyr-Arg-Ala-Phe-Pro-Glu-Asp-Asn-Gly-Arg-Pro-Val-Ile-Lys-Gln-Leu-Leu-Leu-

          80                  85                  90                  95                  100
Thr-Ala-Asp-Asp-Arg-Val-Asn-Pro-Cys-Ile-Gly-Gly-Val-Ile-Leu-Phe-His-Glu-Thr-Tyr-Gln-Leu-Lys-Gly-Gly-

          105                 110                 115                 120                 125
Val-Val-Gly-Ile-Lys-Val-Asp-Lys-Gly-Val-Pro-Leu-Ala-Gly-Glu-Thr-Thr-Thr-Asx-Glx-Gly-Leu-Asp-Gly-Leu-

          130                 135                 140                 145                 150
Ser-Glu-Arg-Cys-Ala-Gln-Tyr-Lys-Lys-Asn-Gly-Ala-Asp-Phe-Ala-Lys-Trp-Arg-Cys-Val-Leu-Lys-Ile-Gly-Glu-

          155                 160                 165                 170                 175
His-Thr-Pro-Ser-Ala-Leu-Ala-Met-Gln-Asn-Ala-Asn-Val-Leu-Ala-Arg-Tyr-Ala-Ser-Ile-Cys-Gln-Glu-Ans-Gly-

          180                 185                 190                 195                 200
Pro-Ile-Glu-Val-Pro-Glu-Ile-Leu-Pro-Asn-Gly-Asn-His-Asp-Leu-Lys-Arg-Cys-Gln-Tyr-Val-Thr-Glu-Lys-Val-

          205                 210                 215                 220                 225
Leu-Ala-Ala-Val-Tyr-Lys-Ala-Leu-Ser-Asn-His-His-Ile-Tyr-Leu-Gln-Gly-Thr-Leu-Lue-Lys-Pro-Asn-Met-Val-

          230                 235                 240                 245                 250
Thr-Pro-Gly-His-Ala-Cys-Thr-Gln-Lys-Tyr-Ser-His-Glu-Gln-Ile-Ala-Met-Ala-Thr-Val-Thr-Ala-Leu-Arg-Gly-

          255                 260                 265                 270                 275
Arg-Thr-Val-Pro-Pro-Ala-Val-Thr-Gly-Val-Thr-Phe-Leu-Leu-Ser-Gly-Glu-Ser-Glx-Glx-Glx-Glx-Gly-Ala-Ser-

          280                 285                 290                 295                 300
Ser-Val-Thr-Pro-Asx-Ile-Ile-Asn-Leu-Asn-Ala-Ile-Asn-Lys-Cys-Pro-Leu-Leu-Lys-Pro-Trp-Ala-Leu-Thr-Phe-

          305                 310                 315                 320                 325
Gly-Ser-Tyr-Gly-Arg-Ala-Leu-Gln-Ala-Ser-Ala-Leu-lys-Ala-Trp-Gly-Gly-Lys-Lys-Glu-Asn-Leu-Lys-Ala-Ala-

          330                 335                 340                 345                 350
Gln-Glu-Glu-Tyr-Val-Lys-Arg-Ala-Leu-Ala-Asn-Ser-Leu-Ala-Cys-Gln-Gly-Lys-Tyr-Thr-Pro-Gly-Ala-Ser-Glu-

          355                 360
Ser-Gly-Ala-Ala-Ala-Gln-Leu-Phe-Ile-Ser-Asn-His-Ala-Tyr.
```

important since according to the data of the literature aminopeptidase-like proteolytic enzymes cannot split those peptide bonds in which proline residues are involved. For this purpose we used synthetic ACTH as a model peptide which contains proline residues. In these experiments we found that proline stops digestion and five dipeptides were removed, i.e.,

Ser-Tyr,Ser-Met, Glu-His, Phe-Arg and Trp-Gly to expose the sequence Lys-Pro-Val.....The results are summerized in table 3. Further degradation of $ACTH_{1-32}$ was prevented by Pro-12 residue, even higher concentration of the enzyme failed to cleave beyond Trp-Gly.

Studies on the N-terminal tryptic peptides of rabbit muscle aldolase (peptide ATS_2 B_{22}) showed that from

Table-2: Terminal sequence and amino acid composition of peptides after cathepsin C digestion.

Model Peptide	Peptide product after Cathepsin digestion	N-terminal residue	2nd residue	Amino Acid Composition
1. ATS_2B_{22}	cd_1	Ser	His	Ser, His, Pro, Ala Leu, Thr, Glu, Lys
	cd_2	Pro	His	Pro, His
2. ATS_3 B_{11}	cdN_{1a}	Glu*	Ala	Glu,Ala
	cdN_{1b}	Ala	Lue	Ala,Leu
	cdB_1	Ala	Leu	Ala,Leu, Glu, Ser, Lys
	cdB_2	Ser	Ala	Ser, Ala, Leu, Lys
	cdB_3	Leu	Lys	Leu,Lys
3. ATS_3N	cdN_1	Phe	Ala	Phe, Ala
	cdA_1	Ala	Asp	Ala,Asp
	cdA_2	Asp*	Gly	Asp,Gly
	cdB	Lys	-	Lys
4. ATS_4N_{12}	cd_1	Glu	Asp*	Glu, Asp
	cd_2	Leu	Lys	Leu,Lys
5. ATS_4N_{13}	cd_1	Val	Asp	Val,Asp
	cd_2	Lys	-	Lys
6. ATS_4 B_1	cd_1	Val	Tyr	Val,Tyr
	cd_2	Val	Leu	Val, Leu
	cd_3	Ala	Ala	Ala
	cd_4	Lys	-	Lys

$ATS_2\ B_2^2$ Pro-His↘-Ser-His-Pro-Ala-Ley-Thr-Pro-Glu-Gln-Lys-Lys.

$ATS_3\ N$ Ans-Gly↘-Ala-Asp↘-Phe-Ala↘-Lys.

$ATS_3\ B_{11}$ Ala-Leu↘-Gln-Ala↘-Ser-Ala↘-Leu-Lys.

$ATS_4 B_1$ Val-Leu↘-Ala-Ala↘-Val-Tyr↘-Lys.

$ATS_4\ N_{12}$ Glu-Asn↘-Leu-Lys.

$ATS_4\ N_{13}$ Val-Asp↘-Lys

$ACTH_{1-32}$ Ser-Tyr↘-Ser-Met↘-Glu-His↘-Phe-Arg↘-Trp-Gly↘-Lys-Pro*-Val-Gly-Lys-Lys-Arg-Arg-Pro-Val-Lys-Val-Tyr-Pro-Asn-Gly-Als-Glu-Asp-Glu-Ser-Ala.

Fig.3: Hydrolysis of tryptic peptides of rabbit muscle aldolase and polypeptide hormone $ACTH_{1-32}$ by bovine spleen cathepsin C.
The attack on these peptides is illustrated and show the particular peptide bond that are cleaved during the consective removal of dipeptides. Asterisks show those positions, where digestion stopped.

Table-3: Terminal sequence and amino acid composition of peptides after cathepsin C digestion of $ACTH_{1-32}$

Peptide Product	N-terminal residue	2nd residue	Amino acid composition
$ACTH/cdB_1$	bis-D-Lys	Pro	Lys, Pro, Val, Gly, Arg Ala, Glu, Ser, Tyr, Asp.
$ACTH/cdB_2$	Phe	Arg	Phe, Arg
$ACTH/cdN_1$	*	Gly	Gly
$ACTH/cdN_2$	Ser	Tyr	Ser,Tyr
$ACTH/cdN_3$	Ser	Met	Ser, Met
$ACTH/cdN_4$	Glu	His	Glu, His

*Trp decomposes during acid hydrolysis.
A set of dipeptides obtained from the substrate ACTH with cathepsin C digestion, were arranged accroding to the known sequences as follows:

Ser-Tys↓-Ser-Met↓-Glu-His↓-Phe-Arg↓-Trp-Gly↓-Lys-Pro*-Val-Gly-Lys-
N_2 N_3 N_4 B_2 N_1 B_1

Arg-Arg-Pro-Val-Lys-Val-Tyr-Pro-Asn-Gly-Ala-Glu-Asp-Glu-Ser-Ala
B

this peptide only the first dipeptide (Pro-His) was cleaved, the second dipeptide (Ser-His) was not split off, because the enzyme could not split the His-Pro peptide bond. Conferring these data we could prove that all peptide bonds are resistant to cathepsin C in which prolyl residue is involved irrespectively to the "position" of the Pro residue in the theoretically expected dipeptide. Cathepsin C can be utilized for sequencing of small amount of peptides. The dipeptide products resulting from hydrolysis can be separated by simple technique i.e. two dimensional electrophoretic and chromatographic methods. Paper electrophoresis at pH 1.9 and thin layer ion exchange chromatography are perfectly convenient for the examination of hydrolysates of dipeptides. Therefore, after cathepsin C digestion, the dipeptide products can be easily identified without use of amino acid analyzer, the molar ratios of amino acids can be estimated by visual inspection or they can be determined by simple densitometry.

This method itself is insufficient for establishing a sequence of peptide. For small peptides, time-course analysis of the digestion alone will often establish the order of the appearance of product dipeptides. For longer peptides where the order of dipeptide release becomes indistinguishable, it is necessary to generate the overlaping alternate dipeptides. These overlaping dipeptides can be obtained by repeating the degradation of the native peptide after removal of N-terminal amino acid by Edman degradation. The identity of these alternate dipeptides will establish the amino acid sequence of the unmodified peptide.

In theoretical example, the original set of dipeptides from $ACTH_{1-32}$ are released sequentially after digestion with cathepsin C i.e. Ser-Tyr, Ser-Met, Glu-His, Phe-Agr, Trp-Gly etc. (see also Fig. 3). The peptide substrate is modified by removing an amino acid residue by Edman method, the modified peptide yields the alternate dipeptides: Tyr-Ser,Met-Glu,His-Phe, Arg-Trp etc.

```
Original dipeptides:- Ser-Tyr,Ser-Met,Glu-His,Phe-Arg,Trp-Gly,
Alternate dipeptides:     Tyr-Ser,Met-Glu,His-Phe,Arg-Trp,
Residue position:-    1 2  3  4  5 6  7  8  9 10
```

It is concluded that cathepsin C digestion can also be usefull complementary tool of the sequence analysis.

References

1. R.J.Planta and M.Gruber, *Biochem.Biophys.Acta;* ***89***, *503 (1964).*
2. R.M. Metrione, Y.Okuda and G. F. Fairclough, *Biochemistry*, ***9***, *2427 (1970).*
3. H.C.Krutzsch and J.J.Pisano, *Biochemistry*, ***17***, *2791 (1978).*
4. J.K.McDonald; P.X. Callahan, B.B.Zeitman and S.Ellis, *J.Biol.Chem.* ***222***, *2693 (1969).*
5. P.X.Callahan; J.A. Shepard; T. J. Reilly, J.K. McDonald and S. Ellis, *Anal.Biochem.*, ***38***, *330 (1970).*
6. J.F.Taylor, A.A.Green and G. T.Cori, *J.Biol.Chem.*, ***173***, *591 (1948)*
7. B.S. Hartley, *Biochem.J.*, ***119***,*805 (1970).*
8. M.Sajgo, *Acta Biochem.Biophys.Acad.Sci. Hung*; ***5***, *231 (1970).*
9. P.Edman and G.Begg, *Europ.J.Biochem.*, ***1***, *80 (1967).*
10. T.Devenyi, I.Hazai, S.Ferenczi and J.Bati, *Acta Biochem.Biophys.Acad.Sci. Hung*, ***6***, *385 (1971).*
11. M. Sajgo and G. Hajos, *Acta Biochem.Biophys.Acad.Sci. Hung*, ***9***, *239 (1974).*