The EMBO Journal Vol.2 No.7 pp.1159–1163, 1983

Amino acid sequence of seminalplasmin, an antimicrobial protein from bull semen

Rainer Theil and Karl Heinz Scheit*

Max-Planck-Institut für biophysikalische Chemie, Abteilung Molekulare Biologie, 3400 Göttingen, Am Fassberg, FRG

Communicated by T.Jovin
Received on 28 March 1983

Analytical ultracentrifugation of highly purified seminalplasmin revealed a molecular mass of 6300. Amino acid analysis of the protein preparation indicated the absence of sulfur-containing amino acids cysteine and methionine. The amino acid sequence of seminalplasmin was determined by manual Edman degradation of peptides obtained by proteolytic enzymes trypsin, chymotrypsin and thermolysin: NH_2-Ser Asp Glu Lys Ala Ser Pro Asp Lys His His Arg Phe Ser Leu Ser Arg Tyr Ala Lys Leu Ala Asn Arg Leu Ser Lys Trp Ile Gly Asn Arg Gly Asn Arg Leu Ala Asn Pro Lys Leu Leu Glu Thr Phe Lys Ser Val-COOH. The number of amino acids according to the sequence were 48, the molecular mass 6385. As predicted from the sequence, seminalplasmin very likely contains two α-helical domains in which residues 8–17 and 40–48 are involved. No evidence for the existence of β-sheet structures was obtained. Treatment of seminalplasmin with the above proteases as well as with amino peptidase M and carboxypeptidase Y completely eliminated biological activity.
Key words: antimicrobial/bull seminal protein/amino acid sequence

Introduction

Seminalplasmin, an antimicrobial protein from bull semen was first isolated by Reddy and Bhargava (1979). It was shown to inhibit effectively the growth of *Escherichia coli*, a variety of other Gram-positive and Gram-negative bacteria, as well as yeast cells like *Saccharomyces cerevisiae* or *Candida albicans*. In *E. coli* (Scheit *et al.*, 1979) and similarly in the other cases investigated so far, seminalplasmin specifically interferes with RNA synthesis and very likely must penetrate cells to be effective (Scheit *et al.*, unpublished data). The highly basic protein with an isoelectric point of 9.8 was purified to apparent homogeneity and was found free from contaminating RNase, DNase and protease activities. Based on amino acid analysis, a minimal mol. wt. of 19 800 was postulated for seminalplasmin (Reddy and Bhargava, 1979). However, attempts to apply different methods such as polyacrylamide gel electrophoresis or ultracentrifugation led to conflicting results and pointed to a much lower mol. wt. of ~8000 (Reddy and Bhargava, 1979).

Recently we made further progress in the purification of seminalplasmin employing h.p.l.c. both analytically and for preparative separation (Theil and Scheit, unpublished data). Amino acid analysis of such preparations yielded a minimal molecular mass of ~6500. Sedimentation velocity and sedimentation equilibrium runs at pH 7 in 0.1 M NaCl similarly indicated a mol. wt. of 6300.

Here we report the amino acid sequence of seminalplasmin, which was purified by h.p.l.c. as the last step of an isolation procedure.

*To whom reprint requests should be sent.

Results

The purity of the seminalplasmin preparation was indicated by detection of only serine as the N-terminal amino acid using dansylation. Amino acid analysis revealed the absence of the sulfur-containing amino acids cysteine and methionine. On the basis of five amino acids: tryptophan, valine, tyrosine, isoleucine and threonine, which were present in a 1:1 relationship, the minimal number of amino acids per molecule would be ~50 (Table I).

For sequence determination, seminalplasmin was hydrolysed by trypsin, chymotrypsin and thermolysin. The resulting peptides were separated by paper electrophoresis and paper chromatography. The two dimensional separation patterns are shown in Figure 1A, B and C. The peptides T14, Th12 and Th13, which had no charge at pH 6.5 were not included. The isolated peptides were subjected to amino acid analysis. The composition and yields of the peptides are summarised in Tables II, III and IV. As can be seen from Table III, digestion of seminalplasmin with chymotrypsin resulted in a considerable number of peptides in low yields. The peptides were manually sequenced by the dansyl-Edman technique (Bruton and Hartley, 1970) (Figure 2). Only two tryptic peptides, namely T1: Ser-Asp-Glu-Lys, and T14: Ser-Val, contained serine in the N-terminal position. Because hydrolysis of seminalplasmin with carboxypeptidase Y yielded valine and only one valine was presumably present in the polypeptide according to the amino acid analysis, it was concluded that peptide T1 was the amino-terminal tryptic peptide and T14 was the carboxy-terminal peptide of seminalplasmin, respectively. Edman degradation of intact seminalplasmin proceeded up to

Table I. Amino acid composition of seminalplasmin

Amino acid residue	mol/100 mol	Minimal numbers[a]	Sequence
Asp	12.4	5.9	6
Thr	2.1	1.0	1
Ser	11.4	5.4	6
Glu	5.04	2.4	2
Gly	5.55	2.6	2
Ala	9.31	4.4	4
Val	1.95	0.9	1
Ile	2.67	1.3	1
Leu	14.38	6.9	6
Tyr	2.78	1.3	1
Phe	4.39	2.1	2
Lys	11.71	5.6	6
His	4.72	2.3	2
Arg	10.46	5	5
Trp	1.06	0.5	1
Pro	5.0	2.4	2
Total		50	48

[a]The minimal numbers of amino acid residues were calculated under the assumption that the protein contained at least one threonine.

© IRL Press Limited, Oxford, England.

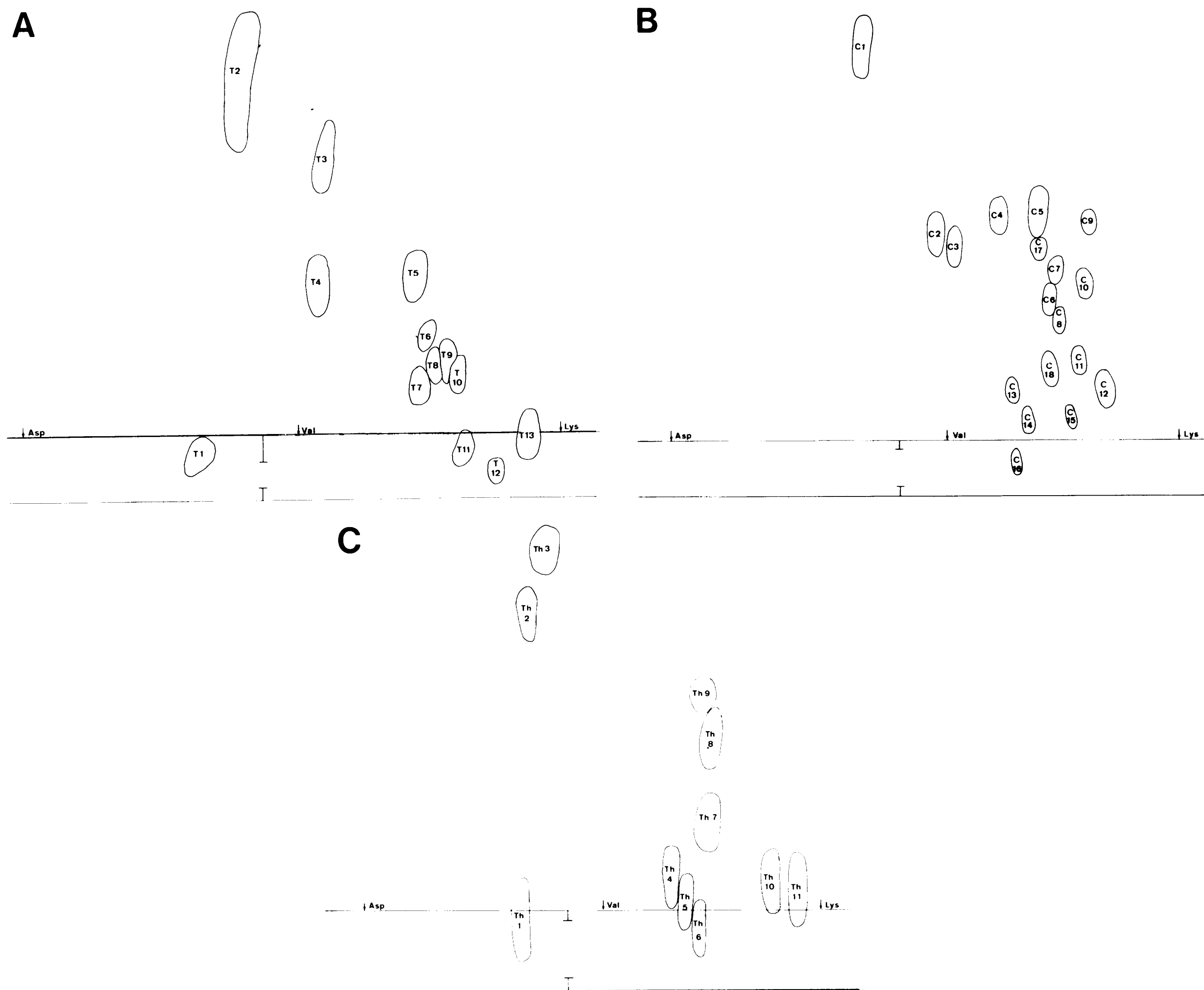

Fig. 1. Separation pattern of peptides. **(A)** Tryptic peptides: **(B)** chymotryptic peptides; **(C)** thermolytic peptides. Val, Lys and Asp indicate the positions of these amino acids run as markers.

residue 6 and allowed us to align peptides T1 and T12. The latter was the only tryptic peptide which could not be sequenced in full. A small tryptic peptide, T13, was found, which resulted from further cleavage of T12. Dansyl-Edman degradation yielded the sequence His-His-Arg. The relationship of this peptide to T12 was established by the overlapping sequence of peptide Th6. Peptide Ch16 was well separated in the peptide map (Figure 1B). It could be isolated in pure form as judged from end group determination. The N-terminal sequence Ser-Asp-Glu, which is unique among all peptide sequences, in conjunction with the characteristic amino acid sequence, led to alignment of peptides T12 and T5 thus providing the sequence of residues 1 – 17. Overlaps of chymotryptic peptides Ch18 as well as Ch8 with peptides T9, T8 and T10 allowed alignment to give the sequence of residues 18 – 27. The overlapping sequences of Ch5 and Th9 confirmed the sequence up to residue 27. The peptides T4, Ch8 and Th8 displayed blue fluorescence on paper and therefore were likely to contain tryptophan. Because amino acid analysis involved hydrolysis by HCl, tryptophan could not be detected due to chemical decomposition. Similarly, dansyl-Edmann degradation did not allow identification of tryptophan residues. From the fact that peptides T4, Ch8 and Th8 displayed blue fluorescence, from the amino acid composition of these peptides, as well as from lack of fluorescence of peptide T4 after the first Edman degradation, it was inferred that tryptophan occupies position 28 of the sequence. The sequential nature of peptides T4 and T11 became evident from the overlapping sequences of peptides Th5 and Ch14. The N-terminal sequence Ala-Asn-Pro-Lys-Leu, as well as the composition of Ch3, aligned the tryptic peptides T7 and T3. The link to Arg (35) of the polypeptide chain was provided by the sequence of Ch12 which overlapped the N termini of peptides T7 and Ch3, thus establishing the sequence of residues 35 – 46. Treatment of seminalplasmin with carboxypeptidase Y furnished amino acids of the composition Leu_2, Glu, Thr, Lys, Ser, Val. It is interesting that further action of the enzyme was blocked at Lys(40). Detailed kinetic analysis of liberation of amino acids led to the C-terminal sequence of residues 43 – 48 (Figure 2, CY), which overlapped sequences

1 5 10 15
Ser-Asp-Glu-Lys-Ala-Ser-Pro-Asp-Lys-His-His-Arg-Phe-Ser-Leu-Ser-Arg-Tyr

T1 T12 T5
T13
Ch16 Ch5
Th1 Th6 Th9

20 25 30 35
Tyr-Ala-Lys-Leu-Ala-Asn-Arg-Leu-Ser-Lys-Trp-Ile-Gly-Asn-Arg-Gly-Asn-Arg-Leu

T9 T8 T10 T4 T11
Ch8 Ch14
Ch18
Th10 Th8 Th5

35 40 45
Arg-Leu-Ala-Asn-Pro-Lys-Leu-Leu-Glu-Thr-Phe-Lys-Ser-Val

T7 T3 T14
T2
CY
Ch3 Ch2
Ch1
Ch12 Th13
Th7 Th2

Fig. 2. Amino acid sequence of seminalplasmin. The abbreviations for the proteolytic peptides are identical to those employed in Tables II, III, and IV. The symbol --> denotes amino acid residues determined by dansyl-Edman degradation. The symbol <-- represents amino acids identified by carboxypeptidase Y (CY) treatment.

of peptides T3 and T14.

We investigated in which way treatment of seminalplasmin with the various proteases affected the biological activity. The latter was assayed by the potential of seminalplasmin to inhibit RNA synthesis in *E. coli*. Action of the proteases trypsin, chymotrypsin, thermolysin, aminopeptidase M and carboxypeptidase Y led to total loss of biological activity. The effect of carboxypeptidase Y is interesting because degradation from the C terminus reproducibly proceeded only to residue Lys(40), leaving the seminalplasmin sequence 1 – 40 intact; but despite this limited hydrolysis, loss of biological activity was total.

Table II. Summary of amino acid composition, yield and net charge of tryptic peptides

Peptide	Yield (%)	Net charge (pH 6.5)	Amino acid composition
T1	96.5	−1	Asp:1.0,Ser:0.8,Glu:1.0,Lys:1.1
T2	33	−1	Thr:0.9,Glu:1.0,Leu:1.7,Phe:0.8
T3	13	0	Thr:0.7,Glu:1.0,Leu:2.6,Phe:0.9,Lys:1.2
T4	80	0	Asp:1.0,Gly:1.0,Ile:0.8,Arg:1.0,Trp
T5	40	+1	Ser:1.8,Leu:1.0,Phe:0.8,Arg:1.0
T6			n.d.
T7	44	+1	Asp:1.1,Ala:1.0,Leu:1.0,Lys:1.3,Pro:1.0
T8	92	+1	Asp:1.0,Ala:1.3,Leu:1.1,Arg:1.0
T9	87	+1	Ala:1.0,Tyr:0.9,Lys:1.2
T10	70	+1	Ser:1.0,Leu:1.0,Lys:1.2
T11	41.9	+1	Asp:1.0,Gly:1.0,Arg:1.0
T12	9.8	+1	Asp:1.0,Ser:0.9,Ala:0.7,Lys:1.4,His:1.9, Arg:1.0
T13	28.6	+1	Arg:1.0,His:1.5
T14	40	0	Val:1.0,Ser:0.7

Table III. Summary, yield and amino acid composition of chymotryptic peptides

Peptide	Yield (%)	Amino acid composition
Ch1	11.7	Thr:1.1,Glu:1.0,Leu:1.9,Phe:0.9
Ch2	17.4	Ser:1.0,Val:1.0
Ch3	7.5	Asp:1.0,Thr:1.0,Glu:0.8,Ala:1.0,Leu:1.6, Phe:0.9,Lys:1.1,Pro:0.8
Ch4	4.5	Leu:1.0,Lys:1.0
Ch5	20.2	Ser:1.9,Leu:1.0,Arg:1.1,Tyr:0.8
Ch6	5.0	Ser:1.0,Leu:1.0,Arg:1.1
Ch8	8.8	Asp:1.1,Ala:0.8,Leu:1.0,Lys:0.9,Arg:1.1, Ser:0.4,Trp
Ch12	9.8	Asp:1.2,Ala:1.0,Leu:1.0,Lys:1.6,Arg:0.9,Pro:0.9
Ch14	11.5	Arg:1.8,Gly:1.5,Ile:1.0,Asp:1.8
Ch15	3.5	Asp:1.0,Ser:0.6,Gly:0.8,Arg:0.9
Ch16	19.3	Asp:1.8,Ser:1.8,Glu:1.0,Ala:0.8,Phe:0.8,Lys:2.2, His:1.7,Arg:1.0,Pro:1.0
Ch17	5.2	Lys:1.0,Phe:0.6
Ch18	10.4	Asp:1.0,Ala:1.5,Leu:1.0,Lys:1.3

Table IV. Summary, yield and amino acid composition of thermolytic peptides

Peptide	Yield (%)	Amino acid composition
Th1	13.8	Asp:1.0,Ser:0.9,Glu:1.2,Lys:1.3
Th2	9.2	Thr:0.8,Glu:0.9,Leu:1.0,Phe:1.0,Lys:1.6
Th3	7.8	Thr:1.0,Glu:1.1,Leu:1.9,Phe:0.9,Lys:0.9
Th5	17.8	Asp:2.2,Gly:2.0,Ile:1.0,Arg:1.9
Th6	40.5	Asp:1.1,Ser:1.2,Ala:1.0,Lys:1.1,His:1.1
Th7	64.6	Asp:1.1,Ala:1.0,Leu:1.0,Lys:0.8,Arg:0.6,Pro
Th8	31.7	Ser:1.0,Leu:0.9,Lys:1.2,Trp
Th9	35.2	Ser:1.1,Leu:1.0,Tyr:0.9,Arg:1.0
Th10	36.6	Ala:1.0,Lys:1.3
Th11	13.3	His:1.0,Arg:1.1
Th12	23.6	Thr:1.0,Phe:0.7
Th13	7.6	Ser:0.8,Val:1.0

Discussion

The basic features of the amino acid sequence can be summarized as follows: the sequence confirmed the absence of sulfur-containing amino acids cysteine and methionine; no clustering of either basic hydrophobic amino acids within the sequence was noticed.

The specific features are worth mentioning: the unusual sequence of residues 9−12 Lys-His-His-Arg and the repetition of a tripeptide in the sequence of residues 30−35 Gly-Asn-Arg-Gly-Asn-Arg.

Quite a number of proteins with the capacity of specific ligand binding and lacking cysteine and methionine are known. Examples are the human salivary protein statherine (Schlesinger and Hay, 1977) and intestinal Ca^{2+}-binding protein (Dayhoff, 1976). However, we are not aware of proteins lacking cysteine as well as methionine and displaying enzymatic activity. Thus, it seems unlikely that seminalplasmin, which lacks both cysteine and methionine, has any enzymatic activity that has escaped our attention.

An interpretation of the amino acid sequence as depicted in Figure 2 employing the formalism of Chou and Fasman (1974, 1975) identified two α-helical domains: the sequences of residues 8−17, Asp-Lys-His-His-Arg-Phe-Ser-Leu-Ser-Arg, as well as the sequence of residues 40−48, Lys-Leu-Leu-Glu-Thr-Phe-Lys-Ser-Val. The existence of β-sheet structures seemed unlikely. Thus, the total contribution of amino acids involved in α-helical formation might account for 39.5% of total residues.

The mol. wt. of seminalplasmin of 6300, as determined by equilibrium sedimentation, agreed well with the molecular mass of 6385 calculated from the sequence. The centrifugation experiments gave no hint of aggregation of seminalplasmin in solution.

Materials and methods

Materials

Trypsin, chymotrypsin, thermolysin, aminopeptidase M and carboxypeptidase Y were obtained from Boehringer (Mannheim, FRG). The commercial enzymes have been used without further purification. Matrix-bound aminopeptidase M and carboxypeptidase Y were from Pierce (Rockford, IL). Dansyl chloride, phenylisothiocyanate, methane sulfonic acid and trifluoroacetic acid were from Pierce (Rockford, IL). All chemicals were of analytical grade.

Inhibition of RNA synthesis in E. coli 160.37

This was employed as an assay for the biological activity of seminalplasmin (Reddy and Bhargava, 1979).

Purification of seminalplasmin

The isolation of seminalplasmin from bull semen basically followed the original protocol of Reddy and Bhargava (1979). The last step of purification was preparative h.p.l.c. on a Nucleosil C18 reversed phase column (Theil, 1983). The isolated seminalplasmin had the following biological activities: growth inhibition of *E. coli* (minimal inhibitory concentration 20 μg/ml); inhibition of RNA synthesis in *E. coli* (50% inhibition at 25 μg/ml).

Proteolytic digestion of seminalplasmin and separation of peptides

Seminalplasmin (1 mg) was dissolved in 1 ml of 0.1 M NH_4HCO_3 (pH 8) and 20 μg of either trypsin, chymotrypsin or thermolysin were added. The mixture was incubated for 10 h at 37°C. After lyophilisation, the residue was dissolved in 0.2 ml of a mixture of 10% pyridine and 0.5% acetic acid in water. The peptides were separated by electrophoresis at pH 6.5 on Whatman 3MM paper using toluene as coolant in the first dimension and by chromatography in butanol-acetic acid-pyridine-water (300:60:240:200) in the second dimension. Peptides which behaved as neutral entities at this pH were eluted and then subjected to electrophoresis at pH 3.5. Peptides containing tryptophan were detected by their intrinsic fluorescence. Other peptides were identified by fluorescamine treatment according to Vandekerckhove and van Montagu (1977). Peptides were eluted from paper with pyridine-acetic acid-

water (10:0.5:90) and characterised by amino acid analysis. For hydrolysis with carboxypeptidase Y, 1 mg of seminalplasmin was dissolved in 2 ml of 0.1 M N-ethylmorpholine, adjusted with acetic acid to pH 6.5. Carboxypeptidase Y-agarose, 2 ml of settled suspension, was added and incubation carried out at 20°C with gentle shaking. Aliquots were removed from the supernatant at timed intervals and subjected to amino acid analysis as well as determination of biological activity. Hydrolysis with aminopeptidase M-agarose was performed in a similar manner except that 0.2 M N-ethylmorpholine buffer of pH 7.3 was employed.

Edman degradation of peptides

Sequencing of the peptides was carried out using the dansyl-Edman procedure described by Bruton and Hartley (1970). Manual Edman degradation followed a protocol published by Petersen *et al.* (1975). Dansylamino acids were detected as described by Weiner *et al.* (1972). Amino acid amides and tryptophan could not be identified by the dansyl technique, although Edman degradation proceeded without difficulties. Tryptophan was therefore detected by the fluorescence of a corresponding peptide before and after Edman degradation and by subtractive methods. Amino acid amides were assigned from the mol. wt. and the electrophoretic mobility of corresponding peptides at pH 6.5 according to Offord (1966).

Analytical methods

Analytical ultracentrifugation was performed with a Beckman model E centrifuge equipped with u.v. optics and a photo-electric scanner. Centrifugation was carried out in 10 mM Tris-HCl, pH 7.2, containing 0.1 M NaCl as solvent. Sedimentation velocity runs were carried out at 44 000 r.p.m. and 25°C; $d(\log x)/dt$ was obtained from the data by least square computer fit with a correlation factor $r = 0.985$. Sedimentation equilibrium centrifugation was performed for 20 h at 30 000 r.p.m. and 25°C; for determination of mol. wt. $d(\log c)/dr^2$ was calculated by least square computer fit with correlation factors of $r = 0.99$ and better. Amino acid analysis was performed with a Kontron Liquimat Pro-1 instrument after hydrolysis with either 4 N methanesulfonic acid (seminalplasmin, 10 nmol) or 6 N HCl (peptides, 1−5 nmol). Samples were dissolved in 50 μl and hydrolysed for 24 h at 115°C. The experimental values in Table I are an average of five independent determinations.

Acknowledgements

This paper is dedicated to F.Cramer on the occasion of his 60th birthday. We appreciate the valuable advice and helpful criticism constantly provided by Dr.N.Geisler. We thank Drs.A.Henschen-Edman and B.Wittmann-Liebold for their competent comments and J.Dichter for his motivated technical assistance. We are grateful to Dr.J.Pfeilsticker (Niedersächsische Rinderproduktion Hannover) for supplying us with bull semen.

References

Bruton,C.J. and Hartley,B.S. (1970) *J. Mol. Biol.*, **52**, 165-173.
Chou,P.Y. and Fasman,G.D.(1974) *Biochemistry (Wash.)*, **13**, 222-245.
Chou,P.Y. and Fasman,G.D. (1975) *J. Mol. Biol.*, **96**, 29-45.
Dayhoff,M.O. ed. (1976) *Atlas of Protein Sequence and Structure*, Vol. **5**, suppl. 3, published by National Biomedical Research Foundation, Washington, DC, p. 282.
Offord,R.E. (1966) *Nature*, **211**, 591-593.
Peterson,J.D., Steiner,D.F., Emdin,S.O. and Falkmer,S. (1975) *J. Biol. Chem.*, **250**, 5183-5191.
Reddy,E.S.P. and Bhargava,P.M. (1979) *Nature*, **279**, 725-728.
Scheit,K.H., Reddy,E.S.P. and Bhargava,P.M. (1979) *Nature*, **279**, 728-732.
Schlesinger,D.H. and Hay,D.I. (1977) *J. Biol. Chem.*, **252**, 1680-1695.
Theil,R. (1983) PhD Thesis, Technical University, Clausthal.
Vandekerckhove,J., Rombauts,W. and Wittmann-Liebold,B. (1977) *Hoppe-Seyler's Z. Physiol. Chem.*, **358**, 989-1002.
Weiner,A.M., Platt,T. and Weber,K. (1972) *J. Biol. Chem.*, **247**, 3242-3251.