CHARACTERISTICS OF HIPROLY BARLEY III. AMINO ACID SEQUENCES OF TWO LYSINE-RICH PROTEINS

by

IB SVENDSEN and BRIAN MARTIN*

Department of Chemistry, Carlsberg Laboratory,
Gamle Carlsberg Vej 10, DK-2500 Copenhagen Valby

and

IB JONASSEN

Department of Biotechnology, Carlsberg Research Laboratory
Gamle Carlsberg Vej 10, DK-2500 Copenhagen Valby

*Present address: Max-Planck Institut, Rheinlanddam 201,
D-4600 Dortmund 1, W. Germany

Keywords: High lysine, Hiproly barley, amino acid sequence

Two lysine-rich components isolated from »Hiproly« barley have been subjected to sequence determination. The complete sequence of the smallest component, designated SP II B, has been determined. This protein consists of 72 amino acid residues and has a molecular weight of 8,072. The largest component, designated SP II A, was blocked to N-terminal Edman degradation, but sequence analyses of cyanogen bromide- and tryptic fragments showed that it was identical to the smallest component, except for the eight to eleven residues of which SP II A is longer in the N-terminus. Of these residues seven have been sequenced, and the following total sequence is reported (Glx, Val, Ser, Ser) - Lys-Lys-Pro-Glu-Gly-Val-Asn-Thr-Gly-Ala-Gly-Asp-Arg-His-Asn-Leu-Lys-Thr-Glu-Trp-Pro-Glu-Leu-Val-Gly-Lys-Ser-Val-Glu-Glu-Ala-Lys-Lys-Val-Ile-Leu-Gln-Asp-Lys-Pro-Glu-Ala-Gln-Ile-Ile-Val-Leu-Pro-Val-Gly-Thr-Ile-Val-Thr-Met-Glu-Tyr-Arg-Ile-Asp-Arg-Val-Arg-Leu-Phe-Val-Asp-Lys-Leu-Asp-Asn-Ile-Ala-Gln-Val-Pro-Arg-Val-Gly.

1. INTRODUCTION

Two water-soluble proteins, designated SP II A and SP II B, have been isolated from »Hiproly« barley (3). These proteins are present in minor quantities in various barley lines, but only

in »Hiproly« they appear in high quantities (4). They are characterized by a high content of lysine and are at least in part responsible for the high content of this amino acid in »Hiproly« when compared with other barley lines. SDS-gel

Abbreviations: DPCC = diphenylcarbonyl chloride, THEED = N,N,N',N'-tetrakis(2-hydroxyethyl)-ethylenediamine, Polybrene = hexadimethrine bromide, PTH = phenylthiohydantoin.

electrophoresis has revealed a molecular weight of 9,000 for SP II B, which is higher than the minimum molecular weight of 8,100 calculated from amino acid composition. With the same technique a molecular weight of 10,500 was found for SP II A, which according to amino acid analysis has a minimum molecular weight of 9,100. This component was found to contain one percent neutral sugar expressed as glucose equivalents (3).

The isoelectric points of SP II A and SP II B was pH 6.34 and pH 7.40, respectively, when determined in 6 M-urea by isoelectric focusing. The two proteins have very similar immunological properties (3) and this finding together with a close similarity in amino acid composition suggests that SP II B may be a proteolytic breakdown product of SP II A. In order to verify this postulate and to further characterize the proteins the determination of their amino acid sequences were undertaken. The result of this investigation is the subject of the present report.

2. MATERIALS AND METHODS

2.1. Materials

The purification of two water soluble proteins from »Hiproly« barley, designated SP II A and SP II B, respectively, is described elsewhere (3). All reagents and solvents used with the sequencer were from Pierce, Eurochemie B.V., Rotterdam, Holland, except N,N,N',N'-tetrakis(2-hydroxyethyl)-ethylenediamine (THEED) which was obtained from ICN-K&K Laboratories, Plainview, New York, ethylacetate which was from Merck Darmstadt, and heptane which was from Fluka, Switzerland. CNBr (purum) was also purchased from Fluka. Trypsin treated with DPCC was from Sigma Chemical Comp., St. Louis, USA. Biogel was a product of Calbiochem AG, Switzerland, and CM-Sephadex of Pharmacia, Uppsala, Sweden. All other reagents were analytical grade products.

2.2. Cyanogen bromide cleavage

The proteins (2–4 mg) were dissolved in 1 ml 70% formic acid and 40–50 mg cyanogen bromide added. The reaction mixture was left overnight in the dark at room-temperature. It was then freeze-dried, redissolved in 250 μ l 30%

acetic acid and applied to a Biogel P6 column (0.9 $\times$ 120 cm) equilibrated with 30% acetic acid. The flow-rate was 10 ml $\cdot$ h⁻¹ and fractions of 1 ml were collected. The peaks were located from absorption at 280 nm, and the peptide containing fractions were pooled.

2.3. Cleavage with trypsin

Tryptic cleavages of either total protein or CNBr-fragments were performed with DPCC-treated trypsin in 0.1 M-ethylmorpholino buffer, pH 8, for 4–8 hours at room-temperature. Enzyme to protein ratio was 1:100 (w/w). Before treatment with trypsin all ϵ -NH₂ groups of lysine were blocked by treatment with acetic anhydride (three additions of 20 μ l with an interval of 20 min). The separation of the resulting tryptic peptides was made either by gel filtration in 30% acetic acid, ion-exchange chromatography or by high-voltage electrophoresis.

2.3.1. Ion-exchange chromatography

Ion-exchange chromatography was made on a Sephadex CM25 column (0.5 $\times$ 15 cm) equilibrated with 0.1 M-pyridine-acetate buffer pH 5.0 and eluted with a linear gradient from 0.1 M to 2.0 M of this buffer. Peptides were located either from absorption at 280 nm or by spot-test for arginine using the phenanthrenequinone method described by YAMADA and ITANO (8).

2.3.2. High-voltage electrophoresis

High-voltage paper electrophoresis was performed on a Savant flat-plate apparatus, model FP30A using a pH 1.9 buffer, for 3 hours at 3,000 V. The peptides were located by a light spraying with ninhydrine and development of the colour overnight at room-temperature. Elution was made with 0.1 M-HCl.

2.4. Amino acid analysis

Amino acid analysis was done on a Durrum amino acid analyzer, Model D500, after acid hydrolysis for 24 hours in 5.7 N-HCl in evacuated, sealed tubes.

2.5. Sequence determinations

For the determination of the amino acid sequences an automatic liquid phase sequencer, Beckmann, model 890 C was used with programme 122974 supplied by this firm. However, 0.1 M-THEED was used instead of 1.0 M-Quadrol, and ethylacetate contained 15% propanol as recommended by BEGG and MORGAN (1). When peptides were sequenced, »Polybrene« was added to the cup as described by TARR et al. (7). Identification of PTH-amino acids was done by high pressure liquid chromatography using a Hewlett-Packard HPLC chromatograph, model 1084 B. A reverse phase (RP-8) column, 4.6×250 nm was packed with LiCrosorb RP-8, 10μ . As the elution system was used a linear gradient from (A) 0.01 M-acetic acid, pH 4.0:acetonitril:methanol (9:1:1) to (B) methanol:water (9:1). Additional information was obtained by thin-layer chromatography on polyamide sheets according to the method of KULBE (5).

2.6. Cleavage with Staphylococcal protease

To 50 nmoles of peptide dissolved in 250 μ l N-ethyl morpholino buffer, pH 8.0, was added 20 μ l of enzyme (stock-solution 0.5 mg·ml⁻¹) and the reaction mixture left for three hours at 40 °C.

3. RESULTS AND DISCUSSION

3.1. Sequence determination of SP II B

The amino acid sequence of the N-terminal part of SP II B was determined for 55 residues with a repetitive yield of 95%. A methionine located in position 48 (see Figure 1) provided a seven residue overlap to the second cyanogen bromide fragment (see below). Three amino acids were present in each cycle, an observation which at first sight pointed towards the presence of three chains or an impure preparation. However, as can be seen from Table I the underlined amino acids were repeated every

Figure 1. The amino acid sequence of lysine-rich proteins SP II A and SP II B from »Hiproly« barley.

»SP II A extension« denotes the blocked N-terminal sequence unique to SP II A. The numbering of the amino acids is based on the sequence of SP II B protein. CNBr-I and CNBr-II are the two cyanogen bromide fragments of SP II B. T₁ to T₆ are the tryptic peptides of acetylated SP II B in the order in which they appear in the sequence counting from the N-terminus. ↓ indicates bonds hydrolyzed during the purification procedure.

Table I.

»Ragged« N-terminal sequence of SP II B.

Cycle										
A.A.	1	2	3	4	5	6	7	8	9	10
Asp	121				58					
Asn		160		110				55		
Thr	54				55		55			
Gly		66		54						
Glu						155		77	131	
Ala			77							
Pro								59		59
Trp							137		87	
Leu			152		81				64	115
Lys				206		158				100
His	110		86				55			
Arg		91				52				
Seq. 1	Thr	Gly	Ala	Gly	Asp	Arg	<u>His</u>	<u>Asn</u>	<u>Leu</u>	<u>Lys</u>
Seq. 2	Asp	Arg	<u>His</u>	<u>Asn</u>	<u>Leu</u>	<u>Lys</u>	<u>Thr</u>	<u>Glu</u>	<u>Trp</u>	<u>Pro</u>
Seq. 3	<u>His</u>	<u>Asn</u>	<u>Leu</u>	<u>Lys</u>	<u>Thr</u>	<u>Glu</u>	<u>Trp</u>	<u>Pro</u>	<u>Glu</u>	<u>Leu</u>

The figures represent the A₂₅₄ of the PTH-amino acids identified by HPLC in the ten first cycles of the automated Edman degradation. The arrows underlining the amino acids in sequences 1–3 indicate the segments identical for the three sequences.

second and sixth step which meant that a peptide with a »ragged« N-terminal end was being sequenced. The longest sequence started with a threonine, the second started with Asp₅ and the third with His₇. With the method used for isolating the protein (3) it is conceivable that a trypsin-like protease has hydrolyzed the bond between Arg₆ and His₇. The reason for the split between Gly₄ and Asp₅ is more difficult to explain. A rough calculation from the figures in Table I of the yields of the three chains gave 50 % of the chain His-Asn-Leu---, 30 % of Asp-Arg-His---, and 20 % of the longest of the chains Thr-Gly-Ala--.

Acetylation and tryptic digestion of the protein produced a long peptide which was separated from the other smaller peptides by gelfiltration on Biogel P30. This peptide started with the sequence His₇-Asn₈ and was sequenced through 35 residues confirming the results discussed above.

According to the amino acid composition (see Table II) SP II B contains only a single methionyl residue. As expected, cleavage at this

residue with cyanogen bromide gave rise to only two fragments. The cleavage was performed on acetylated SP II B and the reaction product was applied directly to the sequencer without prior purification of the relevant CNBr-fragment, but with addition of Polybrene in order to reduce loss of peptide during washing with organic solvents. Nineteen cycles (Glu₄₉-Gln₆₇) were identified and verified the seven overlapping residues to the N-terminal sequence. The repetitive yield was 90 %.

In order to complete the sequence of SP II B another sample of the cyanogen bromide cleaved protein was passed over a Biogel P6 column in 30 % acetic acid and two peaks were obtained (Figure 2, CNBr-I and CNBr-II). The purity of CNBr-II was ascertained from sequence determination which at the same time served as a control (15 residues) of the sequence determination discussed above. The CNBr-II peptide was digested with DPCC-treated trypsin and applied to a Sephadex CM 25 column and eluted with a pyridine-acetate buffer gradient. From the amino acid composition (Table II) five peptides were

Table II.

Amino acid composition of SP II A and SP II B and some relevant fragments.

	SP II B ^{a)}	SP II B ^{b)}	SP II B CNBr-II	SP II A ^{a)}	SP II A ^{c)} extent.	Extent. from seq.	Extent. isolated peptide
Asp/Asn	7.1	7	4	7.8	1	1	1
Thr	3.2	4	0	3.8	0	0	0
Ser	1.1	1	0	3.2	2	(1)	2
Glu/Gln	10.3	9	2	11.45	2	1	2
Pro	4.2	4	1	4.8	1	1	1
Gly	3.8	5	1	5.9	1	1	1
Ala	3.4	4	1	4.0	0	0	0
Cys	0	0	0	0	0	0	0
Val	10.9	10	4	11.2	1	1	2
Met	1.0	1	0	1.0	0	0	0
Ile	6.0	6	2	5.6	0	0	0
Leu	6.4	6	2	5.8	0	0	0
Tyr	1.1	1	1	1.0	0	0	0
Phe	1.0	1	1	1.0	0	0	0
His	0.9	1	0	0.8	0	0	0
Lys	6.2	6	1	7.5	2	2	2
Arg	4.9	5	4	4.9	0	0	0
Trp	0.8	1	n.d.	0.8	0	n.d.	n.d.

a) Composition according to amino acid analysis. The values for SP II A and SP II B are from ref. (3).

b) Composition from sequence studies.

c) Blocked N-terminal sequence of SP II A obtained by subtracting the amino acid composition of SP II B from SP II A.

expected. Three were isolated, two of which were compatible with Glu₄₉-Tyr₅₀-Arg₅₁ and Ile₅₂-Asp₅₃-Arg₅₄, respectively, as judged from amino acid analysis. The largest peptide contained 14 residues (Asp₃, Glu₁, Pro₁, Ala₁, Val₂, Ile₁, Leu₂, Phe₁, Lys₁, Arg₁) and was compatible with the sequence Leu₅₇-Arg₇₀ (T₅). Of this sequence Leu₅₇-Gln₆₇ was already known, leaving positions 68 and 69 undecided with respect to the positioning of the remaining Val and Pro. Being a tryptic peptide the arginine of this peptide must be situated in position 70.

Figure 2. Separation of cyanogen bromide cleaved SP II B on Biogel P6 (0.9 × 120 cm) in 30 per cent acetic acid.

Flow-rate 10 ml · hour⁻¹. Fractions of 1 ml were collected. The bars indicate fractions pooled for further analysis.

Digestion of the peptide Leu₅₇-Arg₇₀ with carboxypeptidase Y released Arg, Pro, Val, Gln and Ala in that order thereby defining Pro in position 69 and Val in position 68.

A comparison of the amino acid composition of CNBr-II with the determined sequence of this peptide leaves only one Val and one Gly at the C-terminal end.

Attempts to determine the C-terminal sequence by hydrolysis with carboxypeptidase Y (2) failed. The reason could be that the C-terminal amino acid was blocked or a glycine was the C-terminal amino acid. Although carboxypeptidase Y in a few cases is known to hydrolyse bonds involving glycine the general observation is that none of the carboxypeptidases (Y, A and B) hydrolyzes this type of peptide bonds. The resistance of SP II B to hydrolysis with carboxypeptidase Y can therefore be explained by placing the remaining glycyl residue as the C-terminal. The expected dipeptide Val₇₁-Gly₇₂ has been isolated by HV-electrophoresis and in an experiment in which a tryptic digest of acetylated SP II B was passed over a Biogel P6 column. The disappearance of Val after dansylation and acid hydrolysis of the dipeptide showed that glycine indeed is the C-terminal amino acid residue.

In Table II a comparison is made between the amino acid composition of SP II B determined by amino acid analysis (3) and sequencing. The discrepancies in the case of Thr, Gly and Ala can be explained by the finding that the N-terminal is »ragged«. These three amino acid residues are cleaved off during the purification procedure together with a smaller amount of Asp.

3.2. Sequence determination of SP II A

An attempt to sequence SP II A from the N-terminal end failed pointing towards a blocked amino acid residue. Amino acid analysis of SP II A showed that this protein also contained only a single methionyl residue. Cyanogen bromide cleavage would therefore give rise to only two fragments, one of which was blocked at the N-terminus. Hence, it was found safe and even advantageous to apply the CNBr-cleavage product directly to the »cup« and sequence CNBr-II as far as possible. A sequence of 16 residues was

determined and found to be identical to the corresponding sequence in SP II B (Glu₄₉-Asn₆₄). Along with the main sequence a minor sequence, amounting to less than 10% of the major one, was obtained. From step nine this sequence was found to be identical with that of SP II B starting from the N-terminus. It thus appears, that the sequence of SP II A extends at least eight residues beyond that of SP II B at the N-terminus. A tentative sequence for these eight residues is as follows: (Ser)-Lys-Lys-Pro-Glu-Gly(-)-Asn-. This finding is in fair agreement with the amino acid analysis according to which SP II A is ten residues longer than SP II B (see Table II). Positions 1 and 7 were empty on the HPLC-chromatograms, but would be accounted for by the two serines found in the amino acid analysis. Isolation of the large CNBr-fragment (CNBr-I) by gel filtration, followed by acetylation and cleavage with trypsin liberated the blocked N-terminal sequence of SP II A (cleavage at Arg₆, Figure 1). The amino acid composition of this peptide is seen to be in good agreement with the one obtained as the difference between the amino acid compositions of SP II A and SP II B (Table II, columns 5 and 7) apart from an additional valine residues. Cleavage of the peptide with staphylococcal protease released a nonapeptide which was sequenced in the presence of Polybrene. The following sequence was obtained: Gly-Val-Asn-Thr-Gly-Ala-Gly-Asp-, thus confirming the overlap to the N-terminal sequence of SP II B and also confirming part of the minor sequence in the sequencing of CNBr-II of SP II A. The vacant position between Gly and Asn is now identified as a valine. With this additional information the sequence of the N-terminal part of SP II A can be extended to (Glx,Ser,Ser,Val)-Lys-Lys-Pro-Glu-Gly-Val-Asn when compared with the amino acid composition of the isolated N-terminal peptide.

It has previously been shown that SP II A contains about one percent carbohydrate expressed as glucose equivalents (3). From the determined sequences it is clear that the carbohydrate must be located within the first four residues of SP II A. Of these one or both serine residues appears to be likely candidates as point(s) of attachment.

The identity of the group blocking the N-terminal residue of SP II A remains to be

determined. However, since deblocking does take place, although to a minor extent, in 70% formic acid, the possibility that the blocking group is a formyl group must be considered. It appears unlikely that an acetyl group will undergo hydrolysis under the same conditions. The possibility of cleavage of a peptide bond releasing a new N-terminal cannot be excluded and would explain the discrepancy between the eight amino acid residues suggested from sequencing and the 10–11 residues found from amino acid composition studies.

The great similarity in amino acid composition between the SP II A and SP II B proteins, the observation that 45 residues in the two sequences are identical and the resistance of both to hydrolytic attack by carboxypeptidase Y together point towards the conclusion that SP II B is a fragment of SP II A produced by limited proteolysis during the purification procedure. A similar suggestion has been offered from studies with tandem crossed immunoelectrophoresis (3).

ACKNOWLEDGEMENTS

We are grateful to dr. MARTIN OTTESEN for critical evaluation of the manuscript. The excellent technical assistance of Mss. BODIL CORNELIUSSEN and LONE SØRENSEN is gratefully acknowledged.

REFERENCES

1. BEGG, G. & F. J. MORGAN: A non-volatile buffer with improved performance in automated protein sequencing. *Febs Lett.* 66, 243–245 (1976)
2. HAYASHI, R., S. MOORE & W. H. STEIN: Carboxypeptidase from yeast. *J. Biol. Chem.* 248, 2296–2302 (1973)
3. JONASSEN, I.: Characteristics of Hiproly barley I. Isolation and characterisation of two water-soluble high-lysine proteins. *Carlsberg Res. Commun.* 45, 47–58 (1980)
4. JONASSEN, I.: Characteristics of Hiproly barley II. Quantification of two proteins contributing to its high lysine content. *Carlsberg Res. Commun.* 45, 59–68 (1980)
5. KULBE, K. D.: Micropolyamide thin-layer chromatography of phenylthiohydantoin amino acids (PTH) at a subnanomolar level. *Anal. Biochem.* 59, 564–573 (1974)
6. NARITA, K., H. MATSUA & T. NAKAJIMA: In *Protein Sequence Determinations* (S. B. Needleman, ed.) pp. 42, Springer-Verlag (1975)
7. TARR, G. E., J. F. BEECHER, M. BELL & D. J. MCKEAN: Polyquarternary amines prevent peptide loss from sequenators. *Anal. Biochem.* 84, 622–627 (1978)
8. YAMADA, S. & H. A. ITANO: Phenanthrenequinone as an analytical reagent for arginine and other substituted guanidines. *Biophys. Biochim. Acta* 130, 538–541 (1966)