Complete Amino Acid Sequence of the N-Terminal Cyanogen Bromide Fragment CNBr-3 from a Microfibrillar Protein from Wool

Keith H. Gough and Adam S. Inglis
Division of Protein Chemistry, CSIRO, Parkville, Vic. 3052.

Abstract
The amino acid sequence of a 59-residue 'non-helical tail' from a low-sulfur microfibrillar protein (component 8c-1) from wool is reported. The amino acid composition shows relatively high contents of serine (17%), half-cystine (14%), proline (8%), glycine (8%) compared to the helical areas of component 8c-1. Although the composition of the cyanogen bromide fragment CNBr-3 resembles the compositions of the high-sulfur proteins extracted from wool there are no obvious sequence homologies present. The Edman degradation yields at an aspartyl-glycyl bond decreased so much that difficulties were encountered in determining the subsequent sequence.

Extra keyword: keratin.

Introduction
The low-sulfur proteins of wool (SCMKA) can be isolated by isoelectric precipitation at pH 4.4 after reduction of the disulfide bonds and subsequent alkylation with iodoacetate. Polyacrylamide gel electrophoresis of component 8, one of the three families of low-sulfur proteins, demonstrates that there are four components, 8a, 8b, 8c-1 and 8c-2 (Crewther et al. 1980). All have a similar amino acid composition being particularly rich in aspartic acid, glutamic acid and leucine. The components are helical in solution (Crewther 1976) and are thought to originate from the microfibrils of the wool fibre (Jones 1976).

The high-sulfur proteins (SCMKB) apparently come from the matrix surrounding the microfibrils (Fraser et al. 1972), are non-helical and are particularly rich in the amino acids half-cystine, serine and proline. The amino acid composition of the CNBr-3 fragment (Thompson and O'Donnell 1967), the smaller of the two cyanogen bromide fragments obtained from the low-sulfur protein component 8c-1, resembles SCMKB in that it is also rich in half-cystine, serine and proline, and suggests that it comes from a non-helical section of component 8c-1. This paper reports the amino acid sequence of the non-helical N-terminal cyanogen bromide fragment of component 8c-1 and compares it with the sequences previously obtained for the non-helical high-sulfur proteins of wool.

Methods
Protein Isolation
Component 8c-1 was isolated as described by Crewther et al. (1976), and the high molecular weight contaminants were removed by precipitation with n-propanol (Dowling and Crewther 1974). A cyanogen bromide cleavage was performed as described by Gross and Witkop (1962) and the two fragments separated on Sephadex G75 in 8 M urea buffer (Thompson and O'Donnell 1967).
**Amino Acid Analysis**

Amino acid analyses were performed as described by Gough et al. (1978) except that tryptophan was analysed after hydrolysis with either p-toluenesulfonic acid or methanesulfonic acid (Dowling et al. 1979).

**Enzymic Digestion**

The digestion with trypsin, thermolysin, pepsin and carboxypeptidase was essentially as described by Gough et al. (1978). For a chymotryptic digestion the protein (23 mg) was dissolved in 0.1 M N-ethylmorpholine acetate, pH 8.0, α-chymotrypsin (0.25 mg) added and incubated at 37°C for 3 h. The reaction was stopped by the addition of diisopropyl phosphofluoridate (2×10 μl), allowed to stand for 30 min at room temperature, and then freeze-dried. The Staphylococcus aureus protease digestion was performed in 0.05 M phosphate buffer, pH 7.8, with an enzyme to protein ratio of 1:30; digestion was carried out at 37°C for 22 h. The mixture was then freeze-dried.

**Cleavage with 2-(2-Nitrophenylsulfenyl)-3-methyl-3-bromoindolenine (BNPS-skatole)**

Cleavage with BNPS-skatole was performed as described by Fontana et al. (1973).

**Peptide Purification**

All enzymic digests and the BNPS-skatole cleavage products were fractionated on columns of Sephadex G50 superfine (300 by 1 cm) coupled to columns of G25 superfine (300 by 1 cm) and eluted with 0.01 M ammonium bicarbonate buffer containing 10% (v/v) isopropanol (Corran and Waley 1974). Mixtures of peptides after column chromatography were further purified by high-voltage paper electrophoresis (HYPE) in Varsol-cooled tanks (Ambler 1963) at each pH 9–0, 6–5 or 4–1 depending on the nature of the peptides.

**Sequence Determination**

Manual sequencing was carried out essentially according to Peterson et al. (1972). The newly released N-terminal amino acid was identified by the dansyl method (Gray 1967) on polyamide sheets (Woods and Wang 1967) except that toluene replaced the benzene in the second solvent (Furth et al. 1974). Automated amino acid sequencing was performed as described by Inglis and Burley 1977. The phenylthiohydantoins (PTH) were identified by thin-layer chromatography (t.l.c.) on aluminium-backed silica gel plates (Inglis and Nicholls 1973), or by back-hydrolysis of the PTH by HI (Inglis et al. 1971), or by high-performance liquid chromatography (Zimmermann et al. 1977).

**Amide Determination**

Where possible the PTH-derivatives of glutamine–glutamic acid and asparagine–aspartic acid were identified by t.l.c. Sometimes the method of Offord (1966) was used to identify amides.

**Dilute HCl Cleavage**

To cleave the CNBr-3 fragment at the Asp-Gly bond (residues 52 and 53) the sample was heated at 108°C in dilute HCl, pH 2–0, for 2 h in vacuo (Inglis et al. 1979). Peptides were cleaved using 0.03 M HCl containing β-mercaptoethanol, ratio 1:2000, at 106°C for 2 h in vacuo (Keutmann et al. 1973).

**Results and Discussion**

O’Donnell and Thompson (1968) and O’Donnell (1969) partially sequenced the CNBr-3 fragment from their preparation of component 8, but were hampered by its gross heterogeneity and suggested that component 8 consisted of a family of similar proteins. Subsequent work has shown that there are in fact four proteins in the component 8 family, 8a, 8b, 8c-1 and 8c-2 (Crewther et al. 1980). Component 8c-1 was chosen for the primary structure determination because it could be purified
relatively easily and there were no problems with heterogeneity in the sequence determination of the CNBr-3 fragment. This 59-residue N-terminal fragment (molecular weight 6750) represents approximately one-seventh of the total molecule (molecular weight 48 000).

N-Ac—SER PHE ASN PHE CMCyS LEU PRO ASN LEU SER PHE ARG SER SER CMCyS SER SER ARG PRO CMCyS

T1  
\[ \begin{array}{c}
\text{TIB} \\
\text{TID} \\
\text{TIE} \\
\text{C1} \\
\text{C1B} \\
\text{C1A} \\
\text{C2} \\
\text{T2} \\
\text{T2A} \\
\text{C3} \\
\text{T3} \\
\text{T4} \\
\text{C3A} \\
\text{P1} \\
\text{P2B} \\
\text{P2C} \\
\text{SA} \\
\text{Dilute HCl}
\end{array} \]

Fig. 1. Amino acid sequence of the CNBr-3 fragment. The positions of cleavage by trypsin (T), chymotrypsin (C), pepsin (P), and staphylococcal protease (SA), are shown by vertical arrows. Letters after each number indicate that the original peptide was subsequently redigested with a second enzyme. Residues underlined by an arrow were identified by sequencing (→ dansyl method or using sequenator; ← carboxypeptidase) and those not underlined were not identified. Continuous underlining indicates residues identified from the amino acid composition.

The three tryptic peptides obtained (Fig. 1) could be easily ordered. T1 was found to have a blocked N-terminus, as does the CNBr-3 fragment, and was thus the N-terminal peptide. T3 contained HSer indicating that this is the C-terminal peptide. The amino acid composition of the N-terminal tryptic dodecapeptide T1 (Fig. 1) corresponded with one of the N-terminal peptides found by O’Donnell (1969). O’Donnell and Thompson (1968) established that the N-terminal serine of the CNBr-3 fragment was acetylated. Peptide T1 was further digested with chymotrypsin and the isolated peptides were sequenced manually (Fig. 1). Chymotryptic and pepsic digests of CNBr-3 fragment (Fig. 1) allowed us to isolate peptides that confirmed the sequences obtained by tryptic digestion. The chymotryptic peptide C1 (Fig. 1) had the amino acid composition Asp1, Ser1, Phe2, and when further digested with thermolysin gave two bands on HVPE, one of which contained the peptide with the
sequence Phe-Asn-Phe. Material from a second ninhydrin-negative, starch–Cl₂-positive band gave no dansyl derivative, but analysed for serine only after hydrolysis with HCl and was thus presumably the N-terminal acetyl serine residue. The analysis of the tryptic peptide T1 by mass spectrometry (Morris and Dell 1975) determined that the N-terminal serine residue was acetylated as well as confirming the sequence of the first 10 residues of peptide T1. The amino acid composition of these peptides are given in Tables 2–4.*

The peptide bonds between residues 12–39 (27 residues) (Fig. 1) could not be cleaved by any of the enzymes trypsin, chymotrypsin, thermolysin, pepsin and S. aureus protease. This region is enriched in half-cystine, proline and serine. Apparently the primary and secondary structure imposed by these amino acids prevents enzymic digestion. Cleavage with thermolysin or trypsin, for example, was unsuccessful because in each case where a hydrophobic residue (Leu, Val or Ile) or a basic residue (Arg) is present, a prolyl residue follows them (Fig. 1). The sequence of residues 12–39 (Fig. 1) was established by the automatic sequenator.

The bond between residues 45–46 (Trp-Phe) was susceptible to digestion by pepsin, chymotrypsin and thermolysin and thus an overlap could not be established directly. Cleavage was therefore carried out with BNPS-skatole, usually specific for tryptophan, and the C-terminal peptide with the amino acid composition Ly50,8, CMCyso,4, Asp2,0, Thr1,1, Ser1,4, Glu2,9, Gly2,1, Phe1,8, HSer.lactone(+) was isolated. The partial sequence, determined both manually and automatically, was Phe-CMCys-Glu-Gly-Ser-Phe-Asp-Gly. This sequence was further supported by data from a thermolytic digest of tryptic peptide T2 (Fig. 1) and peptic digestion of CNBr-3 (Fig. 1), and is consistent with the presence of a Trp-Phe bond at residues 45–46.

The bond Asp-Gly, at positions 52–53, showed the characteristic problem that can occur with this bond (Kohlmiller and Howard 1979); we could not, in most cases, sequence past the aspartyl residue (position 52) and concluded that this was due to the cyclization of the aspartic acid to the imide form (Konigsberg 1972). In some cases a very low yield of the subsequent sequence Gly-Asn-Glu was obtained. A peptide with the amino acid composition corresponding to residues 49–55 was isolated from a S. aureus protease digest of CNBr-3 fragment. The composition Asp1,8, Ser1,0, Glu1,1, Gly1,8, Phe1,0 and a net charge of −2 (Offord 1966) indicated the presence of one amide. When this peptide was sequenced manually the sequence Gly-Ser-Phe-Asp-Gly-Asn-X-Glu was obtained, i.e. after the asparaginyl residue an apparent gap appeared in the sequence but if sequencing continued the expected glutamic acid residue was detected. If the S. aureus peptide was cleaved with dilute HCl (Keutmann et al. 1973) a peptide with the amino acid composition Asp1,0, Glu1,0, Gly1,0 and with a net charge of −1 was isolated, but again the manual sequence procedures apparently did not yield a residue at position 55 and gave the sequence Gly-Asn-X-Glu. This question was resolved by cleavage of the CNBr-3 fragment with dilute HCl (Inglis et al. 1979) and sequence determination of the resultant mixture using the automatic sequenator established the sequence as Gly-Asn-Glu-Lys-Glu. The finding of a ‘gap’ between Asn and Glu seems to be peculiar to the manual sequencing method.

* Tables 2–4 have been lodged as an Accessory Publication, copies of which are obtainable on request from the Editor-in-Chief, Editorial and Publications Service, CSIRO, 314 Albert Street, East Melbourne, Vic. 3002.
The amino acid composition of the CNBr-3 fragment (Table 1) shows similarities to that of the high-sulfur proteins (Gillespie 1962) in that it is rich in half-cystine, proline, serine and glycine but in contrast it also contains higher amounts of phenylalanine and aspartic acid. When compared to component 8 and the helical fragments (Fraser et al. 1972), CNBr-3 is deficient in glutamic acid, lysine and leucine. The four amino acids, half-cystine, proline, serine and glycine, are known to be helix-breakers (Chou and Fasman 1974) and not unexpectedly, a zero value was obtained for the \( \alpha \)-helix content of the CNBr-3 fragment whereas component 8 had a value of 50% (O’Donnell 1969). Thus, the CNBr-3 fragment would correspond to a ‘non-helical tail’ described by Crewther and Harrap (1967). Although the amino acid composition of the CNBr-3 fragment resembles that of the high-sulfur proteins (Gillespie 1962) there are no obvious sequence homologies between the CNBr-3 fragment and the high-sulfur proteins SCMKB IIIA and SCMKB 2 (Crewther 1976). Repeating units are found in SCMKB IIIA and SCMKB 2 with the half-cystine residues spaced by intervals of 5, 10 or 15 residues (Crewther 1976) with some sections of the molecule devoid of half-cystine. But, in the CNBr-3 fragment the half-cystine residues are at intervals of approximately 5 or 10 residues and spread throughout the molecule with no other sign of a repeating unit.

Table 1. Amino acid composition of the cyanogen bromide fragment CNBr-3

Values expressed as number of residues per molecule. The values for threonine and serine were extrapolated to zero time and the values for isoleucine and valine were from a 72-h hydrolysis. Tryptophan was determined from separate hydrolysates with either \( p \)-toluenesulfonic or methanesulfonic acid

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Sequence determination</th>
<th>No. of residues</th>
<th>Amino acid</th>
<th>Sequence determination</th>
<th>No. of residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys</td>
<td>(1)</td>
<td>1·0</td>
<td>Pro</td>
<td>(5)</td>
<td>4·7</td>
</tr>
<tr>
<td>Arg</td>
<td>(2)</td>
<td>2·1</td>
<td>Gly</td>
<td>(5)</td>
<td>4·8</td>
</tr>
<tr>
<td>Trp</td>
<td>(1)</td>
<td>0·7</td>
<td>Ala</td>
<td>(2)</td>
<td>2·0</td>
</tr>
<tr>
<td>CM Cys</td>
<td>(8)</td>
<td>7·2</td>
<td>Val</td>
<td>(2)</td>
<td>2·0</td>
</tr>
<tr>
<td>Asp</td>
<td>(7)</td>
<td>6·3</td>
<td>Ile</td>
<td>(1)</td>
<td>1·0</td>
</tr>
<tr>
<td>Thr</td>
<td>(3)</td>
<td>2·8</td>
<td>Leu</td>
<td>(3)</td>
<td>3·2</td>
</tr>
<tr>
<td>Ser</td>
<td>(10)</td>
<td>7·9</td>
<td>Phe</td>
<td>(5)</td>
<td>4·4</td>
</tr>
<tr>
<td>Glu</td>
<td>(3)</td>
<td>3·4</td>
<td>HSer</td>
<td>(1)</td>
<td>0·9</td>
</tr>
</tbody>
</table>

\(^A\) Total number of residues = 59.

The non-helical tail from the low-sulfur protein 8c-1 is the first such segment to be isolated and completely sequenced, but there is evidence that another segment similar in composition is located near the C-terminus of component 8c-1 (Dowling, unpublished data). As yet, there is insufficient sequence data available for further comparisons with other low-sulfur proteins but it will be of interest to see whether the components 5 and 7 under study have these non-helical tails in both the C- and N-terminal regions as appears to be the case with component 8c-1.

Acknowledgments

We sincerely thank Mr L. M. Dowling for supplying the protein, Mr M. Rubira for technical assistance, and Drs Morris and Dell for the mass spectrometric analysis.
References


Manuscript received 21 December 1979, accepted 17 March 1980