# Bovine Conglutinin Is a Collagen-like Protein†

Alvin E. Davis, III,* and Peter J. Lachmann

ABSTRACT: Conglutinin is a bovine plasma protein which is relatively large and asymmetric with elevated contents of glycine and, to some extent, proline. Although its physiologic function is unknown, conglutinin is known to bind, in the presence of calcium, to yeast cell walls and to the solid-phase-inactivated third component of complement. On sodium dodecyl sulfate–polyacrylamide gel electrophoresis under reducing conditions, isolated conglutinin appeared to consist of a single polypeptide chain ($M_r$ 48 000). Unreduced conglutinin consisted of a single stained band with an apparent molecular weight of approximately 300 000. Cross-linking experiments with glutaraldehyde and dimethyl suberimidate suggested that this $M_r$ 300 000 molecule consists of six of the disulfide-linked polypeptide chains. Amino acid composition revealed hydroxylysine and hydroxyproline together with elevated glycine and proline contents. Digestion of reduced, alkylated conglutinin with bacterial collagenase resulted in formation of a precipitate which consisted of an $M_r$ 24 000 peptide which was digested to $M_r$ 21 000 with large quantities of collagenase. These peptides contained less glycine, proline, hydroxylysine, and hydroxyproline than did the intact protein. The supernatant from this digestion mixture was, however, enriched in these four amino acids, with glycine making up nearly one-third of the total. Prolonged digestion with pepsin at 37 °C resulted in an $M_r$ 20 000 peptide which was enriched in glycine, proline, hydroxyproline, and hydroxylysine. Amino-terminal sequence analysis showed that the glycine-X-Y repeating sequence begins at residue 26. The conglutinin subunit polypeptide chain, thus, consists of an amino-terminal 25 amino acid globular protein-like segment followed by a collagen-like segment ($M_r$ ~20 000), with an $M_r$ 21 000 globular protein-like domain at the carboxy terminus.

Conglutinin is a bovine plasma protein that binds, in the presence of calcium, to yeast cell walls (*Saccharomyces cerevisiae*) or zymosan and to surface-bound-inactivated third component of complement (iC3b) Lachmann, 1967; Lachmann & Müller-Eberhard, 1968). The agglutination by bovine serum of erythrocytes reacted with antibody and complement was first observed by Bordet & Gay (1906). This phenomenon was termed conglutination, and the component in bovine serum responsible for the effect was called conglutinin (Bordet & Streng, 1909). Lachmann & Müller-Eberhard (1968) showed that conglutinin bound to solid-phase C3 which had reacted with a component in serum which they termed conglutinogen activating factor. This factor (subsequently named the C3b inactivator and now known as factor I) has since been shown to be a serine protease which, in the presence of its cofactor (factor H), initially cleaves the $\alpha'$ chain of C3b at two sites (Nagasawa & Stroud, 1977; Pangburn et al., 1977; Harrison & Lachmann, 1980; Davis & Harrison, 1982). These cleavages release an $M_r$ 3000 peptide, leaving an amino-terminal $M_r$ 68 000 peptide and a carboxy-terminal $M_r$ 43 000 peptide, both of which remain disulfide bonded to the C3 $\beta$ chain. This cleavage product, iC3b, is the molecular form of surface C3 to which conglutinin binds (Lachmann & Müller-Eberhard, 1968; Linscott et al., 1978). There is no binding of conglutinin to C3b prior to these cleavages, and there also is no binding to fragments resulting from further degradation of iC3b. Conglutinin, as detected by the agglutination of iC3b-coated sheep erythrocytes, has been demonstrated only in the sera of Bovidae (Lachmann, 1967). Sera of numerous other mammals, including mouse, rat, rabbit, guinea pig, horse, cat, dog, pig, sheep, and human, have all been tested and found to contain no conglutinin activity. The possibility that a conglutinin-like molecule which does not bind to solid-phase iC3b is present in other mammals has not yet been adequately examined.

The physiochemical characteristics of conglutinin were examined in some detail during the 1960's, as were its interactions with both yeast and cell-bound C3 (Lachmann, 1967; Lachmann & Coombs, 1965). Although conglutinin frequently has been used to identify and quantitate immune complexes (Eisenberg et al., 1977; Johny et al., 1980; Manca et al., 1980), there has been virtually no further study of its structural characteristics. On the basis of hydrodynamic criteria, Lachmann & Richards (1964) calculated an apparent $M_r$ of 750 000; the data also indicated that conglutinin was an extremely asymmetric molecule. Electron microscopic examination confirmed this finding and showed elongated rods of approximately 450 Å in length (Lachmann & Coombs, 1965). The amino acid composition of conglutinin revealed a relatively high content of glycine and, to some extent, proline (Lachmann & Coombs, 1965).

The above findings, which are extremely unusual for a plasma protein, prompted us to reinvestigate the properties of conglutinin, with initial emphasis on some of the features of its structure and composition. We report herein that conglutinin, on the basis of its amino acid composition and initial amino acid sequence analysis, is, like the complement component C1q, a collagen-like protein.

## Materials and Methods

Baker's yeast was obtained from the Distiller Co. Sepharose CL6B was purchased from Pharmacia. Dithiothreitol was obtained from Calbiochem and iodoacetamide from Kodak. All sequencing reagents and solvents were from Beckman. Methanol and dimethyl suberimidate were from Pierce, and acetyl chloride was from Sequemat. Sodium dodecyl sulfate ($NaDodSO_4$),[1] acrylamide, *N,N'*-methylenebis(acrylamide),

† From the Division of Immunology and Nephrology, Department of Medicine, Children's Hospital Medical Center, and the Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, and the Mechanisms in Tumour Immunity Unit, Medical Research Council Centre, Cambridge CB22QH, England. *Received October 10, 1983.* Supported in part by U.S. Public Health Service Grants AI 16392 and AI 17896. Parts of this work were performed during a fellowship awarded to A.E.D. from the John Simon Guggenheim Memorial Foundation.

* Address correspondence to this author at the Children's Hospital Medical Center, Boston, MA 02115.

0006-2960/84/0423-2139$01.50/0 

and *N,N,N',N'*-tetramethylethylenediamine were from Bio-Rad. Acetonitrile was from Burdick and Jackson. Glutaraldehyde was from British Drug House and was redistilled as described by Gillett & Gull (1972). Pepsin was from Worthington, and purified collagenase was from advance Biofactures. *N*-Ethylmaleimide and phenylmethanesulfonyl fluoride were from Sigma.

For reduction and alkylation, lyophilized protein was dissolved in 50 mM Tris-HCl containing 6.0 M guanidine, pH 8.0. Dithiothreitol was added to 20 mM; the tube was flushed with nitrogen, sealed, and incubated at 37 °C for 1 h. Iodoacetamide (50 mM) was then added, and the sample was incubated for 15 min in the dark at room temperature. Alkylation was terminated by the addition of excess mercaptoethanol, and the reduced and alkylated protein was dialyzed against water and lyophilized. $NaDodSO_4$–polyacrylamide gel electrophoresis was performed according to Laemmli (1970). Samples were incubated in sample buffer in the presence or absence of 0.1 M 2-mercaptoethanol at 100 °C for 2 min. Following electrophoresis, gels were stained with Coomassie brilliant blue R250.

Amino acid compositions were performed with a Beckman 121 MB amino acid analyzer. Samples were hydrolyzed by incubation of reduced and alkylated protein at 110 °C in 6 NHCl under reduced pressure for 24, 48, and 72 h. Serine and threonine concentrations were determined by extrapolation to zero time. Tryptophan was not determined. Automated Edman degradations were performed with a Beckman 890 C sequencer modified with a cold trap. A 0.1 M Quadrol program was used (Brauer et al., 1975). Conversion was performed with methanolic HCl (1 part acetyl chloride and 7 parts methanol, 65 °C, 10 min). Phenylthiohydantoin derivatives were identified by high-performance liquid chromatography using a Zorbax ODS column (DuPont Instruments) equilibrated in 0.01 M sodium acetate, pH 5.5, containing 20% acetonitrile and developed with an acetonitrile gradient (Zalut et al., 1980).

Digestion with pepsin was performed by incubation of lyophilized protein with 1.75% pepsin (w/w) in 100 mM sodium acetate, pH 4.45. Samples were incubated at 37 °C for varying periods of time; incubation was terminated by adjustment of the pH to 8.0. The major large peptide resulting from this digestion was separated from other peptides by gel filtration on Sepharose 6B in 0.05 M Tris–0.2% $NaDodSO_4$, pH 8.0, followed by high-performance liquid chromatogrphy on an Altex ultrasphere column in 0.1% trifluoroacetic acid with an acetonitrile gradient.

Digestion with collagenase was performed by incubation of reduced and alkylated protein with varying quantities of collagenase (Advance Biofactures Corp., chromatographically purified) (Miller & Udenfried, 1970) for 24 h at 37 °C in 25 mM Tris-HCl–10 mM $CaCl_2$, pH 7.4. Some digestions were done in the presence of *N*-ethylmaleimide (0.25 mM) and phenylmethanesulfonyl fluoride (10 μM). These appeared identical with digests performed in the absence of these reagents. The quantity of collagenase used varied from 5.6 to 70 units per 200 μg of protein. A unit is the amount required to solubilize 1 μmol of leucine equivalents (using a colorimetric ninhydrin method and expressing amino acids liberated as micromoles of leucine) after 24-h incubation at 37 °C with undenatured collagen. Cyanogen bromide digestion was done by incubation of reduced and alkylated protein with cyanogen bromide (~100-fold molar excess over total methionine) in 70% formic acid. The digest was chromatographed on Sepharose CL6B in 0.05 M Tris-HCl–0.2% $NaDodSO_4$, pH 8.0, and on hydroxylapatite equilibrated in 10 mM sodium phosphate–0.1% $NaDodSO_4$, pH 6.4, and eluted with a phosphate gradient. The amino-terminal amino acid sequence was determined for one cyanogen bromide peptide which had an apparent $M_r$ of ~20 000 on $NaDodSO_4$–polyacrylamide gel electrophoresis. On hydroxylapatite, this peptide eluted very early with the phosphate gradient.

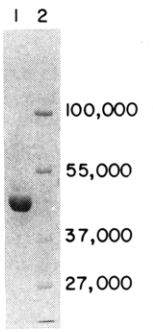


FIGURE 1: $NaDodSO_4$–polyacrylamide gel electrophoresis of conglutinin. Samples were prepared in the presence of 2-mercaptoethanol as described under Materials and Methods. A 10% polyacrylamide gel was used. Track 1, conglutinin; track 2, markers with molecular weights as indicated.

Conglutinin was isolated essentially as described previously by Lachmann (1962, 1967). Briefly, fresh bovine serum (0.5–3 L) was incubated at 56 °C for 30 min and was then incubated with isolated yeast cell walls at 4 °C for 2 h with 2 mM $CaCl_2$ added to the serum. The yeast was then washed 3 times with 0.02 M sodium–potassium phosphate and 0.15 M NaCl, pH 7.2, and the conglutinin was eluted from the yeast by incubation with the above buffer containing 10 mM disodium ethylenediaminetetraacetate for 10 min at room temperature. The conglutinin was then precipitated as a euglobulin by dialysis against 10 mM sodium–potassium phosphate, pH 5.4. The euglobulin precipitate was collected by centrifugation, washed 3 times with euglobulin precipitation buffer, and dissolved in 0.02 M sodium–potassium phosphate and 0.15 M NaCl, pH 7.2. The yeast absorption and euglobulin precipitation were then each repeated. The resulting preparation was then chromatographed on Sepharose CL6B in 0.05 M Tris-HCl–0.2% $NaDodSO_4$, pH 8.0. $NaDodSO_4$ previously was shown to have no effect on the detection of conglutinin activity (Lachmann & Coombs, 1965).

Sheep erythrocytes with iC3b on their surface were prepared by incubation of 10% sheep erythrocytes sensitized with anti-Forssman antibody with an equal volume of C6-deficient rabbit serum for 10 min at 37 °C. Conglutinin activity was detected during isolation by incubation at 4 °C of 50 μL of conglutinin-containing sample with 50 μL of 1% sheep erythrocytes prepared as described above. The conglutinin titer was the highest dilution showing agglutination.

## Results

*Subunit Structure of Conglutinin.* Conglutinin, isolated as described under Materials and Methods, showed only a single minor contaminant ($M_r$ ~45 000) prior to gel filtration (Figure 1). This protein was readily separated by chromatography on Sepharose CL6B in the presence of 0.2% $NaDodSO_4$ (Figure 2). Gel filtration, in the absence of $NaDodSO_4$, resulted in incomplete removal of this contaminant from the conglutinin preparation. The contaminant was present in varying quantities, relative to conglutinin, in different prepa-

[1] Abbreviations: $NaDodSO_4$, sodium dodecyl sulfate; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.

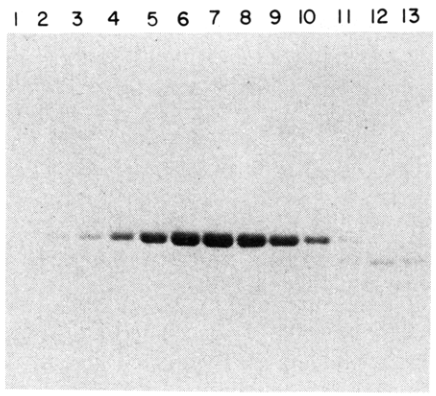


FIGURE 2: Conglutinin isolation. $NaDodSO_4$–polyacrylamide gel electrophoresis of fractions from the Sepharose CL6B elution profile. Samples were prepared in the presence of 2-mercaptoethanol as described under Materials and Methods. The fractions shown in tracks 3–9 were pooled for the conglutinin preparation.

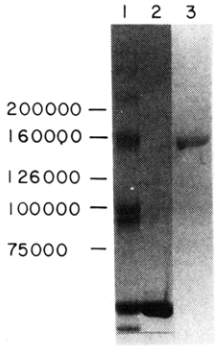


FIGURE 3: $NaDodSO_4$–polyacrylamide gel electrophoresis of cross-linked conglutinin. Track 1, conglutinin polypeptide chains cross-linked with glutaraldehyde as described under Results and subjected to electrophoresis in the presence of 2-mercaptoethanol; track 2, conglutinin in the presence of 2-mercaptoethanol; track 3, conglutinin in the absence of 2-mercaptoethanol. Tracks 1 and 2 were from the same gel with a 5% polyacrylamide concentration; track 3 was from a gel with a 4% polyacrylamide concentration.

rations. In addition, there was no decrease in conglutinin specific activity following removal of the contaminant. It is therefore unlikely that this peptide is a noncovalently associated portion of the conglutinin molecule. The apparent molecular weight of conglutinin under reducing conditions on $NaDodSO_4$–polyacrylamide gel electrophoresis was 48 000 (Figure 1). $NaDodSO_4$–polyacrylamide gel electrophoresis of unreduced conglutinin showed a single band with an apparent molecular weight of approximately 300 000 (Figure 3, track 3). This suggested the possibility that conglutinin consisted of six polypeptide chains of equal size. In order to attempt to confirm this estimate, conglutinin polypeptide chains were cross-linked with glutaraldehyde (0.1%, 15 min, 37 °C, pH 7.2) and dimethyl suberimidate (200 μg/mL, 60 min, 37 °C, pH 9.0). The cross-linked complex was then analyzed by $NaDodSO_4$–polyacrylamide gel electrophoresis under reducing conditions (Figure 3). Cross-linked conglutinin showed a doubling of the more anodal bands, which is probably due to intramolecular cross-linking (Ji, 1983). A total of six sets of bands were visible. Identical results were obtained with glutaraldehyde and with dimethyl suberimidate, except that the efficiency of cross-linking was less with dimethylsuberimidate. The most anodal polypeptide band had the same mobility as reduced conglutinin, while the most cathodal had the same mobility as unreduced conglutinin.

*Amino Acid Composition.* The amino acid composition of conglutinin is shown in Table I. As is apparent, there are high

Table I: Amino Acid Compositions[a]

| amino acid | conglutinin | collagenase supernatant | collagenase precipitate | pepsin $M_r$ 20 000 peptide |
|---|---|---|---|---|
| SCM-Cys[b] | 2.0 | 0.9 | 1.9 | 0.6 |
| Hyp | 3.6 | 8.4 | 2.1 | 8.7 |
| Asp | 8.6 | 4.6 | 8.5 | 4.9 |
| Thr | 4.3 | 3.7 | 4.1 | 3.5 |
| Ser | 6.1 | 4.3 | 6.3 | 5.1 |
| Glu | 11.6 | 7.5 | 14.3 | 8.6 |
| Pro | 8.8 | 13.3 | 6.0 | 9.8 |
| Gly | 18.6 | 27.2 | 13.7 | 30.7 |
| Ala | 8.1 | 6.3 | 8.5 | 6.4 |
| Val | 4.2 | 3.0 | 5.3 | 2.2 |
| Met | 2.0 | 2.1 | 2.1 | 0.5 |
| Ile | 2.3 | 2.1 | 3.1 | 2.1 |
| Leu | 4.9 | 2.5 | 6.9 | 2.6 |
| Tyr | 1.5 | 0.1 | 2.0 | 0.1 |
| Phe | 2.6 | 1.3 | 3.7 | 2.2 |
| His | 1.0 | 0.7 | 1.0 | 0.8 |
| Hyl | 2.4 | 6.8 | 1.5 | 6.9 |
| Lys | 3.4 | 0.9 | 4.8 | 1.5 |
| Arg | 4.0 | 4.3 | 4.2 | 2.8 |

[a] Expressed as residues per 100 residues. [b] SCM-Cys is *S*-(carboxymethyl)cysteine.

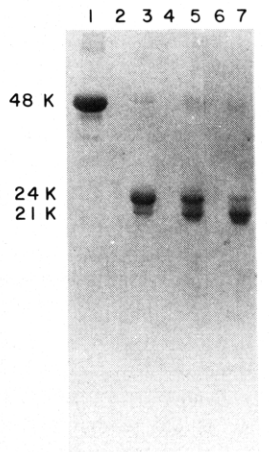


FIGURE 4: Collagenase digestion of conglutinin. The supernatants and washed precipitates from the collagenase digests (as described under Materials and Methods and under Results) were subjected to $NaDodSO_4$–polyacrylamide gel electrophoresis under reducing conditions. Track 1, conglutinin; tracks 2 and 3, the supernatant and precipitate, respectively, following digestion with 5.6 units of collagenase; tracks 4 and 5, the supernatant and precipitate, respectively, following digestion with 28 units of collagenase; tracks 6 and 7, the supernatant and precipitate, respectively, following digesiton with 70 units of collagenase. A 10% polyacrylamide gel was used.

contents of both glycine and proline. In addition, hydroxyproline and hydroxylysine were both detected in significant quantities.

*Digestion with Collagenase.* Reduced and alkylated conglutinin was digested with collagenase as described under Materials and Methods. During digestion with collagenase, a precipitate formed which was separated from the supernatant by centrifugation and was then washed 3 times with 25 mM Tris-HCl–10 mM $CaCl_2$, pH 7.4. The precipitate and supernatant from each digestion were then evaluated by $NaDodSO_4$–polyacrylamide gel electrophoresis (Figure 4). In the supernatant (tracks 2, 4, and 6), no stainable protein bands were visible. The precipitates (tracks 3, 5, and 7), however, show two major polypeptide bands, with apparent molecular weights of 24 000 and 21 000. The $M_r$ 24 000 peptide is more prominent with the lower concentration of collagenase, while the $M_r$ 21 000 peptide is more prominent with the higher concentration. Table I shows the amino acid composition of

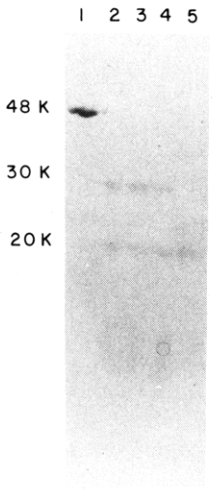


FIGURE 5: $NaDodSO_4$–polyacrylamide gel electrophoresis of conglutinin digested with pepsin. Digestion was performed as described under Materials and Methods, and samples were removed at the indicated times. Track 1, time zero; track 2, 2 h; track 3, 3 h; track 4, 4 h; track 5, 16 h. A 10% polyacrylamide gel was used, and samples were subjected to electrophoresis under reducing conditions.

the supernatant and precipitate from a collagenase digest performed with the same quantity of collagenase as in the digest shown in Figure 4, tracks 6 and 7. On $NaDodSO_4$–polyacrylamide gel electrophoresis, this digest appeared identical with that shown in tracks 6 and 7, except that there was somewhat more undigested conglutinin remaining in the precipitate (data not shown). As shown in Table I, the supernatant from the digest was enriched in both glycine and proline, with glycine making up nearly one-third of the amino acids present. In addition, both hydroxylysine and hydroxyproline were enriched as compared with the intact protein. The composition of the precipitate, on the other hand, showed a decrease in both glycine and proline, as compared with the intact protein, and is thus more similar to the composition of a globular protein. Hydroxylysine and hydroxyproline were both decreased, but significant quantities of each remained detectable.

*Digestion with Pepsin.* Native conglutinin was digested with 1.75% pepsin as described under Materials and Methods. Figure 5 shows a $NaDodSO_4$–polyacrylamide gel of aliquots removed from the digest at several time points. At earlier times, two major peptides were visible, with apparent molecular weights of 28 000 and 20 000. By 18 h, only the $M_r$ 20 000 peptide remained. Also, at earlier times, several faint smaller peptides were visible, which do not reproduce well in the figure. The 18-h digestion mixture was reduced and alkylated, and the $M_r$ 20 000 peptide was separated from the smaller contaminating peptides by gel filtration and high-performance liquid chromatography, as described under Materials and Methods. Its composition is shown in Table I. This peptide shows enrichment in glycine, proline, hydroxyproline, and hydroxylysine. It, thus, has a composition very similar to collagens.

*Amino-Terminal Amino Acid Sequence.* The amino-terminal amino acid sequence of conglutinin is shown in Figure 6. The sequence was determined with 50 nmol of reduced and alkylated protein. The repetitive yield was 95%, on the basis of the alanine residues at positions 1, 12, and 14, the glutamic acid residues at positions 2 and 24, and the methionine residues at positions 3 and 19. This sequence is of interest in that a cysteine residue is present at position 6 and that a repeating glycine-X-Y sequence begins at residue 26. Unfortunatley, no identifications could be made at positions

```
 1  Ala-Glu-Met-Thr-Cys-Phe-Ser-Gln-Lys-Ile-
11  Leu-Ala-Asn-Ala-His-Thr-Leu-Val-Met-His-
21  Ser-Pro-Leu-Glu-Ser-Gly-Leu-Hyp-Gly- X -
31   X -Gly
```

FIGURE 6: Amino-terminal amino acid sequence of conglutinin.

```
Gly-Pro-Thr-Gly-Thr-Pro-Gly-Pro- X -Gly-Glu-Thr-Gly
```

FIGURE 7: Amino-terminal amino acid sequence of one conglutinin cyanogen bromide peptide (CBIB).

30 and 31. However, the presence of glycine at positions 26, 29, and 32, together with a hydroxyproline at position 28, strongly indicates the beginning of a region of a repeating triplet sequence. Amino acid sequence analysis of the cyanogen bromide peptides of conglutinin is currently in progress. The one cyanogen bromide peptide which has thus far been analyzed confirms that conglutinin contains the glycine-X-Y repeating sequence (Figure 7). This amino-terminal sequence was determined with 20 nmol of peptide; the repetitive yield was 93%.

## Discussion

In retrospect, earlier studies suggested very strongly that bovine conglutinin might be a collagen-like protein. Both the electron microscopy showing an elongated rodlike molecule and the previous amino acid composition showing elevated glycine and proline contents (Lachmann & Coombs, 1965) were consistent with this possibility. The data presented in this paper indicate that conglutinin is in fact a collagen-like protein and that it contains a region or regions of repeating glycine-X-Y triplet sequence. The amino acid composition, which agrees fairly closely with that reported by Lachmann & Coombs (1965), shows an elevated glycine and proline content together with hydroxyproline and hydroxylysine in quantities similar to those found in collagens. The results of the collagenase and pepsin digestions support and extend the composition data. Nearly all Y–gly peptide bonds in the Gly-X-Y sequence of collagen are hydrolyzed by collagenase. The finding that the $M_r$ 48 000 conglutinin polypeptide is degraded to an $M_r$ 21 000 collagenase-resistant peptide is an extremely strong indication that conglutinin contains a region or regions with a total molecular weight of approximately 20 000–30 000 which consists of the triplet sequence. The amino acid compositions of the resistant peptide and the small peptides released by collagenase are in agreement with this interpretation. There are two likely possible explanations for the fact that the collagenase-resistant peptide still contained hydroxyproline and hydroxylysine. One is that the peptide contains a region of collagenase-resistant collagen-like sequence, which is consistent with the finding that a slightly larger peptide is also seen in the digests, particularly at lower collagenase concentrations. The second is that some undigested conglutinin remained with $M_r$ 21 000 peptide, which in fact was also true.

Digestion of conglutinin with pepsin results in a resistant peptide with a molecular weight of approximately 20 000, the composition of which is collagen-like. It consists of nearly one-third glycine and is enriched in hydroxylysine, hydroxyproline, and proline. These data agree with the previous data and further indicate that the collagen-like portion of the molecule is in one region, or at least is primarily in one region. It is also of interest to note that previous data showed that conglutinin activity was very resistant to digestion with pepsin (Lachmann & Coombs, 1965). This may suggest that the binding of conglutinin to iC3b is via the collagen-like portion

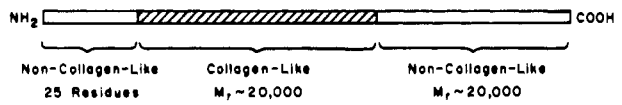


FIGURE 8: Proposed structure of the conglutinin subunit polypeptide.

of the molecule. Current studies are evaluating this possibility.

The amino-terminal sequence of conglutinin, together with the above data, indicates that the probable overall structure of the conglutinin subunit polypeptide chain is as shown in Figure 8. The finding of a single stainable band on NaDodSO$_4$–polyacrylamide gel electrophoresis together with the finding of a single amino-terminal sequence indicates that the conglutinin polypeptide chains are identical. There is a 25-residue ($M_r$ ~3400) amino-terminal segment with a globular protein-like composition. This is followed by a collagen-like region beginning at residue 26 which has a molecular weight of ~20 000. The true size of this segment in fact is probably much smaller ($M_r$ 12 000–15 000) since NaDodSO$_4$–polyacrylamide gel electrophoresis overestimates the molecular weight of collagen-like proteins (Furthmayer & Timpl, 1971). At the carboxy terminus of the polypeptide chain is an $M_r$ 21 000 segment with a globular protein-like composition. Although it is possible that a small globular, but pepsin-resistant, region might be contained within the collagen-like region or that a small collagenase-resistant collagen-like segment might be contained within the non-collagen-like region, these both seem unlikely.

Aside from collagens themselves, only two other proteins, C1q and acetylcholinesterase, contain regions of collagen-like sequence. Conglutinin has some obvious similarities to C1q, but also striking differences. The structure of the polypeptide chain of conglutinin, described above, is very similar to that of C1q, except that in C1q the amino-terminal non-collagen-like region is much shorter (Reid, 1974, 1976). The C1q polypeptide chains themselves are also smaller. The C1q chains, like conglutinin, have a cysteine residue very near the amino terminus. In C1q, this cysteine provides the only interchain disulfide bonds (Reid, 1976). The disulfide bonding in conglutinin must be more complicated since six identical chains appear to be disulfide linked (see below). Aside from the above-mentioned cysteine, there is no sequence homolgy between the amino-terminal sequence of conglutinin and either the A, the B, or the C chain of C1q (Reid et al., 1982). Conglutinin also shares properties with some of the recently described collagens which contain non-collagen-like regions. Intima (or type VI) collagen, for example, consists of disulfide-linked subunit polypeptides which are roughly similar in size to the conglutinin polypeptide chain ($M_r$ 40 000–70 000) and have similar amino acid compositions (Furuto & Miller, 1980; Odermatt et al., 1983). They are also similar in that each have amino and carboxy-terminal non-collagen-like domains. These regions, however, and the collagen-like regions appear to differ in size from the equivalent regions in conglutinin. Type M collagen isolated from cartilage is also somewhat similar to conglutinin in that it consists of relatively small ($M_r$ 33 000) disulfide-linked chains (Shinokomaki et al., 1980). The composition of a collagen-like glycoprotein derived from bovine aorta and nuchal ligament, and which may be related to intima collagen, is strikingly similar to, but not identical with, that of conglutinin (Gibson & Cleary, 1982). The overall structure of the conglutinin polypeptide chain shows some similarity to that of the procollagen molecules. The carboxy-terminal globular portion of conglutinin is similar in size to the carboxy-terminal $\alpha$1(I) and $\alpha$2(I) procollagen peptides (Fuller & Boedker, 1981; Kühn, 1982). Similar to the procollagens, most of the cysteine is in the carboxy-terminal globular portion of the conglutinin molecules; as with the procollagens, this region of the molecule contains interchain disulfide bonds (data not shown). However, the overall amino acid composition of this region, as inferred from the peptide which precipitates during digestion with collagenase, is not closely similar to that of the carboxy-terminal procollagen peptides (Fuller & Boedker, 1981; Bernard et al., 1983). Less similarity is seen in comparison of the amino-terminal globular portion of conglutinin with the amino terminus of either the procollagens or the $\alpha$1(I), $\alpha$2(I), or $\alpha$1(III) collagen polypeptide chains (Hürlein et al., 1979; Rauterberg et al., 1972; Fietzek et al., 1974; Glanville & Fietzek, 1976).

The conglutinin molecule, under denaturing conditions, appears to consist of six identical disulfide-linked polypeptide chains. NaDodSO$_4$–polyacrylamide gel electrophoresis gives apparent molecular weights of 48 000 and 300 000 for reduced and unreduced conglutinin, respectively. This suggested the possibility that it was made up of six chains, but since a molecular weight determination as high as 300 000 is not very reliable, it was felt that additional confirmation of this estimate was desirable. The cross-linking experiments with either gluaraldehyde or dimethyl suberimidate also indicate that conglutinin consists of six polypeptide chains. Previous data indicated that, under nondenaturing conditions, conglutinin had an apparent molecular weight of 750 000 (Lachmann & Richards, 1964). This estimate was based on diffusion and ultracentrifugation data. It may be therefore, that native conglutinin consists of a noncovalently linked dimer or trimer, with each subunit consisting of the six covalently linked polypeptide chains. It is also likely, on the basis of the amino acid composition, the sensitivity to collagenase, and the presence of a pepsin-resistant region, that conglutinin contains a collagen-like triple-helical region. This, also, is consistent with the above data. Physical studies, including ultracentrifugation, gel filtration, and electron microscopy, are currently in progress in order to attempt to resolve these questions.

## Acknowledgments

We thank Andre Dauphinais for excellent technical assistance and Dr. Richard Harrison and Dr. David Eyre for helpful discussions and suggestions.

## References

Bernard, M. P., Chu, M.-L., Myers, J. C., Ramirez, F., Eikenberry, E. F., & Prockop, D. J. (1983) *Biochemistry 22*, 5213.

Bordet, J., & Gay, F. P. (1906) *Ann. Inst. Pasteur Paris 20*, 467.

Bordet, J., & Streng, O. (1909) *Zintralbl. Bakteriol., Parasitenkd., Infektionskr. Hyg., Abt.* 1: *Orig. 49*, 260.

Brauer, A. W., Margiolies, M. N., & Haber, E. (1975) *Biochemistry 14*, 3029.

Davis, A. E., III, & Harrison, R. A. (1982) *Biochemistry 21*, 5745.

Eisenberg, R. A., Theofilopoulous, A. N. & Dixon, F. J. (1977) *J. Immunol. 118*, 1428.

Fietzek, P. P., Breitkreutz, D., & Kühn, K. (1974) *Biochim. Biophys. Acta 365*, 305.

Fuller, F., & Boedker, H. (1981) *Biochemistry 20*, 996.

Furthmayr, H., & Timpl, R. (1971) *Anal. Biochem. 41*, 510.

Furuto, D. K., & Miller, E. J. (1980) *J. Biol. Chem. 255*, 290.

Gibson, M. A., & Cleary, E. G. (1982) *Biochem. Biophys. Res. Commun. 105*, 1288.

Gillett, R., & Gull, K. (1972) *Histochemie 30*, 162.

Glanville, R. W., & Fietzek, P. P. (1976) *FEBS Lett. 71*, 99.

Harrison, R. A., & Lachmann, P. J. (1980) *Mol. Immunol. 17*, 9.
Hürlein, D., Fietzek, P. P., Wachter, E., Lapiere, C. M., & Kühn, K. (1979) *Eur. J. Biochem. 99*, 31.
Ji, T. H. (1983) *Methods Enzymol. 91*, 580.
Johny, K. V., Dasgupta, M. K., Singh, B., & Dosseter, J. B. (1980) *Clin. Exp. Immunol. 40*, 459.
Kühn, K. (1982) in *Immunochemistry of the Extracellular Matrix* (Furthmayr, H., Ed.) Vol. 1, p 1, CRC Press, Boca Raton, Fl.
Lachmann, P. J. (1962) *Immunology 5*, 687.
Lachmann, P. J. (1967) *Adv. Immunol. 6*, 480.
Lachmann, P. J., & Coombs, R. R. A. (1965) in *Complement, Ciba Foundation Symposium* (Walstenholme, G. E. W., & Knight, J., Eds.) p 242, Little, Brown and Co., Boston, Ma.
Lachmann, P. J., & Müller-Eberhard, H. J. (1968) *J. Immunol. 100*, 691.
Laemmli, U. K. (1971) *Nature (London) 227*, 680.
Linscott, W. D., Ranken, R., & Triglia, R. P. (1978) *J. Immunol. 121*, 658.
Mauca, R., Gigliorini, P., Bombardieri, S., & Celada, R. (1980) *Clin. Immunol. Immunopathol. 16*, 131.
Miller, R. L., & Udenfriend, S. (1970) *Arch. Biochem. Biophys. 139*, 104.
Nagasawa, S., & Stroud, R. M. (1977) *Immunochemistry 14*, 749.
Odermatt, E., Risteli, J., Van Delden, V., & Timpl, R. (1983) *Biochem. J. 211*, 295.
Pangburn, M. K., Schreiber, R. D., & Müller-Eberhard, H. J. (1977) *J. Exp. Med. 146*, 257.
Rauterberg, J., Timpl, R., & Furthmayr, H. (1972) *Eur. J. Biochem. 27*, 231.
Reid, K. B. M. (1974) *Biochem. J. 141*, 189.
Reid, K. B. M. (1976) *Biochem. J. 155*, 5.
Reid, K. B., Gagnon, J., & Frampton, J. (1982) *Biochem. J. 203*, 559.
Shinokomaki, M., Duance, V. C., & Baily, A. J. (1980) *FEBS Lett. 121*, 51.
Zalut, C., Henzel, W. S., & Harris, H. W. (1980) *J. Biochem. Biophys. Methods 3*, 11.

# $^1$H Nuclear Magnetic Resonance Spectroscopic Study of the Polypeptide Toxin I from *Anemonia sulcata*†

Paul R. Gooley, Laszlo Beress, and Raymond S. Norton*

ABSTRACT: High-resolution $^1$H nuclear magnetic resonance (NMR) spectroscopy at 300 MHz has been used to study the conformation in aqueous solution of the polypeptide toxin I (ATX I) from the sea anemone *Anemonia sulcata*. Resonances from a number of aromatic and methyl groups have been assigned, in many cases to specific amino acid residues in the sequence. Unusual splitting patterns due to virtual coupling are observed for the aromatic protons of one Trp residue and $\gamma$-$CH_3$ of one Thr residue. ATX I appears to be a flexible molecule but possesses a stable, nonrandom tertiary structure. The latter is characterized by a hydrophobic region encompassing Trp-23 and -31 and several aliphatic residues. A limited region of the molecule coexists in two equally populated structural forms, which are detected by splitting of the methyl resonances from Met-18 and a Thr residue. The aromatic rings of Phe-25 and Tyr-42 are not involved in strong interactions with other residues. The pH and temperature dependencies of the spectrum have been analyzed. $pK_a$ values are obtained for Gly-1 (8.3), Lys-7 and -45 (11.2), Glu-35 (4.1), Tyr-42 (10.4), and one of the two remaining carboxylates, tentatively assigned to Asp-9. Protonation of the latter, at around pH 3, is accompanied by an extensive conformational change.

Sea anemones contain a number of polypeptides and proteins which they use for the capture of prey and for defense. These molecules display a wide variety of biological activities, ranging from cardiostimulants, cardiotoxins, and neurotoxins to hemolysins and protease inhibitors (Beress, 1978, 1982; Norton et al., 1978; Alsen, 1983).

Of particular interest are a series of homologous polypeptides of molecular weight about 5000, isolated from the anemones *Anthopleura xanthogrammica*, *A. elegantissima*, and *Anemonia sulcata*. These molecules exert potent effects on the mammalian heart and on mammalian and crustacean nerves, apparently by binding to the sodium channels of these tissues in such a way as to delay the inactivation of this channel and prolong the action potential (Romey et al., 1976; Bergman et al., 1976; Kodama et al., 1981). This activity has been the subject of intensive investigation, but our knowledge of the electrophysiological properties of these molecules is not matched by an understanding of the structural basis for their observed activities. One of these polypeptides, anthopleurin-A, has been studied by laser Raman, circular dichroism, and fluorescence spectroscopy (Ishizaki et al., 1979), as well as by natural-abundance $^{13}$C NMR[1] spectroscopy (Norton & Norton, 1979; Norton et al., 1982). These studies showed that anthopleurin-A is a globular protein containing numerous $\beta$-bends and some $\beta$-pleated sheet regions and provided limited information on the environments of several amino acid side chains. A second member of this series, *Anemonia sulcata* toxin II (ATX II), has been studied by laser Raman spec-

† From the School of Biochemistry, University of New South Wales, Kensington NSW 2033, Australia (P.R.G. and R.S.N.), and the Department of Toxicology, University of Kiel, D-2300 Kiel 1, Federal Republic of Germany (L.B.). *Received November 10, 1983.* This work was supported in part by grants from the Australian Research Grants Scheme and the Deutsche Forschungsgemeinschaft (Grant BE 55418).

[1] Abbreviations: ATX I, *Anemonia sulcata* toxin I; ATX II, *Anemonia sulcata* toxin II; NMR, nuclear magnetic resonance; DSS, sodium 4,4-dimethyl-4-silapentane-1-sulfonate.

0006-2960/84/0423-2144$01.50/0 