Localization of Carbohydrate Attachment Sites and Disulfide Bridges in Limulus α2-Macroglobulin

EVIDENCE FOR TWO FORMS DIFFERING PRIMARILY IN THEIR BAIT REGION SEQUENCES*

Lise B. Husted‡, Ebbe S. Sørensen‡, Peter B. Armstrong§§, James P. Quigley§§
Lene Kristensen‡, and Lars Sottrup-Jensen‡‡

From the §Department of Molecular Biology, Science Park Division, University of Aarhus, 8000 Aarhus C, Denmark,
¶Marine Biological Laboratory, Woods Hole, Massachusetts 02543, the ΔDepartment of Molecular and Cellular Biology,
University of California, Davis, California 95616, and the ΔDepartment of Vascular Biology,
The Scripps Research Institute, La Jolla, California 92037

The primary structure determination of the dimeric invertebrate α2-macroglobulin (α2M) from Limulus polyphemus has been completed by determining its sites of glycosylation and disulfide bridge pattern. Of seven potential glycosylation sites for N-linked glycosylation, six (Asn725, Asn807, Asn866, Asn906, Asn0987, and Asn1145) carry common glucosamine-based carbohydrate groups, whereas one (Asn890) carries a carbohydrate chain containing both glucosamine and galactosamine. Nine disulfide bridges, which are homologues with bridges in human α2M, have been identified (Cys226–Cys269, Cys346–Cys380, Cys512–Cys599, Cys587–Cys807, Cys840–Cys876, Cys874–Cys910, Cys946–Cys1328, Cys1104–Cys1155, and Cys1450–Cys1475). In addition to these bridges, Limulus α2M contains three unique bridges that connect Cys269 and Cys302, Cys1129 and Cys1174, respectively, and Cys719 in one subunit with the same residue in the other subunit of the dimer. The latter bridge forms the only interchain disulfide bridge in Limulus α2M. The location of this bridge within the bait region is discussed and compared with other α-macroglobulins. Several peptides identified in the course of determining the disulfide bridge pattern provided evidence for the existence of two forms of Limulus α2M. The two forms have a high degree of sequence identity, but they differ extensively in large parts of their bait regions, each α2M probably controls a particular set of proteinases, although information on this is fragmentary (2, 6).

Limulus α2M is the most extensively studied invertebrate αM. It is a 360-kDa dimer (7, 8), and its proteinase-binding characteristics (1, 9), shape (7, 8), amino acid sequence deduced from its cDNA sequence (10), and carbohydrate composition (10) have been determined. In the set of peptides generated to initiate cDNA cloning (11, 12), it was observed that residues in several positions were at variance with those predicted from the cDNA (10), suggesting that the Limulus α2M used, which was purified from pooled hemolymph, was a mixture of two or more forms. Like most other known αMs, Limulus α2M contains internal thiol esters (13). When activated during proteinase complex formation, the αM-thiol esters rapidly react with nucleophiles on the attacking proteinase and other available nucleophiles (14). This process results in efficient self-cross-linking reaction within the Limulus α2M dimer contributes to the tight binding of trypsin (17).

Human α2M is the only αM for which complete information on the arrangement of its intrachain and interchain disulfide bridges and positions of Asn-based carbohydrate groups is available (18, 19). Apart from the Cys residue being part of the thiol ester site, the human α2M subunit contains 24 Cys residues of which 22 engage in 11 intrachain bridges and 2 engage in interchain disulfide bridges thereby aligning the two subunits of the α2M dimer in an antiparallel fashion. For vertebrate αMs of known sequence most positions of disulfide bridges can readily be predicted from the data on human α2M.

* This work was supported by grants from the Danish Science Research Council, by a basic research grant from the University of Aarhus, by the Novo Nordic Foundation (to L. S. J.), and by National Science Foundation Grant MCB 97-28771 (to P. B. A.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** To whom correspondence should be addressed: Dept. of Molecular Biology, Science Park Division, University of Aarhus, Gustav Wieds Vej 10, 8000 Århus C, Denmark. Tel.: 45-89-42-50-51; Fax: 45-89-42-50-77; E-mail: lj@mb.au.dk

† The abbreviations used are: α2M, α2-macroglobulin; αM, α-macroglobulin; IAA, iodoacetamide; MESSA, mercaptoethanesulfonic acid; MS, mass spectrometry; RP-HPLC, reverse phase-high pressure liquid chromatography; PTH, phenylthiohydantoin.

Received for publication, August 12, 2002, and in revised form, September 5, 2002
Accepted for publication, JBC Papers in Press, September 5, 2002, DOI 10.1074/jbc.M208236200

43698 This paper is available on line at http://www.jbc.org
In addition to the Cys residue engaging in thiol ester formation, the Limulus \( \alpha_2 \)M subunit contains 23 Cys residues of which 18 would be expected to form nine disulfide bridges equivalent with those found in human \( \alpha_2 \)M. However, for Limulus \( \alpha_2 \)M the pattern of disulfide bridges is ambiguous with regard to five positions (10). The Cys residues that have no counterpart in human \( \alpha_2 \)M are located at positions 360, 381, 719, 1370, and 1434. Because no free –SH groups can be detected in native Limulus \( \alpha_2 \)M, they are all likely to be paired, and importantly, the two subunits of the Limulus \( \alpha_2 \)M dimer must be connected by an uneven number of interchain bridges in contrast to the human dimer.

From the carbohydrate composition given earlier (10) Limulus \( \alpha_2 \)M contains both glucosamine and galactosamine (19.4 and 2.9 residues/mol subunit, respectively). As galactosamine is not present in human \( \alpha_2 \)M and probably other mammalian \( \alpha_2 \)Ms, it was also of interest to locate the carbohydrate groups in Limulus \( \alpha_2 \)M. Limulus \( \alpha_2 \)M contains seven candidate Asn residues for attachment of glucosamine-based glycan groups (Asn\(^{80} \), Asn\(^{275} \), Asn\(^{397} \), Asn\(^{866} \), Asn\(^{896} \), Asn\(^{1089} \), and Asn\(^{1145} \)) (10). In contrast, carbohydrate groups containing galactosamine are frequently bound to Ser and Thr residues which, however, cannot readily be predicted.

Here we report the determination of the complete disulfide bridge pattern of Limulus \( \alpha_2 \)M consisting of 11 intrachain bridges and one interchain bridge. Curiously, the single interchain bridge engages a Cys residue located in the bait region. We also report the localization of six Asn residues carrying glucosamine-based carbohydrate groups, and one Asn residue carrying a carbohydrate group containing both glucosamine and galactosamine. We further provide evidence from sequencing of a number of peptides for a second Limulus \( \alpha_2 \)M, the sequence of which differs from that reported earlier (10) particularly in its bait region. The two forms (Limulus \( \alpha_2 \)M-1 and -2) are most likely present in each animal and can be partially separated by ion exchange chromatography on Mono Q at pH 7.4.

**EXPERIMENTAL PROCEDURES**

**Proteins and Other Materials—**Limulus \( \alpha_2 \)M was prepared as reported previously (20). Tosyl-phenylalanine chloromethyl ketone-treated trypsin, thermolysin, angiotensin II, and adrenocorticotropic hormone, fragment 18 treated trypsin, thermolysin, and cyanogen bromide-oxidized proteins were from Sigma, Merck, and The Separations Group, and LC packings. \(^{14} \)C-Labeled iodoacetamide (IAA) was from Amersham Biosciences, Machery-Nagel, Shandon, The Separations Group, and LC Biosciences, and standard chemicals were from Sigma, Merck, and Rathburn.

**Analytical Procedures—**Amino acid analysis was performed by cation exchange using established procedures (21). Automated peptide sequencing and PTH-derivative analysis were carried out as reported (22). SDS-PAGE was performed in 10–20% and 20% slab gels using the standard Tris glycine system. Mass spectra were acquired with a Bruker BIFLEX matrix-assisted laser desorption/ionization time-of-flight instrument equipped with a 1-m flight tube, a reflector, a 337-nm nitrogen laser, and a 1-GHz digitizer. Thin film matrix surfaces were prepared using the fast evaporation technique from \( \alpha \)-cyan-4-hydroxy-cinnamic acid dissolved in acetonitrile/water (99:1) to 30 \( \mu \)g/\( \mu \)l. A 0.5-\( \mu \)l volume of analyte (0.1–10 \( \mu \)mol/\( \mu \)l) was deposited on the matrix surface, and allowed to dry on to the crystals. Spectra were obtained by averaging 20–50 single-shot spectra and calibrated externally by using the charged constants of angiotensin II and adrenocorticotropic hormone, fragment 18. Theoretical peptide masses (MH\(^{+} \)) were calculated using the GPMAW program (Lighthouse Data, Odense, Denmark).

**Main Digest—**Approximately 30 mg of Limulus \( \alpha_2 \)M was used as start material for the main digest. SDS-PAGE revealed that almost complete bait region cleavage had occurred during preparation and/or storage, and the material was therefore treated at pH 5.5 with 20 mM iodoacetamide to block the thiol group appearing upon thiol ester cleavage. After removal of excess reagent by gel chromatography on Sephadex G-25 in 10% formic acid, the material was water-dried, redissolved in 70% formic acid, and treated with 50 mg of CNBr for 20 h. After freeze-drying the degraded Limulus \( \alpha_2 \)M was resolved in 300 \( \mu \)l of formic acid, and 10 volumes of water was added. Addition of aliquots of 5 M NaOH was used to raise the pH of the solution, and at pH 4–5 the solution turned turbid. Precipitation appeared to be complete at pH 7–8.

In analytical experiments trypsin or thermolysin was added at pH 4, and digestion was attempted after raising the pH to 7–8. In analytical experiments showed only extensive degradation with thermolysin. Therefore, to a solution of \( \sim 18 \) mg of CNBr-degraded Limulus \( \alpha_2 \)M 1:50 (w/w) thermolysin was added at pH 4 and the pH subsequently raised to 7 by addition of Tris. After incubation for 90 min at 55 °C with stirring, the suspension had nearly cleared. To separate the larger carbohydrate-containing and disulfide peptides from the small peptides, the digest was fractionated on a Superdex peptide column using 0.1% trifluoroacetic acid, 25% acetonitrile as eluent. The column effluent was monitored by measuring the absorbance at 225 nm and by determining the amount of half-cystine, glucosamine (GlcN), and galactosamine (GalN) in fractions combined in pairs. The elution conditions and the fraction size were the same as in A.
Preparation and Digestion of Partially Reduced Methylamine-reacted Limulus α2M—Prior to this experiment Limulus α2M was treated with methylamine and IAA. By using 8 mM mercaptoethanesulfonic acid (MESA) at pH 8.0 for 20 min, the interchain disulfide bridge(s) in the Limulus α2M dimer could be reduced to an extent of >90% as evaluated the material was lost. An additional preparative tryptic digest as above was subjected to gel chromatography on Superdex 75 using 50% formic acid (Fig. 2B) was separated on a 4 × 250-mm Nucleosil C18 column. The column was equilibrated with 5% solvent B (90% acetonitrile, 0.08% phosphoric acid) and 95% solvent A (0.1% trifluoroacetic acid) and eluted with a gradient formed by solvent A and solvent B (dashed line). The separation was performed at 50 °C at a flow rate of 1 ml/min, and fractions of 0.5 ml were collected. Peptides were detected at 215 nm (solid line), and the amount of half-cystine in fractions having an absorbance >0.05 was determined (vertical bars). By MS and sequence analyses of half-cystine-containing fractions it was found that the fractions indicated by a horizontal bar contained a disulfide-bridged cluster involving Cys577, Cys578, and Cys579. B, the fractions shown by the horizontal bar in A were digested with chymotrypsin and separated on a 2 × 250-mm Nucleosil C18 column. The column was operated at 50 °C at a flow rate of 0.2 ml/min and eluted with a gradient formed from the same solvents as in A (dashed line). The separation was monitored at 215 nm (solid line), and 0.1 ml fractions were collected. Fractions 26–35 (bar) were analyzed by MS.

In most cases the evidence for assigning the disulfide bridges was based on the amino acid composition of performic acid-oxidized relatively short pure peptides and sequence analysis of intact peptides. No peptides contained an internal disulfide bridge, and two sequences in near equimolar yield were seen, occasionally on a background of several minor irrelevant components. In cases where Cys125 was released after less than 5–8 cycles of Edman degradation, bis-PTH-Cys125 was normally seen in the RP-HPLC analysis of the PTH-derivatives eluting as a low yield peak near PTH-Tyr (24); after more than 8 cycles no signal was observed. When performed, mass spectrometry (MS) confirmed the assignment based on sequence analysis. However, in two cases MS provided the full evidence for assignment.

The evidence for locating carbohydrate groups to particular positions was based on the presence of glucosamine (and in one case also galactosamine) in hydrolysates coupled with the lack of a PTH-derivative when encountering Asn residues located in the sequence Asn-Xaa-Ser/Thr (Xaa not Pro).

Preparation and Digestion of Partially Reduced Methylamine-reacted Limulus α2M—Prior to this experiment Limulus α2M was treated with methylamine and IAA. By using 8 mM mercaptoethanesulfonic acid (MESA) at pH 8.0 for 20 min, the interchain disulfide bridge(s) in the Limulus α2M dimer could be reduced to an extent of >90% as evaluated by guest on August 31, 2017 http://www.jbc.org/ Downloaded from
by non-reducing SDS-PAGE (not shown). Methyamine-reacted Limulus αM (30 mg, 4 mg/ml in 100 mM Tris-HCl, pH 8.0) reduced with MESA was freed of MESA by gel chromatography on a column of Sephacryl G-25 equilibrated and eluted with degassed 20 mM sodium acetate, 100 mM NaCl, pH 4.5. The pH of the Limulus αM solution was then raised to 9.0 by addition of Tris, and 50 μCi of 14C-labeled IAA was added. After reacting for 30 min the solution was made 20 mM in unlabeled IAA and allowed to react for a further 30 min. Then the excess reagents were separated from Limulus αM by a second round of gel chromatography using 20 mM Tris-HCl, 100 mM NaCl, pH 8.0, as eluent, and the protein was subsequently freeze-dried.

The labeled protein was redissolved in 2 ml of 6 M guanidinium chloride; the pH was adjusted to 9.0 by addition of Tris, and dithioerythritol was added to 10 mM to fully reduce Limulus αM. After reduction for 30 min IAA was added to 30 mM, and after 30 min of reaction the reduced and carboxamidomethylated protein was recovered by gel chromatography on Sephacryl G-25 using 20 mM Tris-HCl, 100 mM NaCl, pH 9.0, as eluent. The preparation was digested with 1:50 (w/w) trypsin for 3 h at 37 °C. Then the fractions containing radioactivity were located by scintillation counting, and the major peptides were further purified by RP-HPLC on a column of Hypersil C18 or by cation exchange chromatography as above (not shown).

**Determination of the Amino Acid Sequence of the Bait Region of Limulus αM-2**—One mg of Limulus αM was treated with ~2 mg of CNBr in 70% formic acid for 20 h at room temperature. After drying the degraded material was redissolved in 500 μl of 6 M guanidinium chloride, 50 mM Tris-HCl, pH 9.0, and reduced for 30 min with 10 mM dithioerythritol. The peptide solution was then acidified with trifluoroacetic acid and loaded on a 4.6 × 250-mm column equilibrated with 4.5% 2-propanol in 0.1% trifluoroacetic acid. The column was eluted with a gradient of 2-propanol in 0.1% trifluoroacetic acid (Fig. 5), and the bait region peptides eluted in fractions 25 to 35 ml containing fragments were identified by SDS-PAGE and sequence analysis (bar).

**Isolation of the Two Forms of Limulus αM from Single Animal Hemolymph**—Hemolymph (80- and 100-ml samples) was separately drawn from two animals and processed as described earlier (20). Five mg of material depleted in hemocyanin and pentraxin was subjected to gel chromatography on a Superose 6 HR 10/30 column equilibrated and eluted with 50 mM Tris-HCl, pH 7.4. The Limulus αM containing fractions near the void volume of the column were pooled and loaded on a Mono Q HR 5/5 column equilibrated with the above Tris buffer and eluted with a shallow gradient of NaCl. Two partially separated peaks appeared at [NaCl] = 190 and 220 mM, representing Limulus αM-2 and -1, respectively (Fig. 6).

**RESULTS**

**Assignment of Disulfide Bridges and Location of Carbohydrate Groups**—When determining the sequence of Limulus αM by a combination of peptide sequencing and cDNA cloning (10), we found a number of positions in which residues deviating from those determined from the cDNA sequence were present. As detailed below additional partial peptide sequence information has been obtained that does not conform with the sequence deduced from the cDNA sequence. In fact, the preparation of Limulus αM used for analysis is a mixture of two closely related forms, αM-1 representing the published sequence (10) and αM-2 representing the set of deviating partial sequences.

In order to localize the glycosylation sites and disulfide bridges in Limulus αM, a digest of Limulus αM termed the main digest was made. First Limulus αM, which had been treated with IAA, was degraded with CNBr with the aim of generating a limited number of relatively large peptide clusters.
which could then be further enzymatically digested. However, in contrast to human αM (25), most of the material appeared in one large cluster. Subsequent digestion of the material with trypsin was unsuccessful due to limited solubility at pH 7–8, but treatment with thermolysin brought most of the material into solution. Upon gel chromatography on a Superdex peptide column, the effluent was divided into two pools, one consisting of the large peptides including those containing carbohydrate (pool A) and one consisting of the small peptides (pool B) (see “Experimental Procedures”). The material in each pool was subjected to RP-HPLC (Fig. 1, A and B). In a few cases the material in the fractions obtained after RP-HPLC was of sufficient purity to allow assignment of the disulfide bridges and to locate the sites of carbohydrate attachment by compositional and sequence analysis. In general, the RP-HPLC pools made on the basis of their content of half-cystine and amino sugars were further separated by cation exchange chromatography (not shown).

Because no peptide material could be isolated from the main digest in sufficient purity to assign two particular disulfide bridges, and because evidence for localizing two carbohydrate groups originating from both carbohydrate-containing pools made on the basis of their content of half-cystine and amino sugars were further separated by cation exchange chromatography (not shown).

The evidence for the location of carbohydrate groups to Asn residues is summarized in Table I. As seen in Fig. 1A glucosamine was present in many fractions in the elution profile, whereas galactosamine was only present in fractions 163–180. The major component in these fractions was a peptide with the N-terminal sequence LYANGS (77–82) apparently containing both glucosamine and galactosamine. In subjecting the material in these fractions to lectin affinity chromatography on concanavalin A and Jacalin-Sepharose to possibly separate the glucosamine- and galactosamine-containing peptides, the material was lost. However, from subdigestion of material from the ancillary digest (pool 2 in Fig. 2A) the peptide LYANGSYSSPSSNDFFFE containing both amino sugars was isolated.

It was further established from sequence analysis of fractions obtained from the RP-HPLC separation of pool A from the main digest (Fig. 1A) that Asn1145, Asn307, Asn896, Asn998, and Asn1146 all carry glucosamine-based carbohydrate groups. The composition of the material found in pool A26–A28 indicated that a short peptide containing Asn998 was present. However, no evidence from sequence analysis for carbohydrate on Asn996 could be obtained possibly due to cyclization of Glun962 during treatment with CNBr in formic acid. Upon subdigestion of the material from pool 1 in Fig. 2A originating from the ancillary digest, a disulfide-bridged peptide cluster composed of residues 853–853, 864–864, and 889–921 was obtained (pool 1 in Fig. 2A). Sequence analysis of this peptide provided evidence for carbohydrate on Asn996 as well as confirmatory evidence for carbohydrate on Asn896.

The sequences of peptides containing the glycosylated Asn residues shown in Table I all represent αM-1, and no evidence for peptides containing Asn residues in different sequences was obtained. This indicates that αM-2 has the same pattern of glycosylated Asn residues as αM-1.

In Table II the evidence for the assignment of all disulfide bridges in Limulus αM is summarized. The bridges Cys1155 to Cys1434 were all identified from the main digest. Most of the peptides containing these disulfide bridges were recovered as several cleavage variants, some of which are listed in Table II. With the exception of the bridges considered below, the sequences of the peptide mates originated from αM-1. In the case of the bridges Cys1104–Cys1155 and Cys1362–Cys1475, the mates containing Cys1155 and Cys1475 did not originate from αM-1, but rather from αM-2, and the stretches containing Cys1362 and Cys1370 were all identified from the main digest. Most of the peptides containing these disulfide bridges were recovered as several cleavage variants, some of which are listed in Table II. With the exception of the bridges considered below, the sequences of the peptide mates originated from αM-1. In the case of the bridges Cys1104–Cys1155 and Cys1362–Cys1475, the mates containing Cys1155 and Cys1475 did not originate from αM-1, but rather from αM-2, and the stretches containing Cys1362 and Cys1370 were all identified from the main digest. Most of the peptides containing these disulfide bridges were recovered as several cleavage variants, some of which are listed in Table II. With the exception of the bridges considered below, the sequences of the peptide mates originated from αM-1. In the case of the bridges Cys1104–Cys1155 and Cys1362–Cys1475, the mates containing Cys1155 and Cys1475 did not originate from αM-1, but rather from αM-2, and the stretches containing Cys1362 and Cys1370 were all identified from the main digest. Most of the peptides containing these disulfide bridges were recovered as several cleavage variants, some of which are listed in Table II. With the exception of the bridges considered below, the sequences of the peptide mates originated from αM-1. In the case of the bridges Cys1104–Cys1155 and Cys1362–Cys1475, the mates containing Cys1155 and Cys1475 did not originate from αM-1, but rather from αM-2, and the stretches containing Cys1362 and Cys1370 were all identified from the main digest.
The pools indicated with prefix A are from the separations shown in Fig. 1A and Fig. 1B, respectively. Pools named with the prefix Ct are from the separation shown in Fig. 3B. For masses below 1200-Da monoisotopic masses were calculated and above 1200-Da average masses are shown. The underlined amino acid residues deviate from the amino acid sequence deduced from cDNA cloning.

### Table II

<table>
<thead>
<tr>
<th>Bridge position</th>
<th>Peptide set</th>
<th>Present in pool</th>
<th>MH(^+) found</th>
<th>MH(^+) calc.</th>
<th>Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>228–269</td>
<td>ICAQ/IDGC</td>
<td>B131–138</td>
<td>KIC/IDGC (B149–152)</td>
<td>QICAQ/LIDGC (A203–206)</td>
<td></td>
</tr>
<tr>
<td>361–382</td>
<td>LCRP/RACKE</td>
<td>B183–186</td>
<td>LCRP/RACKE (A and B287–290)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>456–580</td>
<td>IEGF/KVKC</td>
<td>B121–124</td>
<td>TEEINC/KVKC (B159–162)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>612–799</td>
<td>IGC/IC</td>
<td>B113–118</td>
<td>LSC/ICQGKSDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>657–707</td>
<td>TRPCP/KP/GF/YCEDYK</td>
<td>C28</td>
<td>1811.31</td>
<td>1811.0</td>
<td></td>
</tr>
<tr>
<td>719–719</td>
<td>EDGGRPCPQ/FEDGGRPCPQFDE</td>
<td>C31</td>
<td>2697.91</td>
<td>2697.8</td>
<td></td>
</tr>
<tr>
<td>849–876</td>
<td>LDLSCP/VCYGGKSDT</td>
<td>A251–254</td>
<td>1410.7</td>
<td>1411.6</td>
<td></td>
</tr>
<tr>
<td>874–910</td>
<td>TSC/ICGNQD</td>
<td>B73–76</td>
<td>LSECLPN/VCYGGKSDTDTTRWM (A283–270)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>946–1328</td>
<td>FACPKDQ/YNGK/ATQSGCG</td>
<td>A139–144, B135–138</td>
<td>1658.1</td>
<td>1657.8</td>
<td></td>
</tr>
<tr>
<td>1104–1155</td>
<td>SNGC/ATBC</td>
<td>B49–52</td>
<td>FACPKDQ/GNK/ATQSGCGL (A219–222)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1362–1475</td>
<td>LYRDCNN/IDENCEKLP</td>
<td>A239–246</td>
<td>2053.3</td>
<td>2053.3</td>
<td></td>
</tr>
<tr>
<td>1370–1434</td>
<td>IATC/LC</td>
<td>B153–156</td>
<td>IATC/LTDQ/KQC (A139–144)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The fragment set was subdigested with *S. aureus* proteinase after partial separation from several other fragments on a Superdex 75 column that was equilibrated and eluted in 50% formic acid (Fig. 2B). The *S. aureus* proteinase digest was fractionated by RP-HPLC (Fig. 3A), and a smaller fragment was obtained that contained the three Cys residues. This fragment was further digested with chymotrypsin, and the digest was again fractionated by RP-HPLC (Fig. 3B). By MS analysis of fractions from this RP-HPLC separation, it was demonstrated that Cys\(^729\) is connected to Cys\(^727\) (expected peptide set TRPCP/KP/GF/YCEDYK bound to YCEDYK (a\(\alpha\)M-1, Table II) and that Cys\(^719\) is engaged in forming an interchain bridge within the *Limulus* a\(\alpha\)M dimer. In this case no MS data were obtained from the peptide expected from a\(\alpha\)M-1 (dimer of \(719^{\text{EDGGRPCPQ}}/\text{EDGGRPCPQFDE}\)), but rather from \(\alpha\)M-2 as the dimeric peptide set \(719^{\text{EDGGRPCPQ}}/\text{EDGGRPCPQFDE}\) was identified (Table II). That sequence was known from peptides isolated from partial reduction experiments described below.

### Partial Reduction of Methylamine-treated Limulus \(\alpha\)M—

Due to the particular location of the 24 half-cystine residues in human a\(\alpha\)M, whose pairing to 12 bridges was relatively easily established (18), the identification of two of these as being engaged in interchain bridges was made by partial reduction using the highly solvated MESA and radiolabeling (19). As judged from SDS-PAGE (not shown), the subunits of methylamine-reacted *Limulus* a\(\alpha\)M were likewise essentially fully separated using 8 m \(\text{MESA}\) for 30 min at room temperature. Following alkylation with radiolabeled IAA and removal of excess reagents, the preparation was fully reduced and alkylated with unlabeled IAA. The material was then digested with trypsin and separated by RP-HPLC (Fig. 4). The fractions containing the major part of the radioactivity were further purified, and two sets of peptides were characterized from these fractions as summarized in Table III.

Two major radiolabeled peptides present in pools 166–168 and 187–188 (Fig. 4), respectively, originated from unexpected cleavage by trypsin at Arg-Gln (705–706). Upon sequence analysis they were found to contain the same N-terminal sequence PCKPSGFEDGGRPCPQ. The radiolabel was solely found in position Cys\(^719\), hence demonstrating that the interchain disulfide bridge containing this Cys residue is indeed solvent-exposed. However, downstream of Cys\(^719\) the sequences diverged, with that of the peptide from pool 166–168 being identical to the sequence expected from the bait region of a\(\alpha\)M-1 and that of pool 187–188 being different (a\(\alpha\)M-2, Table III). This shows that a fairly long stretch of the bait regions of a\(\alpha\)M-1 and a\(\alpha\)M-2 differs markedly in sequence.

The cleavage of partially reduced *Limulus* a\(\alpha\)M by trypsin was incomplete as evidenced by the distribution of a major part of the label in several late eluting fractions. The peptide in pool 204–205, originating from *Limulus* a\(\alpha\)M-2, apparently was that of pool 187–188 having a long C-terminal extension which, however, could not be identified due to poor sequencing yields. Two major labeled peptides present in pools 303–305 and 312–315, respectively, had N termini reflecting cleavage at Arg-Gln (663–664). Upon subdigestion with chymotrypsin several peptides, which could not be adequately purified, indicated that they each contained the bait region and hence represented N-terminally extended versions of the bait region peptides. The specific activity of all peptides containing the bait region was ~40,000 cpm/nmol.

From two other pools peptides having an ~10-fold lower specific activity than the bait region peptides described above were isolated. These were mates of the C-terminal disulfide bridge Cys\(^1362\)–Cys\(^1475\) pointing to solvent exposure of that bridge located in the part of *Limulus* a\(\alpha\)M presumably being equivalent with the receptor-binding domain of human a\(\alpha\)M (26, 27). In addition, two versions of the mate containing Cys\(^1362\) originating from both a\(\alpha\)M-1 and a\(\alpha\)M-2 were found (Table III).

### Determination of the Complete Bait Region Sequence of *Limulus* a\(\alpha\)M-2—

In order to determine the entire bait region sequence of *Limulus* a\(\alpha\)M-2 and to be able to distinguish between the two forms of *Limulus* a\(\alpha\)M, we took advantage of the fact that the partial bait region sequence of a\(\alpha\)M-2 contains a Met residue at position 728 not present in a\(\alpha\)M-1 (Table III). By assuming that a\(\alpha\)M-2 like a\(\alpha\)M-1 has a Met residue at position 864, CNBr degradation would generate ~15- and 18-kDa fragments, respectively. After incubation with CNBr the fragments generated were reduced and separated on a Vydac C4 column (Fig. 5). Both the 18-kDa fragment containing the entire bait region of a\(\alpha\)M-1 and the 15-kDa fragment containing the C-terminal part of the bait region of a\(\alpha\)M-2 were found to elute in fractions 40 and 41. By Edman degradation of samples
electroblotted onto polyvinylidene difluoride, the remaining part of the sequence of the bait region of α₂M-2 was obtained from the 15-kDa fragment. The complete sequence of the bait region of α₂M-2 is aligned with the bait region sequences of Limulus α₂M-1 (10) and tick αM (28) in Fig. 7.

**Isolation and Characterization of the Two Forms of Limulus α₂M from Single Animal Hemolymph**—By having established that Limulus α₂M prepared from pooled hemolymph contains two related forms, we investigated whether each form was present in individual animals. α₂M was prepared from two animals using the procedure of Ref. 20. In the final ion exchange step of each experiment (Fig. 6), two partially separated peaks appeared. Each peak represented pure and active Limulus α₂M as judged from the presence of the 180-kDa subunit and the heat fragments of ~55 and 125 kDa upon reducing SDS-PAGE. Samples were removed from the flanking parts of the partially separated proteins, treated with CNBr, and separated by SDS-PAGE. Electroblotting and sequence analysis as above identified the material eluting at [NaCl] = 190 mM as α₂M-2 (containing the 15-kDa bait region fragment), and the material eluting at [NaCl] = 220 mM as α₂M-1 (containing the 18-kDa bait region fragment). α₂M-2 had the same N-terminal sequence as α₂M-1 (15 residues determined). Based on the heights of the two peaks in Fig. 6, α₂M-1 and α₂M-2 were present in an approximate 2:1 molar ratio in both animals investigated.

**Discussion**

In this work we have completed the primary structure determination of Limulus α₂M by determining its disulfide bridge pattern and localizing its sites of carbohydrate attachment. Furthermore, evidence for the occurrence of two forms of Limulus α₂M (α₂M-1 and α₂M-2) has been obtained.

It was found that all seven potential N-linked glycosylation sites are occupied. As shown in the schematic comparison of the localization of disulfide bridges and glycosylation sites in Limulus and human α₂M in Fig. 8, the only N-linked glycosylation site that is conserved in human α₂M is Asn<sup>306</sup>. Even though Limulus α₂M contains galactosamine (10), we could not identify any Ser or Thr residues containing carbohydrate. Instead we found that Asn<sup>306</sup> carries a glycan containing both glucosamine and galactosamine. The presence of galactosamine-containing N-linked carbohydrate chains has been reported previously in several other proteins including pituitary glycoprotein hormones (29), human urinary kallidinogenase (30), human tissue factor pathway inhibitor (31), bovine component PP3 (32), snake venom batroxobin (33), and hemocyanin from the freshwater snail <i>Lymnaea stagnalis</i> (34) but not in other αM.

Regarding the disulfide bridge pattern of Limulus α₂M, we have confirmed the existence of nine disulfide bridges that are located in a similar position in human α₂M as shown in Fig. 8. The remaining five Cys residues in Limulus α₂M, which are not part of the thiol ester site, were shown to be engaged in three disulfide bridges. One of these bridges (Cys<sup>382</sup>-Cys<sup>383</sup>) is located in the N-terminal part of the protein. Another bridge connecting Cys<sup>1370</sup> and Cys<sup>1434</sup> is located in the region corresponding to the receptor-binding domain of the mammalian αMs. The location of a bridge at this position is compatible with the three-dimensional structure of the receptor-binding domain of bovine α₂M, because the side chains of the equivalent residues in bovine α₂M are located in close proximity to each other on β-strands 2 and 7, respectively (35). Furthermore, a disulfide bridge is located in a similar position in human C3 (36). The third bridge, which is unique to Limulus α₂M, is an interchain disulfide bridge that connects Cys<sup>719</sup> in one subunit with the same residue in the other subunit of the dimer. In a previous alignment of the bait region of Limulus α₂M with the bait regions of other αM, this Cys residue was located at the N-terminal border of the bait region (10). However, this location is not in accordance with our determination of the disulfide bridge pattern, because we found that the bridge, which defines the N-terminal boundary of the bait region, is Cys<sup>657</sup>-Cys<sup>707</sup> (Fig. 7). This implies that the bait region is 12 residues longer than previously thought (10) and that the above-mentioned interchain disulfide bridge is located in the bait region.

To investigate whether Limulus α₂M contains other interchain disulfide bridges than the above-mentioned, Limulus α₂M was partially reduced with MESA, and the freed thiol groups were radiolabeled with IAA. Because MESA is highly solvated it preferably reduces solvent-exposed disulfide bridges. As expected, it was found that the interchain bridge linking the two bait regions was easily reduced. However, the bridge Cys<sup>1362</sup>-Cys<sup>1475</sup> was also reduced by MESA. Because this bridge is conserved in human α₂M and is located in the region corresponding to the receptor-binding domain of the mammalian αMs and, furthermore, is reduced only to an extent of ~10% compared with the bait region interchain bridge, we conclude that the bridge linking the bait regions is the only interchain disulfide bridge in Limulus α₂M.

The location of an interchain disulfide bridge within the bait region of Limulus α₂M shows that parts of the two bait regions of the Limulus α₂M dimer are located in close proximity to each other. This observation is in line with studies on recombinant bait region variants of human α₂M. By deleting parts of the C-terminal end of the bait region of human α₂M, it was shown that the bait regions are involved in forming the interface between its non-covalently associated dimers (37). This was further examined by mutating single residues in the bait region to Cys residues. These cysteine-containing variants formed disulfide-linked tetramers demonstrating that at least two bait regions are located close to each other at the interface between the non-covalently associated dimers of human α₂M (38).

The two interchain bridges in human α₂M are located in a completely different part of the primary structure than the single interchain bridge of Limulus α₂M, namely in the N-terminal part. This difference in location of the interchain bridges in tetrameric human and dimeric Limulus α₂M seems to be reflected in the functionality of the disulfide-linked dimers. Evidence suggests that the disulfide-linked dimers of...
human $\alpha_2$M are non-functional by consisting of two halves of different traps (37, 39, 40). In contrast, the Limulus $\alpha_2$M dimer is known to be an inhibitory unit. Because the interchain bridges are located in different areas of the primary structure in Limulus and human $\alpha_2$M, it is likely that the Limulus dimer resembles the non-covalently linked dimers of human $\alpha_2$M, supporting the view that these dimers are the functional dimers.

By using PCR primers based on the peptide sequences $^{905}$PNEAICG and $^{1210}$YYWQNSI present in two of a set of peptides isolated from Limulus $\alpha_2$M (11, 12), the cDNA and deduced protein sequence of Limulus $\alpha_2$M was determined (10). However, upon aligning the sequences of 24 isolated peptides covering 499 residues with the deduced sequence of Limulus $\alpha_2$M, it was observed that 23 positions differed as shown in the left part of Table IV. This suggests that the preparations from pooled hemolymph used for peptide isolation consisted of at least two related proteins. Additional evidence supporting this was provided by several peptides isolated during the determination of the disulfide bridge pattern. In six cases we have isolated Cys-containing peptides that covered the same peptide stretch but differed at one or more positions within this stretch (Tables II and III). This demonstrates that two forms of Limulus $\alpha_2$M actually exist.

As described under “Results,” both forms probably have the same glycosylation and disulfide bridge pattern, but their bait region sequences differ greatly in an ~25-residue stretch (Fig. 7). A comparison of the bait region sequences of $\alpha_2$M-1 and $\alpha_2$M-2 with that of tick $\alpha_2$M (28) shows that apart from a few scattered short motifs there is no extensive sequence similarity between either bait region of the two forms of Limulus $\alpha_2$M and tick $\alpha_2$M (Fig. 7). This is in line with data on the bait regions of vertebrate $\alpha$Ms that differ greatly both in length and sequence (6). The lack of a Cys residue in the bait region of tick $\alpha_2$M is in agreement with tick $\alpha_2$M most likely being a dimer of non-covalently associated processed subunits (28).

In contrast to the dissimilarity between the bait regions of tick $\alpha_2$M and the two forms of Limulus $\alpha_2$M, there is a high degree of sequence identity between the N-terminal and C-terminal parts of the bait regions of the two forms of Limulus $\alpha_2$M. Despite this sequence identity, it is likely that the two forms are able to inhibit different proteinases because there is no similarity in the central part of the bait region. Previously, a significant degree of sequence identity within the bait region has only been observed among closely related $\alpha$Ms such as carp $\alpha_2$Ms (41) and the two isoforms of rat $\alpha_1$-inhibitor 3 (42, 43), suggesting that the two forms of Limulus $\alpha_2$M have a high degree of sequence identity too. This is further supported by the fact that even though peptide stretches covering almost 900 amino acid residues have been sequenced only 52 differences between the two forms have been observed (Table IV). In line with this agreement with tick $\alpha_2$M (28) shows that apart from a few scattered short motifs there is no extensive sequence similarity between either bait region of the two forms of Limulus $\alpha_2$M and tick $\alpha_2$M (Fig. 7). This is in line with data on the bait regions of vertebrate $\alpha$Ms that differ greatly both in length and sequence (6). The lack of a Cys residue in the bait region of tick $\alpha_2$M is in agreement with tick $\alpha_2$M most likely being a dimer of non-covalently associated processed subunits (28).

In contrast to the dissimilarity between the bait regions of tick $\alpha_2$M and the two forms of Limulus $\alpha_2$M, there is a high degree of sequence identity between the N-terminal and C-terminal parts of the bait regions of the two forms of Limulus $\alpha_2$M. Despite this sequence identity, it is likely that the two forms are able to inhibit different proteinases because there is no similarity in the central part of the bait region. Previously, a significant degree of sequence identity within the bait region has only been observed among closely related $\alpha$Ms such as carp $\alpha_2$Ms (41) and the two isoforms of rat $\alpha_1$-inhibitor 3 (42, 43), suggesting that the two forms of Limulus $\alpha_2$M have a high degree of sequence identity too. This is further supported by the fact that even though peptide stretches covering almost 900 amino acid residues have been sequenced only 52 differences between the two forms have been observed (Table IV). In line with this agreement with tick $\alpha_2$M (28) shows that apart from a few scattered short motifs there is no extensive sequence similarity between either bait region of the two forms of Limulus $\alpha_2$M and tick $\alpha_2$M (Fig. 7). This is in line with data on the bait regions of vertebrate $\alpha$Ms that differ greatly both in length and sequence (6). The lack of a Cys residue in the bait region of tick $\alpha_2$M is in agreement with tick $\alpha_2$M most likely being a dimer of non-covalently associated processed subunits (28).

In contrast to the dissimilarity between the bait regions of tick $\alpha_2$M and the two forms of Limulus $\alpha_2$M, there is a high degree of sequence identity between the N-terminal and C-terminal parts of the bait regions of the two forms of Limulus $\alpha_2$M. Despite this sequence identity, it is likely that the two forms are able to inhibit different proteinases because there is no similarity in the central part of the bait region. Previously, a significant degree of sequence identity within the bait region has only been observed among closely related $\alpha$Ms such as carp $\alpha_2$Ms (41) and the two isoforms of rat $\alpha_1$-inhibitor 3 (42, 43), suggesting that the two forms of Limulus $\alpha_2$M have a high degree of sequence identity too. This is further supported by the fact that even though peptide stretches covering almost 900 amino acid residues have been sequenced only 52 differences between the two forms have been observed (Table IV). In line with this agreement with tick $\alpha_2$M (28) shows that apart from a few scattered short motifs there is no extensive sequence similarity between either bait region of the two forms of Limulus $\alpha_2$M and tick $\alpha_2$M (Fig. 7). This is in line with data on the bait regions of vertebrate $\alpha$Ms that differ greatly both in length and sequence (6). The lack of a Cys residue in the bait region of tick $\alpha_2$M is in agreement with tick $\alpha_2$M most likely being a dimer of non-covalently associated processed subunits (28).

In contrast to the dissimilarity between the bait regions of tick $\alpha_2$M and the two forms of Limulus $\alpha_2$M, there is a high degree of sequence identity between the N-terminal and C-terminal parts of the bait regions of the two forms of Limulus $\alpha_2$M. Despite this sequence identity, it is likely that the two forms are able to inhibit different proteinases because there is no similarity in the central part of the bait region. Previously, a significant degree of sequence identity within the bait region has only been observed among closely related $\alpha$Ms such as carp $\alpha_2$Ms (41) and the two isoforms of rat $\alpha_1$-inhibitor 3 (42, 43), suggesting that the two forms of Limulus $\alpha_2$M have a high degree of sequence identity too. This is further supported by the fact that even though peptide stretches covering almost 900 amino acid residues have been sequenced only 52 differences between the two forms have been observed (Table IV). In line with this agreement with tick $\alpha_2$M (28) shows that apart from a few scattered short motifs there is no extensive sequence similarity between either bait region of the two forms of Limulus $\alpha_2$M and tick $\alpha_2$M (Fig. 7). This is in line with data on the bait regions of vertebrate $\alpha$Ms that differ greatly both in length and sequence (6). The lack of a Cys residue in the bait region of tick $\alpha_2$M is in agreement with tick $\alpha_2$M most likely being a dimer of non-covalently associated processed subunits (28).
with this, in the separation systems devised so far, e.g. in Ref.
20, the two forms were found to co-purify.

Hypothetically the two forms might be encoded by a single
gene pair or by two gene pairs. To settle this question we
prepared Limulus αM from the hemolymph of two animals.
We obtained in both cases a partial separation of the two forms
in ion exchange experiments, strongly suggesting that they are
expressed by two gene pairs, similar to many other species.

REFERENCES
Chem. 264, 15781–15789
259, 12719–12724
10. Armstrong, P. B., Mangel, W. F., Wall, J. S., Hainfield, J. F., van Holde, K. E.,
Chem. 259, 12719–12724
J. Biol. Chem. 266, 19426–19431
Chem. 259, 8318–8327
15863–15869
chem. J. 347, 679–685
chem. J. 347, 679–685
Anal. Biochem. 176, 253–260
24. Bendixen, E. Halkier, T., Magnusson, S., Sottrup-Jensen, L., and Kristensen,
25. Sottrup-Jensen, L., Stepanik, T. M., Jones, C. M., Lønblad, P. B., Kristensen,
20–24
12117–12120
30. Tomiya, N., Aways, J., Kurono, M., Hanzawa, H., Shimada, I., Arata, Y.,
31. Smith, P. L., Skelton, T. P., Fiete, D., Dharmesh, S. M., Beraneck, M. C.,
267, 19140–19146
33. Tanaka, N., Nakada, H., Itoh, N., Mizuno, Y., Takamishi, M., Kawakami, T.,
68–74
35. Sottrup-Jensen, L., Stepanik, T. M., Jones, C. M., Lønblad, P. B., Kristensen,
38. Armstrong, P. B., Mangel, W. F., Wall, J. S., Hainfield, J. F., van Holde, K. E.,
Chem. 259, 12719–12724
Localization of Carbohydrate Attachment Sites and Disulfide Bridges in *Limulus* α₂-Macroglobulin: EVIDENCE FOR TWO FORMS DIFFERING PRIMARILY IN THEIR BAIT REGION SEQUENCES

Lise B. Husted, Esben S. Sørensen, Peter B. Armstrong, James P. Quigley, Lene Kristensen and Lars Sottrup-Jensen

doi: 10.1074/jbc.M208236200 originally published online September 5, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M208236200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 43 references, 22 of which can be accessed free at
http://www.jbc.org/content/277/46/43698.full.html#ref-list-1