Available online at www.sciencedirect.com
ScienceDirect
Developmental Biology 310 (2007) 54–70
DEVELOPMENTAL BIOLOGY
www.elsevier.com/developmentalbiology

The toposome, essential for sea urchin cell adhesion and development, is a modified iron-less calcium-binding transferrin

Hans Noll [a,*,1], Joy Alcedo [a,b,2], Michael Daube [b], Erich Frei [b], Emile Schiltz [c], John Hunt [a], Tom Humphreys [d], Valeria Matranga [e], Martin Hochstrasser [b,3], Ruedi Aebersold [f,g], Hookeun Lee [f], Markus Noll [b,*]

[a] *Department of Cell and Molecular Biology, University of Hawaii, School of Medicine, 1960 East-West Road, Honolulu, HI 96822, USA*
[b] *Institute for Molecular Biology, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[c] *Institute for Organic Chemistry and Biochemistry, Albert-Ludwigs-University Freiburg, Albertstr. 21, D-79104 Freiburg i. Br., Germany*
[d] *Pacific Biosciences Research Center, University of Hawaii, 1960 East-West Road, Honolulu, HI 96822, USA*
[e] *Istituto di Biomedicina e Immunologia Molecolare "Alberto Monroy", Via Ugo La Malfa 153, 90146 Palermo, Italy*
[f] *Institute of Molecular Systems Biology, ETH-Hönggerberg, Wolfgang-Pauli-Strasse 16, CH-8093 Zürich, Switzerland*
[g] *Faculty of Science, University of Zürich, CH-8057 Zürich, Switzerland, and Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103, USA*

Received for publication 23 March 2007; revised 9 July 2007; accepted 14 July 2007
Available online 24 July 2007

Abstract

We describe the structure and function of the toposome, a modified calcium-binding, iron-less transferrin, the first member of a new class of cell adhesion proteins. In addition to the amino acid sequence of the precursor, we determined by Edman degradation the N-terminal amino acid sequences of the mature hexameric glycoprotein present in the egg as well as that of its derived proteolytically modified fragments necessary for development beyond the blastula stage. The approximate C-termini of the fragments were determined by a combination of mass spectrometry and migration in reducing gels before and after deglycosylation. This new member of the transferrin family shows special features which explain its evolutionary adaptation to development and adhesive function in sea urchin embryos: (i) a protease-inhibiting WAP domain, (ii) a 280 amino acid cysteine-less insertion in the C-terminal lobe, and (iii) a 240 residue C-terminal extension with a modified cystine knot motif found in multisubunit external cell surface glycoproteins. Proteolytic removal of the N-terminal WAP domain generates the mature toposome present in the oocyte. The modified cystine knot motif stabilizes cell-bound trimers upon Ca-dependent dissociation of hexamer-linked cells. We determined the positions of the developmentally regulated cuts in the cysteine-less insertion, which produce the fragments observed previously. These fragments remain bound to the hexameric 22S particle in vivo and are released only after treatment of the purified toposome with reducing agents. In addition, some soluble smaller fragments with possible signal function are produced. Sequence comparison of five sea urchin species reveals the location of the cell–cell contact site targeted by the species-specific embryo dissociating antibodies. The evolutionary tree of 2-, 1-, and 0-ferric transferrins implies their evolution from a basic cation-activated allosteric design modified to serve multiple functions.

Keywords: Cell adhesion; Iron-less transferrins; Membrane synthesis; Posttranslational processing; Sea urchin; Thiol-protease; Toposome; WAP domain

* Corresponding authors. M. Noll is to be contacted at fax: +41 44 635 6829.
E-mail addresses: hansnoll1@comcast.net (H. Noll), Markus.Noll@molbio.unizh.ch (M. Noll).
[1] Current address: 6716 36th Ave NW, Seattle, WA 98117, USA.
[2] Current address: Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation, Maulbeerstrasse 66, CH-4058 Basel, Switzerland.
[3] Current address: Cilag AG International, Landis+Gyr-Strasse 1, CH-6300 Zug, Switzerland.

0012-1606/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2007.07.016

Introduction

A potential model system for the study of embryonic cell adhesion and positional information (Wolpert and Gustafson, 1961; Noll et al., 1979) was introduced upon Herbst's discovery (1900), more than a century ago, that in calcium-free seawater sea urchin blastulae dissociate into single cells, which, when returned to normal seawater, reaggregate into normally developing embryos. The molecule responsible for this Ca^{2+}-dependent adhesive activity was isolated by the use of an assay (Noll et al., 1979, 1981; Matranga et al., 1986) in which reaggregation was inhibited by Fabs specific for the cell–cell adhesion site and the reversal of this inhibition by the addition of the active membrane component (Beug et al., 1970). In addition, the activity could be removed from the cell surface with non-cytolytic concentrations of butanol in seawater, a treatment that renders the dissociated cells incapable of reaggregation and development in seawater, unless the cells were resupplied with the dialyzed extract (Noll et al., 1979) or with the purified adhesion glycoprotein, called toposome (Noll et al., 1985).

Surprisingly, the active component was similar to a previously known 22S glycoprotein complex (Malkin et al., 1965; Infante and Nemer, 1968; Ii et al., 1978; Ozaki, 1980; Harrington and Easton, 1982) whose function had been controversial. Past work, lacking a functional assay and assuming its exclusive occurrence in cytoplasmic "yolk granules", sought to assign the 22S complex some nutritional role (Williams, 1967; Shyu et al., 1986). However, no depletion of the 22S particles is observed in cell homogenates of embryos raised under conditions of starvation, a treatment that was expected to force utilization of stored nutrients, but, contrary to expectations, stunted further development (Scott et al., 1990). Hence, this observation is inconsistent with a proposed role in nutrition, but confirms the proposed function as a stored membrane component (Gratwohl et al., 1991). Furthermore, the 22S particles do not decrease during embryonic development, contrary to what is expected for nutrients required for development (Kari and Rottmann, 1985).

Although previous efforts in cloning the gene that generates the 22S particle from *Strongylocentrotus purpuratus* (Shyu et al., 1986, 1987) suggested that it is related to the vitellogenin family, the fragments that were sequenced were too short for an unambiguous identification of the protein structure. Here we describe the sequence of a full-length cDNA of the 22S particle from *Tripneustes gratilla*. Its coding region of 1344 amino acids has no similarity to any known vitellogenin. The cognate protein belongs to the transferrin family, whose members are characterized by two internal repeats that form two lobes connected by a short alpha-helical peptide (Anderson et al., 1987). The 22S particle has the additional novel features of a 280-amino acid intervening sequence within its second lobe and a further C-terminal extension of 240 amino acids. Furthermore, the protein lacks most of the five iron-binding amino acids D, Y, R, Y, and H present at specific positions in iron-transporting transferrins (Baker et al., 1987; Legrand et al., 1988; Baker and Lindley, 1992), which is consistent with the Ca^{2+}-binding function of the 22S protein in cell adhesion rather than transport. We also provide here the entire highly conserved sequence of its homolog in *Paracentrotus lividus*. The iron-less nature of the toposome protein is not unusual for a member of the transferrin family, since other iron-less members of this family continue to be discovered. Two of these are also membrane-associated (Morabito and Moczydlowski, 1994; McNagny et al., 1996), but their functions remain unknown.

The most remarkable feature of the 280-amino acid intervening sequence within the transferrin structure is its central function in early sea urchin development. Developmentally regulated proteolytic modifications in this region, which fail to disrupt the hexameric particle in vivo, have been elucidated and are described below. To understand their significance, it is necessary to relate these post-translational modifications to earlier observations.

The cDNA-derived amino acid sequences reported here are not those of the 22S particle but of its precursor, which is synthesized exclusively in the gonads and gut of the adult animal as a 180–190 kDa glycoprotein (Shyu et al., 1986). From the unfertilized eggs, the toposomes are isolated as 22S particles (Noll et al., 1985), which on reducing SDS gels appear as a single 170 kDa band (Kari and Rottmann, 1985; Noll et al., 1985). We now show that the N-terminus corresponding to this band results from a cut between amino acids 86 and 87 of the precursor, thus generating the mature toposome. By contrast, after fertilization, the particles, while still sedimenting at 22S, are further modified proteolytically. When isolated from the cytoplasm or from purified membranes of blastulae and analyzed by SDS-PAGE, these particles give rise, in addition to the 170 kDa protein, to four fragments ranging in molecular mass from about 70 kDa to 110 kDa with some species-dependent variations. All of the smaller fragments derive from the 170 kDa subunit (Kari and Rottmann, 1985; Noll et al., 1985; Matranga et al., 1986; Armant et al., 1986; Yokota and Kato, 1988; Lee et al., 1989; Scott and Lennarz, 1989; Gratwohl et al., 1991; Mallya et al., 1992), since new synthesis cannot be detected either in the unfertilized eggs or in the embryo up to the pluteus stage (Heifetz and Lennarz, 1979; Shyu et al., 1986). This developmentally controlled modification by proteases in the "yolk" granules (Yokota and Kato, 1988; Lee et al., 1989; Scott et al., 1990) is detectable in *T. gratilla* 6 h after fertilization and reaches its peak at the early gastrula stage (Noll et al., 1985). However, these proteolytic cuts fail to disrupt the hexameric particle, in which the peptide chains are held together by the extensive disulfide bonds characteristic of the transferrin family. Blocking these cuts by inhibitors of thiol-proteases arrests development at the blastula stage (Mallya et al., 1992; H.N. and A.J., unpublished results). Here we present the N-terminal sequences derived by Edman degradation of the five proteolytic fragments from *T. gratilla* in addition to three internal peptides of the major C-terminal fragment of *P. lividus* determined by amino acid sequencing. The approximate C-terminal ends of the fragments from *T. gratilla* were estimated by their band positions in reducing gels and mass spectrometry. An extensive analysis of the structural details of the mature toposome confirms and extends our previous conclusions concerning its function.

Materials and methods

Isolation of toposome cDNAs from T. gratilla and P. lividus

A cDNA expression library of poly(A)$^+$ RNA, isolated from the hindgut of *T. gratilla*, was constructed in the l UNI-ZAP XR vector with the Stratagene ZAP-cDNA synthesis kit, which generates cDNAs subcloned between the *Eco*RI site (5′ end) and *Xho*I site (3′ end) of the plasmid vector pBluescript SK(−). Initially, 8 positive cDNA clones were isolated from 260,000 plaques that were screened, according to the protocol supplied by Stratagene, with a 1:1500-fold diluted rabbit antiserum for expression of the toposome precursor protein. The antiserum was elicited against purified 22S particles isolated from *T. gratilla* blastula embryos (Noll et al., 1985). It was specific for the 22S particle and reacted with no other component when tested by immunoblotting of total embryo extract. Moreover, this antiserum recognized polypeptide epitopes of the 22S glycoprotein, as evident from its reaction with preparations from which the carbohydrate had been removed enzymatically (H.N. and C. Kellner, unpublished results). Similarly, seven monoclonal antibodies reacting with butanol extract and obtained in response to immunization with whole cells or butanol extract from blastula membranes of *T. gratilla*, immunoprecipitated only the 22S particle from total embryo extracts (Noll et al., 1985). It follows that the 22S glycoprotein shares no epitopes with other soluble proteins of the embryo. In subsequent screens, cDNAs including the entire coding region of the toposome precursor were isolated by hybridization of a P^{32}-labeled 0.62 kb *Eco*RI(vector)–*Hin*dIII DNA probe derived from the 5′ end of the longest toposome cDNA, ctp1, isolated in the immunoscreen. The two longest toposome cDNAs, cU5 and cU6, have identical 5′ ends, exceeding that of ctp1 by 926 bp, and hence probably correspond to the full-length mRNA.

To isolate toposome cDNAs from *P. lividus*, a cDNA fragment encoding the polypeptide between the tryptic peptides P2 and P3 (Fig. 1B) was synthesized by PCR of total hindgut RNA from *P. lividus* and the use of primers corresponding to the *T. gratilla* toposome sequence of P2 and P3. Subsequently, overlapping cDNAs of the 5′- and 3′-portion were synthesized with the SMART RACE cDNA amplification kit (Clontech) and primers of the *P. lividus* sequence between P2 and P3, as described by the user manual. The overlapping cDNAs comprised 4740 bp without the polyA tail, a length similar to the corresponding 4721 bp of the *T. gratilla* toposome cDNA sequence and in excellent agreement with the length of a single band of 5.0 kb observed by Northern analysis (V.M., unpublished result).

DNA sequence analysis

DNA sequences were analyzed on both strands with a DNA sequencer model 373A using dye terminators (Applied Biosystems Inc.). Searches for homologous DNAs and corresponding amino acid sequences and their alignments were performed with Blast programs (Blast and PSI Blast) as implemented by NCBI (http://www.ncbi.nlm.nih.gov/BLAST/), while alignments were carried out with the help of the pileup module of the GCG Wisconsin package (http://www.accelrys.com/products/gcg_wisconsin_package/index.html).

Microsequencing of tryptic peptides of toposomes from P. lividus

Toposomes, isolated from *P. lividus* blastula embryos as described previously (Noll et al., 1985; Matranga et al., 1986), were subjected to SDS-PAGE. After transfer to immobilon and staining with Coomassie blue, an attempt was made to determine the N-terminal sequence of each major band, which was not successful and thus suggested that the N-termini of the toposome bands might be blocked. Therefore, after separation by SDS-PAGE and transfer to immobilon, the 80 kDa band was eluted and the protein digested with trypsin. The resulting peptides were separated by HPLC on a C-18 reverse phase column, and the sequence of three peptides, P1, P2 and P3, was determined by Edman degradation at the Molecular Biology Core Laboratory, Case Western Reserve University, Cleveland. The sequences of P1 to P3 are highly homologous to the corresponding amino acid sequences derived from *T. gratilla* toposome cDNAs, as shown in Fig. 1B. It should be noted that the microsequence of P2, YTNPLELSPK, deviates from that derived from *P. lividus* toposome cDNAs, YTNPLWLSPK (Fig. 5). Although a number of polymorphisms were found among the *P. lividus* toposome cDNAs, no polymorphism was detected at the sixth position of P2.

Purification of toposomes

Toposomes were prepared essentially as previously described (Noll et al., 1985; Matranga et al., 1986; Cervello et al., 1992). *P. lividus* embryos of the hatching blastula stage were dissociated in Millipore-filtered calcium- and magnesium-free sea water containing 2 mM EDTA. The cells, collected by centrifugation, were suspended in 10 volumes of 3% 1-butanol in seawater. After repeated centrifugations, the clarified supernatant was dialyzed against phosphate-buffered saline (PBS) and concentrated by Amicon filters (30 kDa exclusion) to about 6 mg/ml. For further purification, 8 ml concentrate (8.6 A_{280}/ml) was applied to a Sephacryl column (200 cm long, 1.8 cm wide) containing 480 ml of S-400 (Pharmacia). Elution, monitored at A_{280}, was with PBS at a flux of 48 ml/h. Toposomes, verified by sucrose gradient analysis and reducing SDS-PAGE (Noll et al., 1985), emerged in a sharp peak after two broader peaks. Each 8 ml-fraction collected across the peak was dialyzed against distilled water and concentrated to 2–5 A_{280}/ml as described above. Analysis of these peak fractions by reducing SDS-PAGE revealed protein bands running at the same positions as those of *T. gratilla* blastula embryos shown in Fig. 6A.

T. gratilla unhatched blastula, hatched blastula, or prism stage was harvested at room temperature in 250 ml bottles by centrifugation for 3 min at 1000 rpm in a Sorvall centrifuge. The resulting 10–20 ml of packed embryos were resuspended in 200 ml of dissociation medium, DM [Ca-, Mg-free sea water, CMFSW (450 mM NaCl, 10 mM KCl, 25 mM Na_2SO_4, 2.5 mM $NaHCO_3$, 50 mM Tris–HCl, pH 8.0), containing 1 mM EDTA], and washed twice by centrifugation and resuspension in DM as before. The progress of dissociation was followed by inspection of resuspended samples after each washing step. After 15 min of constant gentle manual agitation, the dissociated cells were centrifuged again and resuspended in 200 ml of Millipore-filtered sea water (MFSW) containing 2.5% 1-butanol, and extracted for 5 min with constant gentle manual agitation. This suspension was spun at 9000 rpm four times or until the supernatant was clear. The extract was further concentrated with repeated additions of PBS in a Millipore Amicon cell (Membrane: Regenerated Cellulose NMWL 30,000) until free of butanol. The final concentrate, 10–15 mg in 5 ml or less, was applied to a Sephacryl 16/60 S-300 HR column attached to a FPLC ÄKTA prime (GE Healthcare, former Amersham Biosciences) UV-monitored fraction collector. Elution was with PBS at constant pressure below 0.15 MPa (1.48 atm) at a flow rate of 0.5 ml per min. The 22S complex emerged as a sharp peak after a single even sharper peak eluting at the void volume. (When the purified material was reapplied to the column, it emerged at the same position without the preceding peak.) The purified material was stored frozen at −85°C.

Treatment of purified toposomes with PNGase F and analysis by PAGE

To remove N-linked oligosaccharides, toposomes were incubated with PNGase F (Enzymatic CarboRelease kit from QA-Bio). No O-linked oligosaccharides were detected in toposomes, as no difference in mass was detectable after treatment with *O*-Glycosidase (Enzymatic CarboRelease kit from QA-Bio). Untreated or treated toposomes were analyzed by 8% SDS-PAGE under reducing conditions in a Hoefer SE 260 Mighty Small electrophoresis apparatus. The running distance for the resolving gel was 8 cm. Protein molecular weight standards (Biorad) were Myosin (200 kDa), β-galactosidase (116.25 kDa), phosphorylase b (97.4 kDa), serum albumin (66.2 kDa), and ovalbumin (45 kDa). Gels were stained with silver (Biorad Silver Stain Plus kit). To test which protein bands after deglycosylation with PNGase F were derived from which *N*-glycosylated band, the gel was stained with imidazole/Zn, the stained glycosylated bands were cut out, destained (Zn Stain and Destain Kit from Biorad), eluted in PNGase F buffer (Enzymatic CarboRelease kit from QA-Bio), treated with PNGase F, and again analyzed with PNGase F-treated toposomes by SDS-PAGE.

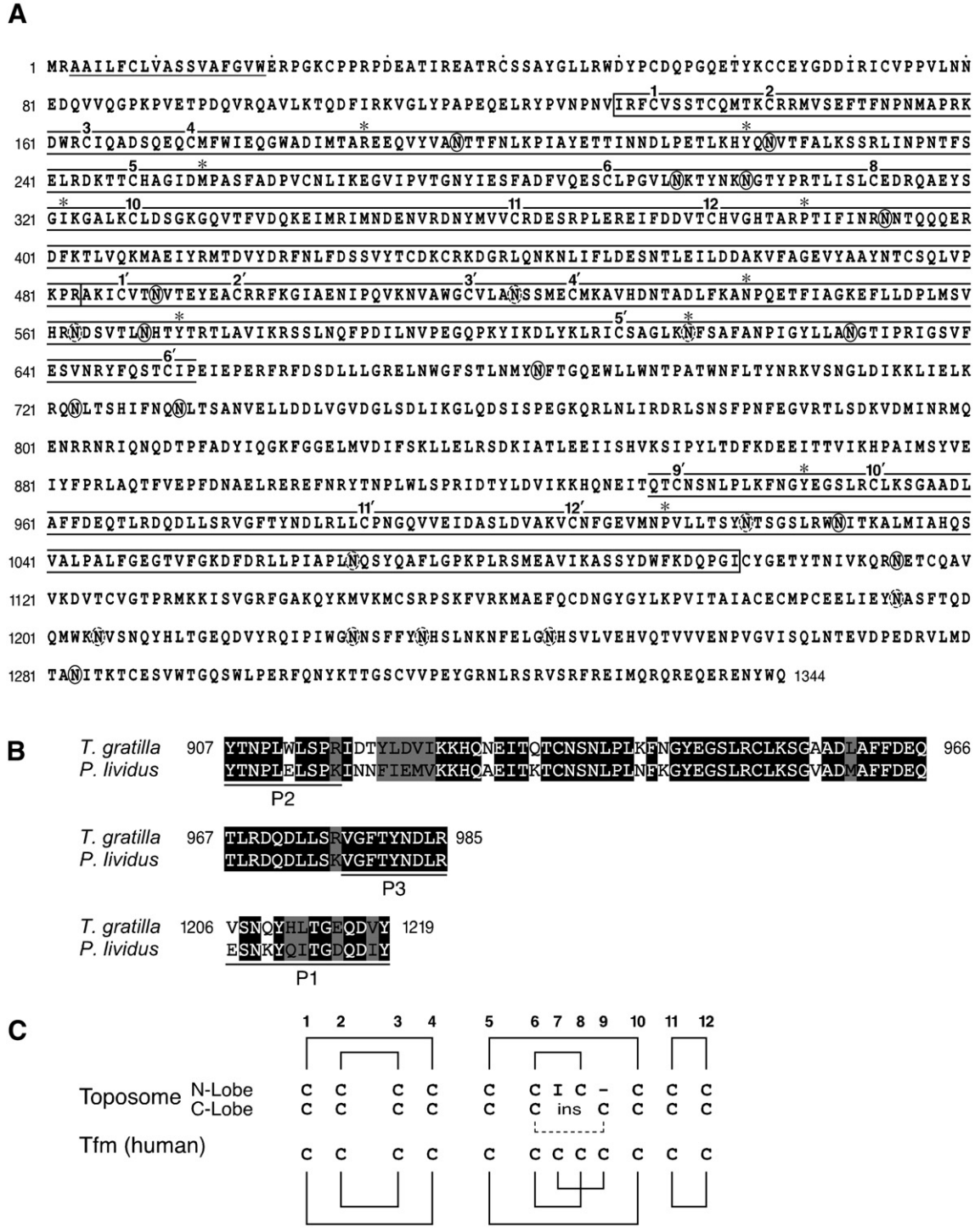

Fig. 1. Amino acid sequence of toposome subunit precursor with details of its derivation and substructure. (A) Amino acid sequence derived from *T. gratilla* cDNA encoding toposome precursor. The sequence homologous to the transferrin family is framed, the additions (amino acids 1–129 and 1101–1344) and the insertion (654–932) are not framed. The canonical cysteines (1–6, 8, 10–12 in the N-lobe and 1′–6′, 9′–12′ in the C-lobe of the transferrin part of the toposome) are indicated, and the positions corresponding to the iron-binding ligand amino acids in the diferric transferrins are marked by asterisks. The signal sequence is underlined, and potential glycosylation sites are marked by solid circles for the strong NXT type and dotted circles for the weak NXS type. The GenBank accession number for the full-length toposome cDNA of *T. gratilla* is AY026514. (B) Confirmation of cDNA-derived amino acid sequence of toposome subunit from *T. gratilla* by comparison with near identical sequences of three tryptic peptides (underlined) from *P. lividus* determined by microsequencing. The *P. lividus* sequence between peptide P2 and P3 was derived from the PCR product of genomic *P. lividus* DNA amplified by the use of primers corresponding to the *T. gratilla* sequence. Identities in amino acids are shaded in black, similarity in gray. (C) Disulfide bridges between conserved cysteines in the transferrin domains of toposomes. The disulfide bridges linking the canonical cysteines are shown by brackets. In the N-lobe of the transferrin part, cysteine 7 is replaced by isoleucine and the section corresponding to cysteine 9 is deleted. In the C-lobe, the sequence including C7 and C8 is deleted and replaced by an insertion, while C6 may be linked to C9 as indicated by a dashed line.

Mass spectrometry of mature 170 kDa toposomes and its proteolytically modified products

For mass spectrometric analysis, purified *T. gratilla* toposomes from 16 h-old unhatched blastula embryos were deglycosylated with PNGase F and toposome bands separated by SDS-PAGE (Fig. 6A). Each Coomassie-stained protein band was excised, destained, reduced and alkylated, digested with trypsin, and eluted from the gel by the use of the ProteoPrep Reduction and Alkylation Kit (Sigma) and the Trypsin Profile IGD Kit (Sigma).

The tryptic peptides of each deglycosylated toposome band were subjected to reversed-phase microcapillary chromatography (RPLC)-electrospray ionization (ESI) tandem mass spectrometry (MS/MS). The set-up of the RPLC system was as described previously (Lee et al., 2004). ESI-based RPLC-MS/MS (LTQ Thermo Finnigan San Jose, CA) analyses were carried out with a binary HP1100 pump (Agilent Technologies, Palo Alto, CA) attached to a 100 μm × 2.0 cm sample-trapping precolumn and a 75 μm × 10.5 cm fused-silica microcapillary reversed phase column. The columns were in-house packed with a C18 resin (5 μm, 200 Å Magic C18AQ, Michrom BioResources Inc., Auburn, CA). Sample volumes of 4 μl in solvent A (0.1% formic acid) were loaded onto the precolumn at a flow rate of 4 μl/min for 5 min. After loading, tryptic peptides were separated and analyzed at a 250 nl/min flow rate with a linear binary solvent composition gradient of 5% to 35% solvent B (100% acetonitrile) over 50 min as described (Lee et al., 2004). Peptides eluted from the microcapillary column were ionized at an electrospray voltage of 2.0 kV and subjected to mass spectrometry (MS). The three most abundant peptide ions with peak intensities above 10,000 ion counts in each MS scan were selected for collision-induced dissociation (CID) and subjected to a second MS analysis (MS/MS). The procedure alternated between one MS scan over the m/z range 400–1800 and three MS/MS scans. The specific m/z value of the peptide selected for second MS analysis was excluded from MS/MS for 2 min.

The spectra of the second MS were searched against a nonredundant yeast protein sequence database and the predicted *T. gratilla* toposome precursor amino acid sequence by the use of the computer algorithm, SEQUEST (Eng et al., 1994). The intensity value of each tryptic peptide was derived from a high-resolution Pep3D Image of the first MS analysis with a selected intensity range below saturation (Li et al., 2004). The integrated intensity of the manually selected peak corresponding to each peptide identified in the second MS analysis with high confidence was measured by the use of the software ImageJ. The results have been compiled in Table S1. As only peptides with $m/z \geq 400$ and $z \geq 2$ could be identified, no peptides shorter than octapeptides could be detected and hence listed in Table S1.

N-terminal sequencing of protein fragments

After PAGE as described above, the proteins were transblotted from the gel to a PVDF membrane in Towbin buffer containing 0.1% SDS by means of a Biorad Semi-Dry Transblot Transfer Cell for 1 h at 15 V (1–2 mA/cm^2). After Ponceau S-staining, the bands were cut out for protein sequencing. The samples were analyzed by Edman-degradation in a gas-phase sequencer (model 477A; Applied Biosystems, Foster City, CA, USA), with on-line HPLC-identification (model 120A) of the amino acid phenylthiohydantoins performed according to the manufacturer's recommendations. To determine the N-termini of T, N1, C1, N2, C2, and N3 (Fig. 6A) listed in Table S2, 6–14 residues were sequenced by Edman-degradation.

Results

The subunits of the toposome precursor are modified transferrins lacking most iron ligands

The amino acid sequence of the *T. gratilla* toposome precursor derived from its full-length cDNA is shown in Fig. 1A, which conspicuously includes a signal peptide and a cysteine-rich internal repeat. The signal peptide (amino acids 3–19) was expected for a secreted protein made maternally in the gut and transported to the oocyte. The protein also contains 24 potential *N*-glycosylation sites, 14 strong sites of the NXT type and 10 weak NXS sites (Fig. 1A). Extensive glycosylation of the 22S particle has been reported previously (Harrington and Easton, 1982), and is exclusively of the N-linked polymannose type (Armant et al., 1986; Scott and Lennarz, 1989). Of the strong sites, 12 are identical in all species listed in Fig. 5. Interestingly, in the mature toposome, some sections of the N- and C-terminus as well as most of the large cysteine-free section in the middle are free of glycosylation sites (Fig. 1). As will be shown below, some of these sections are the targets of the developmentally programmed proteolytic modifications described later.

A genomic DNA sequence of the 5′ end, including the first two introns, of the *S. purpuratus* gene encoding the 22S particle precursor has been published previously (Shyu et al., 1987). The first 13 amino acids of the *Tripneustes* toposome precursor are identical to the derived amino acid sequence of the *Strongylocentrotus* gene product, and the identity throughout the first two exons is still above 60%. The cDNA-derived amino acid sequence of *T. gratilla* also has extensive regions of homology with the amino acid sequence of the 22S particle from *P. lividus* determined subsequently (Fig. 5). Independent confirmation was obtained by microsequencing of three tryptic peptides, P1–P3, derived from the 80 kDa fragment of *P. lividus* (Fig. 1B). These peptides, 9–14 amino acids in length, at positions 907–916, 977–985, and 1206–1219, are all part of the C-terminal third of the molecule (Fig. 1A).

While no sequence similarity to vitellogenins could be detected, the alignments in Fig. 2 clearly identify toposomes as members of the transferrin (Tf) gene family. All known vertebrate and some insect Tfs are characterized by two internal repeats of about 330 amino acids, which form two lobes connected by a short alpha-helical peptide. Each lobe consists of two domains separated by a cleft containing the iron-binding site (Legrand et al., 1988; Baker and Lindley, 1992). The twofold internal homology is expressed by 12 cysteine residues in conserved positions (Fig. 2), evidently the result of an ancestral gene duplication. The disulfide bridges linking the canonical cysteines (Williams, 1974; Williams et al., 1982a; MacGillivray et al., 1982, 1983; Mazurier et al., 1983; Metz-Boutigue et al., 1984) are responsible for the structural and evolutionary stability of the molecule (Williams, 1982) and its 3D conformation. One of these bridges, C7–C9, and most of the connecting loop is absent in the N-lobe of the toposome. Similarly, two bridges are not present in the C-lobe because of a deletion involving the canonical positions 7 and 8 (Fig. 2). Apparently, the overall structure is conserved with 5 out of 6 disulfide bridges still able to form (Fig. 1C). In the absence of an X-ray crystallographic analysis, we reconstructed in Fig. 4B the 3D folding pattern in analogy to that of Baker et al. (1987) as a working hypothesis that explains many of our observations.

In the cases listed for comparison in Fig. 2, the five iron ligands, D, Y, R, Y, and H, have been invariant both with respect to the amino acid involved as well as to their position in the polypeptide chain. Based on this comparison, toposomes lack most of the Fe-ligands. Only present are the first tyrosine in the

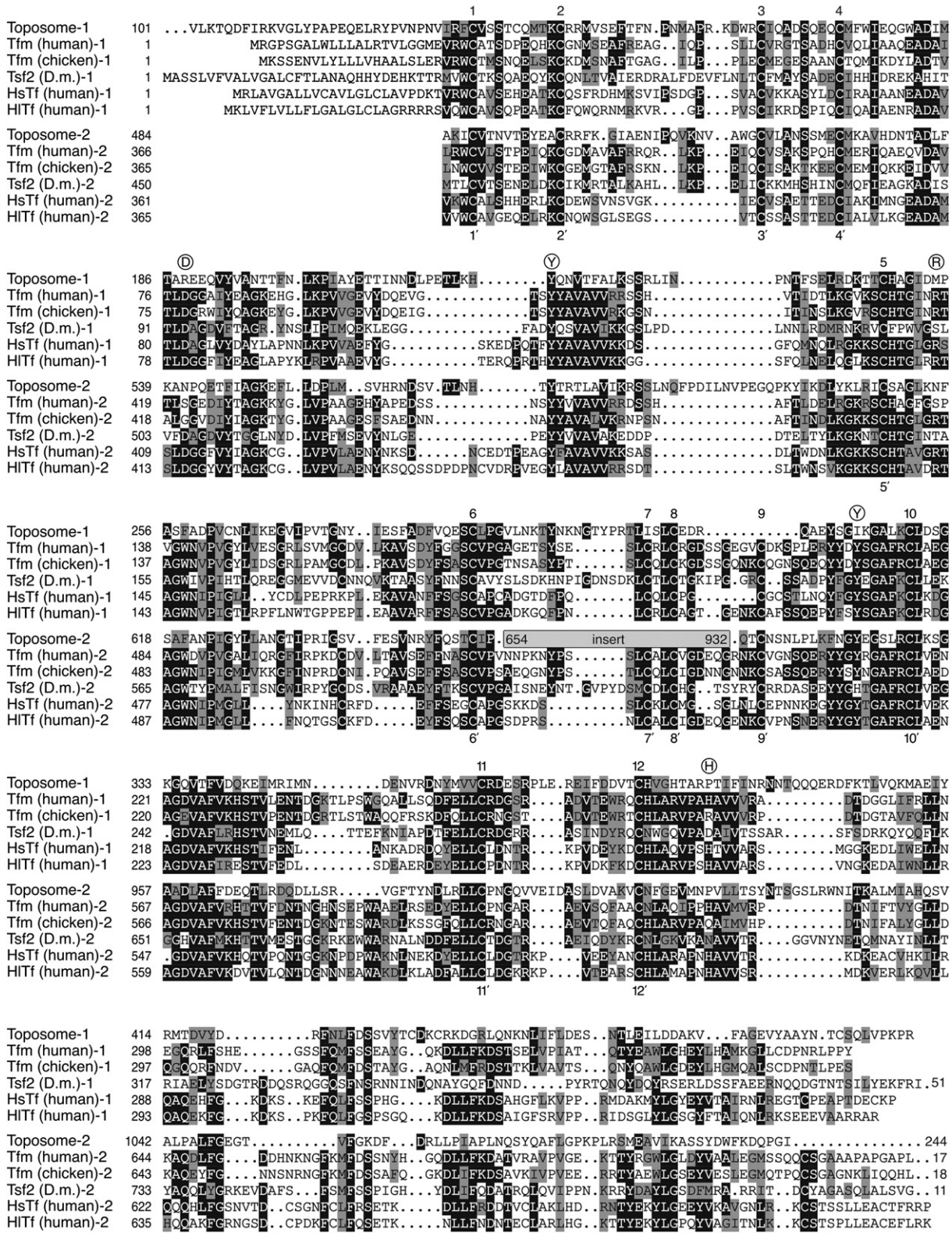

Fig. 2. Identification of the toposome subunit polypeptide as a member of the transferrin gene family. Alignment of the predicted amino acid sequence of the *T. gratilla* toposome with melanotransferrin (Tfm), transferrin Tsf2 of *D. melanogaster* (CG 10620), and serum- (HsTf) and lactotransferrin (HlTf). The N-lobe of each group (upper blocks) is aligned with its cognate C-lobe (lower blocks). The canonical cysteines are numbered 1–12 in the N-lobe and 1′–12′ in the C-lobe, and the positions of the diferric ligand amino acids are marked by the corresponding symbols D, Y, R, Y, and H in circles. The *T. gratilla* sequence begins with amino acid 101 and ends with 1100. The remaining 244 residues are listed in Fig. 1. The number of residues similarly cut off in Tfm (human), Tfm (chicken), and Tsf2 is indicated at the right margin. The position of the insertion in the C-lobe of the toposome subunit (654–932) is shown as a gray rectangle. Identities in amino acids are shaded in black, similarities in gray.

N-lobe and both tyrosines in the C-lobe (Fig. 2). While this seems to be insufficient to support Fe-binding, the significance of these tyrosines for Ca-binding is uncertain in the absence of any identification of the Ca-ligands by crystal structure analysis. Most mammalian Tfs, i.e., the serum- and lacto-Tfs, are diferric except for the monoferric melanotransferrin, which is expressed in human melanomas as well as in cells of certain fetal tissues (Rose et al., 1986; Alemany et al., 1993; Richardson, 2000). Other monoferric Tfs have been isolated from a frog, *Xenopus laevis* (Moskaitis et al., 1990), and a crustacean, the crayfish *P. leniusculus* (Liang et al., 1997). The function of most of these mono-ferric Tfs is unknown. By contrast the function of the zero-ferric pICA is known. Like the toposome, it binds Ca^{2+} instead of Fe^{3+} (Wuebbens et al., 1997).

Structural modifications are consistent with known toposome functions

The most unusual feature of the toposome polypeptide is the insertion of some 280 amino acids in the second lobe and two additions, one of 129 amino acids at the N-terminus of the precursor molecule and another of about 240 amino acids at the C-terminus (Figs. 2 and 4A). As shown in Fig. 1A, the insertion interrupts the C-lobe after the canonical cysteine 6 and replaces the section containing cysteines 7 and 8. Therefore, C6 might form a disulfide bridge with C9 in the C-lobe of the toposome (Fig. 1C). This is attractive because the insertion carries no cysteines (Figs. 1A and 5).

In order to evaluate the significance of the almost 50% of the toposome sequence not homologous to transferrins, we searched for known motifs in the N-terminal section, in the insertion, and in the C-terminal peptide. A WAP domain was found near the N-terminus (amino acids 21–77). This domain, which is shared by certain protease inhibitors (Schalkwijk et al., 1999), is characterized by 8 cysteines in fixed positions that are linked as indicated in Fig. 3A. In toposomes, C2 and C7 are missing and hence the bridge linking these two cysteines. The presence of the WAP domain in the precursor makes sense because the postulated protease inhibiting activity is thought to protect the precursor during its transport from the gut to the oocyte.

In addition to the N-terminal WAP domain, a cystine knot (Meitinger et al., 1993; Murray-Rust et al., 1993) was detected near the C-terminus (amino acids 1101–1189; Fig. 4A). Again the 3D structure of this domain, which occurs in some multimeric extracellular glycoproteins, is determined by cysteines in specific positions and linked in a characteristic pattern (Fig. 3B). Here too, the toposomes and two other glycoproteins listed lack Cb, which is thought to be involved in intermolecular disulfide bonds (Isaacs, 1995; Katsumi et al., 2000; Bell et al., 2001). In toposomes C3 and C4 are missing as well, which leaves C6 and C1 without partners. The Norrie disease protein NDP (Fig. 3B) forms trimers via disulfide bonds between Cx, Cb, and Cd of the cystine knot (Katsumi et al., 2000). We therefore propose that Cx and Cd, which are conserved in toposomes, form disulfide bonds Cx1–Cd2, Cx2–Cd3, and Cx3–Cd1 between toposome subunits to generate the trimeric form of toposomes. In any case, the presence of such a domain in toposomes is consistent with the strikingly conserved positions of the 8 cysteines present in all 5 species and with the known effect of this domain in promoting subunit assembly when located near the C-terminus of glycoproteins.

We have been unable to find any domains in the 280 amino acid insertion that may shed light on its function. Although alignments suggestive of a partial MAP-kinase domain could be obtained, its significance is not clear. However, this cysteine-less domain is the target of the developmentally regulated post-translational modifications described below.

Extracellular transferrins bind to the cell surface either by way of the known transmembrane Tf-receptor (Newman et al., 1982) or, as in the case of melano-Tf (Alemany et al., 1993), by a lipid-inositol (GPI) anchor (Low, 1987). For toposomes such an anchor, however, is unlikely because of the absence of the requisite stretch of hydrophobic amino acids (Low, 1987).

To gain a better idea of how the 280 amino acid insertion and N- and C-terminal additions would affect the transferrin 3D structure, as defined by the conserved S-S links, we added these regions in Fig. 4B to the appropriate positions of the 3D diagram of Baker et al. (1987) assuming, for simplicity, globular structures for these domains. Even if these turn out not to be globular, the picture gives an idea of the size of these added domains in space and what to expect for the hexamer resulting from the interaction of two trimers, each attached to the surface of a different cell.

Extent of sequence identity of protoposomes from five different species is remarkable

In Fig. 5, the derived amino acid sequence of the *P. lividus* toposome precursor (protoposome) is aligned with those of *T. gratilla* and the two Japanese species *Pseudocentrotus depressus* (Unuma et al., 2001) and *Hemicentrotus pulcherrimus* (Yokota et al., 2003). We have added to this alignment the sequence of *S. purpuratus* (Brooks and Wessel, 2002) after correction for obvious sequencing mistakes.

The extraordinary degree of identity between these sequences (Fig. 5) is not seen in the iron-binding members of the transferrin family (Fig. 2). The deviations from identity are shown by their individual amino acids in Fig. 5 to highlight probable sites of antigenic specificity. We also marked the WAP and Cys-knot domains to show that the domain-specific conserved amino acids are present in all five toposome precursors, thus further corroborating the identification of these domains.

Edman degradation and mass spectrometry of the deglycosylated toposomes confirm the expected removal of the WAP domain and establish the N-terminal sequence of the mature toposome as well as those of the proteolytic fragments produced post fertilization

To analyze the mature toposomes and their developmentally regulated proteolytic processing (Kari and Rottmann, 1985; Noll et al., 1985; Armant et al., 1986; Yokota and Kato, 1988;

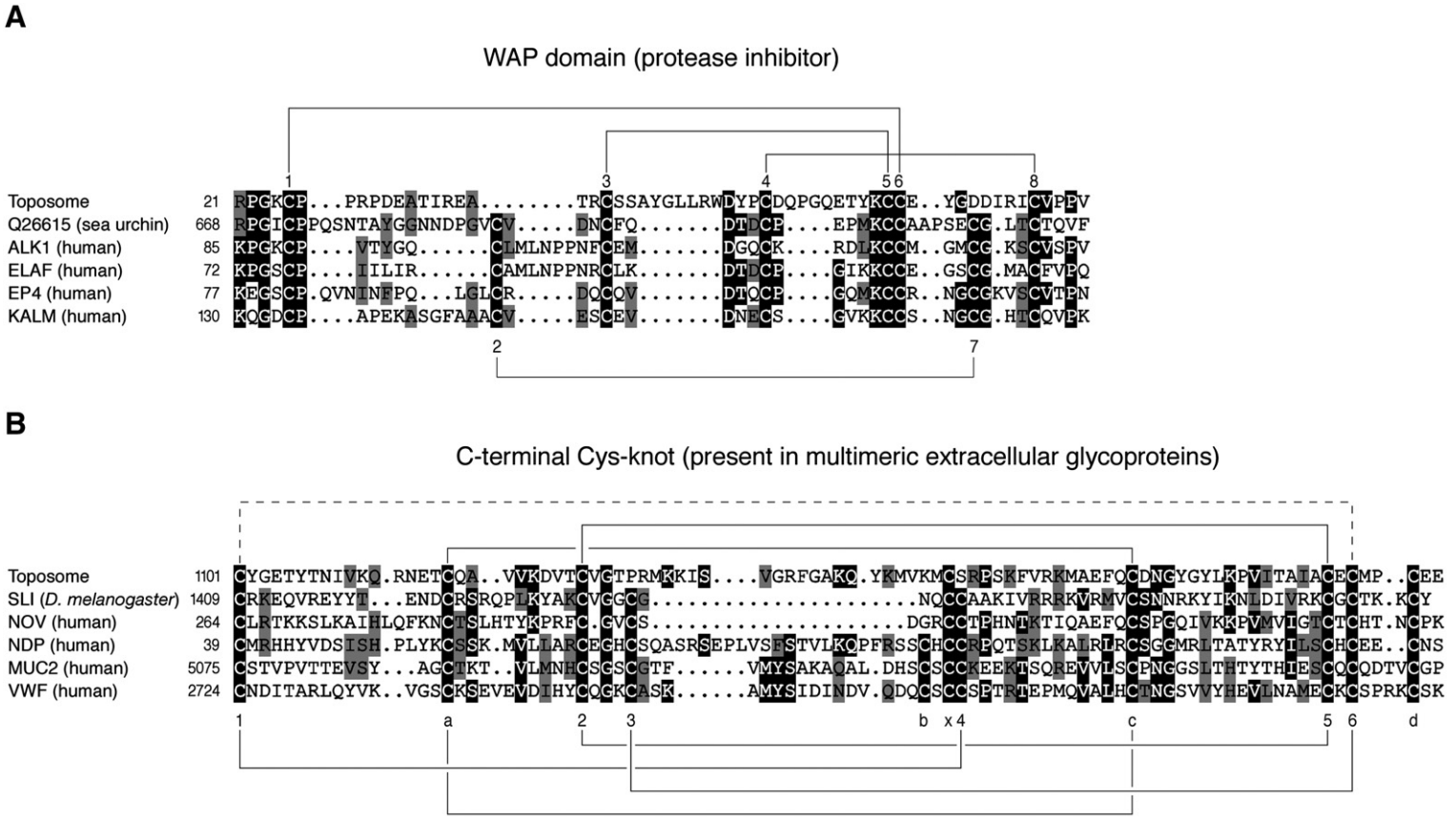

Fig. 3. The toposome subunit includes a WAP domain at its N-terminus and a C-terminal cystine knot. (A) WAP domain. Alignment of the conserved WAP domain of the toposome (*T. gratilla*) with the following 5 other representatives: Q266615 sea urchin cortical granule protein (Wessel, 1995), ALK1 mucus proteinase inhibitor (Stetler et al., 1986), ELAF elastase-specific inhibitor (Saheki et al., 1992), EP4 major epididymis-specific protein E4 (Kirchhoff et al., 1991), KALM Kallmann syndrome protein (Legouis et al., 1991). The WAP-specific cysteines, numbered 1–8, are linked as shown. The toposome lacks the linked cysteines 2 and 7. Identities in amino acids are shaded in black, similarities in gray. (B) C-terminal cystine knot. Alignment of the Cys-knot domain of the toposome (*T. gratilla*) with the following 5 other representatives: SLI slit protein (Rothberg et al., 1990), NOV human Nov protein (Martinerie et al., 1994), NDP Norrie disease protein (Meitinger et al., 1993), VWF von Willebrand factor (Marti et al., 1987; Katsumi et al., 2000), MUC2 human intestinal mucin (Gum et al., 1992; Bell et al., 2001). The Cys-knot-specific cysteines, numbered 1–6, a–d, and x are linked as shown. In the toposome, cysteines 3, 4, and b are absent, leaving 6 and 1 without their partner. We propose that 1 is linked to 6, while x and d are involved in inter- rather than intramolecular disulfide bridges stabilizing a trimeric toposome. Identities in amino acids are shaded in black, similarities in gray.

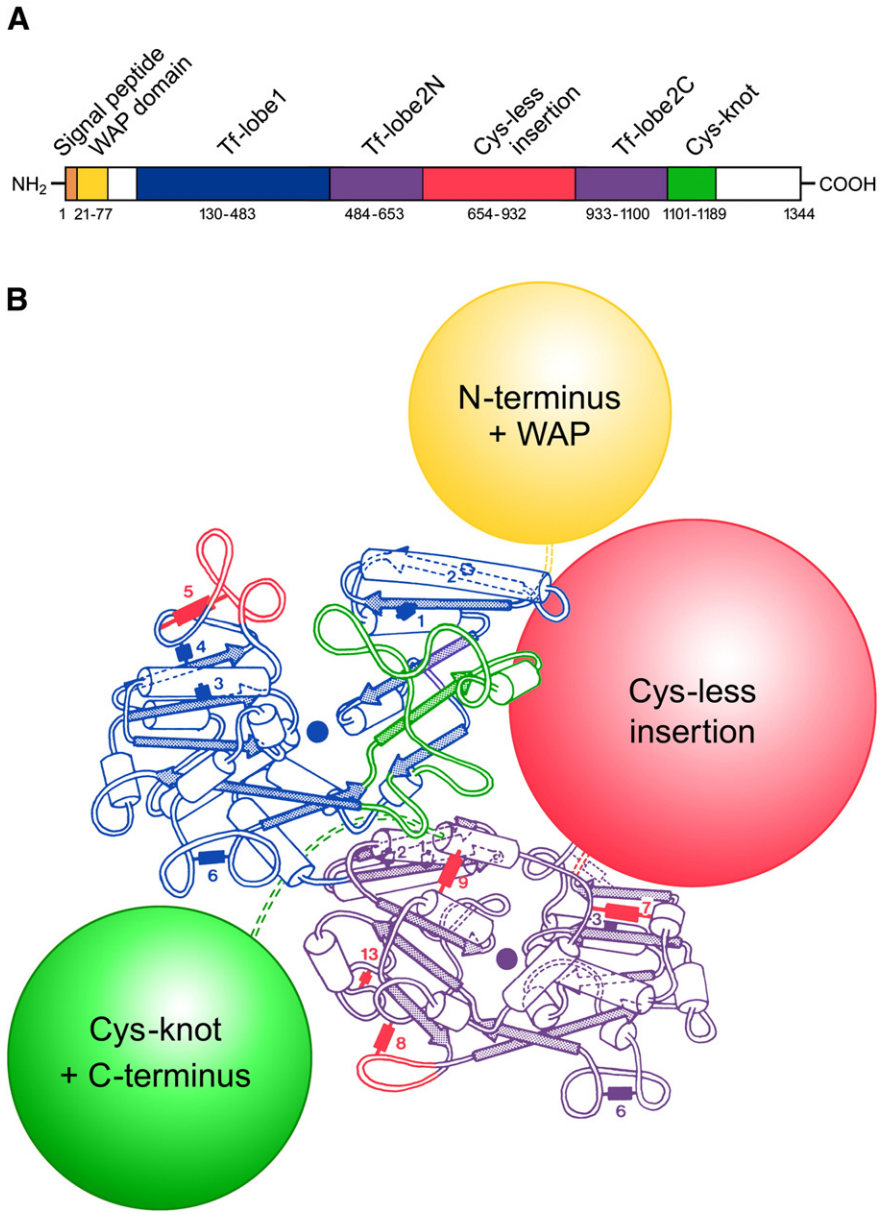

Fig. 4. Toposome domain structure in linear and 3D representation. (A) Linear domain structure of the toposome precursor. The various domains are illustrated in different colors with the beginning and end indicated by the numbers of the corresponding amino acids of the *T. gratilla* toposome precursor. (B) 3D diagram of *T. gratilla* toposome precursor modified from that of lactoferrin (Baker et al., 1987). α-helices are shown as cylinders, β-strands as arrows, position of iron atoms for ferric Tfs as solid circles, disulfide bridges (numbered according to Williams, 1982) as solid bars. The N-terminal half (N-lobe) is at top, the C-terminal half (C-lobe) at bottom, their relative orientations are related to a twofold screw axis. The three colored spheres illustrate the two additions and the insertion shown linearly in panel A, estimated under conditions of globular chain folding. The attachment points to the canonical transferrin framework are shown by stippled double lines. The external small loop and the large loop, including an α-helix and β-strand, in green correspond to the variable sequence predicted to be the epitope for Fabs capable of dissociating embryos into single cells. Comparison with Fig. 2 shows that the loops in red are not present in toposomes and that the disulfide bridges marked in red are missing because one or both cysteines are absent in toposomes.

Lee et al., 1989; Scott and Lennarz, 1989; Scott et al., 1990; Gratwohl et al., 1991; Mallya et al., 1992), 22S particles from *T. gratilla* were purified on Sephacryl columns and analyzed on a reducing SDS-polyacrylamide gel (Fig. 6A). The N-terminal sequences of the toposomes and their proteolytic fragments were determined by Edman degradation (cf. Materials and methods). In addition, the purified fragments were sequenced by mass spectrometry. The results are summarized in Fig. 6. The SDS gel shows three stages of development: unhatched blastula, hatched blastula, and early prism stage before and after deglycosylation by PNGase F (Fig. 6A). In confirmation of previous results (Noll et al., 1985), early stages of embryonic development show three strong bands corresponding to 170, 111, and 80 kDa (lanes 1a and 2a in Fig. 6A). The corresponding bands after deglycosylation in lanes 1b and 2b are T, the mature protein, and N1 and C1, the products of the first cut by the developmentally regulated enzyme in the "yolk" vesicles. In addition, there is a weak doublet at 105/101 kDa and

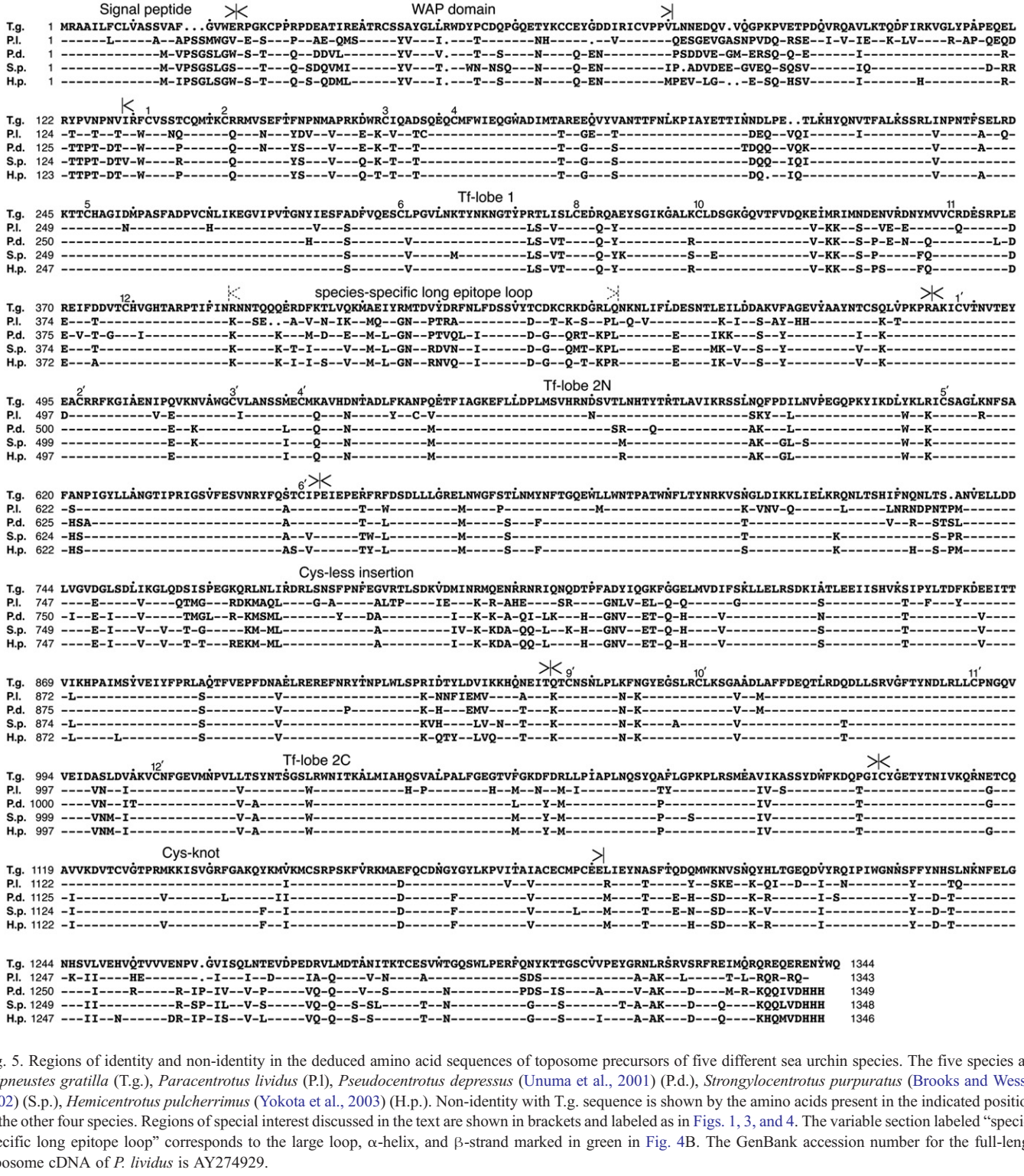

Fig. 5. Regions of identity and non-identity in the deduced amino acid sequences of toposome precursors of five different sea urchin species. The five species are: *Tripneustes gratilla* (T.g.), *Paracentrotus lividus* (P.l), *Pseudocentrotus depressus* (Unuma et al., 2001) (P.d.), *Strongylocentrotus purpuratus* (Brooks and Wessel, 2002) (S.p.), *Hemicentrotus pulcherrimus* (Yokota et al., 2003) (H.p.). Non-identity with T.g. sequence is shown by the amino acids present in the indicated positions of the other four species. Regions of special interest discussed in the text are shown in brackets and labeled as in Figs. 1, 3, and 4. The variable section labeled "species-specific long epitope loop" corresponds to the large loop, α-helix, and β-strand marked in green in Fig. 4B. The GenBank accession number for the full-length toposome cDNA of *P. lividus* is AY274929.

a weak band at 71 kDa, which becomes C2 after deglycosylation, while the doublet converts to a single 68 kDa band N2 (Fig. 6A and Table S2), which suggests that the members of the original doublet differed only in their carbohydrate content.

As development proceeds, the split products increase at the expense of the mature glycoprotein until this has disappeared at the prism stage (Fig. 6A, lane 3a). By contrast, the originally weak bands resulting from secondary cuts strongly increase after gastrulation at the expense of the bands produced by the first cleavage (cf. lanes 2a and 3a of Fig. 6A).

The scheme in Fig. 6B summarizes the time course of the proteolytic processing from unhatched blastula embryos to early prisms. The N-termini have been sequenced by Edman degradation of the deglycosylated bands (Fig. 6A, bands T, N1, N2, C1, C2, and N3), while their C-termini have been estimated from the molecular mass associated with the gel band

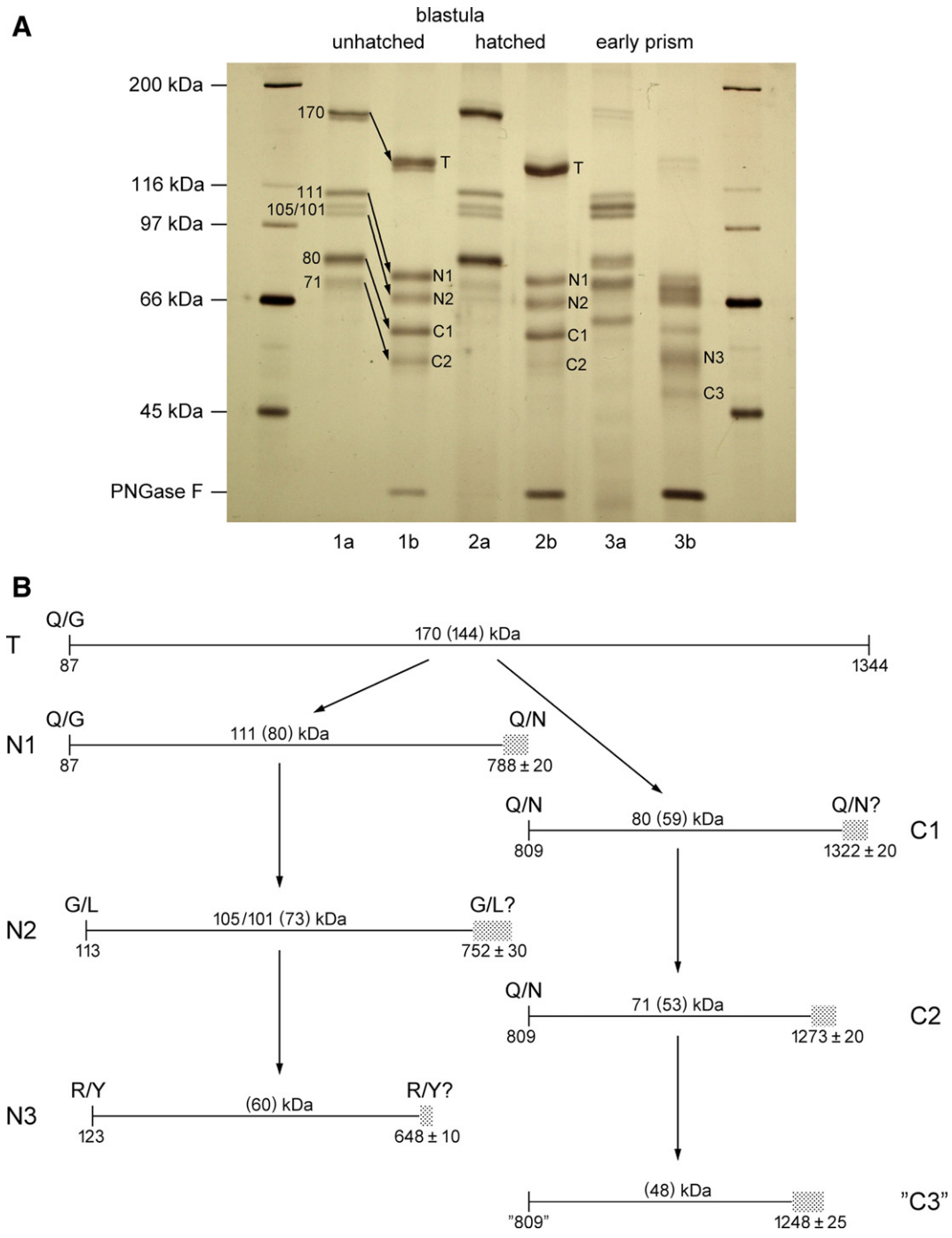

Fig. 6. (A) Analysis of mature and proteolytically modified toposomes from blastula and early prism stage. Purified toposomes from 16 h-old unhatched and 23 h-old hatched blastula, and from 44 h-old early prism stage embryos were analyzed before (a lanes) and after (b lanes) deglycosylation with PNGase F in an 8% SDS-polyacrylamide gel. For calibration, markers of the known molecular weights indicated were run on the left and right. Toposome bands are labeled as indicated in B. Arrows show the origin of deglycosylated from fully glycosylated toposome bands. (B) Scheme of proteolytic processing of toposomes. Mature toposomes (T) are processed by proteolytic cleavage within the Cys-less insertion of the C-lobe sequentially, as indicated by arrows. The molecular weights of mature toposomes and their proteolytically processed N-terminal (N1, N2, N3) and C-terminal (C1, C2, C3) moieties are indicated in kDa with those of the deglycosylated forms in parentheses. N-termini determined by Edman degradation are marked by position as well as N-terminal and preceding amino acid, while C-terminal positions are indicated by the range within which they are most probable to be found. For a more detailed explanation, see text.

positions and the mass spectrometry of the tryptic peptides of each band (Tables S1 and S2). While the N-termini, indicated in Fig. 6B by their position in the amino acid sequence, are precise, the C-termini are shown as the range of the best estimate of the two methods. Although the C-terminal ends presumably are the result of equally precise cuts by the vesicle enzyme, sequential amino acid sequencing of the fragments has not been attempted, and the cutting point is difficult to predict without knowing the exact specificity of the enzyme. The only experimentally verified cuts are Q/N, G/L, and R/Y (Fig. 6B).

From the data in Fig. 6, the following picture emerges. Before arriving in the oocyte, the mature toposome is produced by the "maturation" enzyme by a Q/G cut. The resulting mature toposome T extends from amino acid 87 to the C-terminus at amino acid 1344. This extrapolation appears reasonable, as the last peptide determined by MS/MS analysis at a high level of confidence ends at amino acid 1318 and the tryptic peptides covering the 26 C-terminal amino acids are all shorter than hexapeptides and hence not detectable by MS/MS analysis (Table S1). Thus, the 1258 amino acids of the toposome have a molecular mass of 144 kDa (Table S2).

The first cut after fertilization by the enzyme present in the cytoplasmic vesicles produces the 111 kDa and 80 kDa protein bands (Noll et al., 1985; Fig. 6B). From Edman degradation and mass spectrometry, we know that the larger fragment contains the N-terminus and the smaller the C-terminal end of the parent molecule. Moreover, microsequencing of three peptides (P1, P2, P3) from the 80 kDa fragment of *P. lividus* yielded sequences (Fig. 1B) localized within the C-terminal moiety of the parent molecule. The cut occurs in the unprotected cysteine-less insertion. As development progresses, these two fragments are further reduced, as obvious from Fig. 6 and Tables S1 and S2.

The vesicle enzyme cuts between Gln808 and Asn809 (Q/N in Fig. 6B). Subsequently, the fragment N2 is produced by a 112/113 G/L cut, possibly by the same enzyme. The C-terminus of this fragment has also been further reduced, conceivably also by a G/L cut at 749/750. Similarly, the C-terminal fragment C1 is reduced at its C-terminus, possibly by a Q/N cut at 1303/1304, which is within the limits of the determined C-terminus and the only Q/N in the entire C-terminal portion. Further shortening of C1 yields C2, ending near position 1273 (Fig. 6B).

As the deglycosylated proteolytically processed toposomes, purified from early prism stage, result in less well resolved bands (Fig. 6A, lane 3b), not all of their N-termini could be determined by Edman degradation. Thus, the broad band near 50 kDa was found to consist of the C2 and a new N3 fragment starting at position 123 and ending near 648, which would fit well with another R/Y cut at 645/646 (Fig. 6B and Table S2). Therefore, with the exception of C2, the C-termini of all fragments end near one of the three Edman degradation-derived vesicle enzyme specificities Q/N, G/L, and R/Y (Fig. 6B).

A new, well resolved band appeared at about 62 kDa (Fig. 6A, lane 3a), possibly giving rise to C3 after deglycosylation (Fig. 6A, lane 3b and Table S2). In addition, a strong band is visible between the N1 and N2 bands in lane 3b of Fig. 6A. Attempts to elute it were unsuccessful. However, its position in the gel suggests that it is derived from N1. As it is still larger than N2, it might be an intermediate product between N1 and N2. We emphasize that the uncertainty of the C-terminal ends (Fig. 6B and Table S2) reflects that of the mass spectrometry rather than that of the biological process. On the contrary, the cuts produced by the vesicle enzyme are few, precise, and well defined as obvious from the N-termini of the fragments determined by Edman degradation and the sharp and reproducible gel bands (Fig. 6A). The most important result of these experiments is that in the course of development portions of the insertion are successively removed without affecting the cohesion of the 22S particle. The significance of this developmental process is discussed below.

Discussion

This paper on structure and function of the toposome challenges the general view of the role and location of this molecule, also called major yolk protein or MYP. Despite evidence against a nutritional role (Kari and Rottmann, 1985; Scott et al., 1990), the belief in its exclusive localization in cytoplasmic vesicles, called "yolk granules", still persists (Wessel et al., 2000). So far the only reliable test for cell adhesion has been the ability of the active molecule to reverse the inhibition produced by Fabs specific for the contact site (Beug et al., 1970). The exclusive control of cell adhesion by the 22S toposome particle up to the pluteus stage is dramatically illustrated by the bare skeletons remaining after exposure of live *P. lividus* and *Arbacia lixula* plutei to homologous anti-toposome Fab in contrast to the intact live controls of the same plutei treated with eightfold higher concentrations of the heterologous Fab (Noll et al., 1981). These Fab preparations, specific for the 22S particle (Noll et al., 1985; Gratwohl et al., 1991), recognized the protein but not the carbohydrate part (C. Kellner and H. N., unpublished results).

The toposome binds calcium not iron

The binding of calcium is the key to toposome function (Herbst, 1900; Noll et al., 1979). The calcium concentration-dependent dissociation and association of sea urchin embryonic cells can now be explained on basis of the transferrin structure. Apparently, regulation of this reversibility is crucial for morphogenetic cell movements. The binding to the receptor on the cell surface has a decisive effect on the conformation of the glycoprotein complex, as it results in reversible binding of Ca^{2+}, which favors association of trimers to hexamers at the Ca^{2+} concentration of sea water, thus promoting cell adhesion by linking the trimers attached to two different cells (Noll et al., 1985). By contrast, after removal from the cell surface, dissociation of the isolated 22S hexamer into its 15S trimers requires EDTA (Noll et al., 1985; Cervello et al., 1992). For this reason biologically relevant Ca^{2+}-binding studies would have to be carried out with receptor-bound toposomes under in vivo conditions, a most difficult endeavor. According to a recent study (Hayley et al., 2006), binding of Ca^{2+} to purified toposomes changed the shape to a more compact form, as has been known for Fe^{3+}-binding in transferrins (Baker et al., 1987; Yajima et al., 2000). In both cases removal of the stabilizing metal renders the molecule susceptible to proteolytic attack (Williams et al., 1982b) and loss of activity (Legrand et al., 1984; Noll et al., 1985).

The known ability of melanotransferrin to bind Zn^{2+} might explain the animalization effect of this metal on sea urchin embryos (Mitsunaga and Yasumasu, 1984). Thus, if the stabilizing action of Zn^{2+}, by displacing Ca^{2+}, shifts the delicate equilibrium toward membrane association, the cell movements involved in morphogenesis would be inhibited. Whether the

binding of Zn^{2+}, Con A, or toposome-specific Fab also inhibit the post-fertilization processing required for development beyond the blastula stage remains to be determined.

The strongest evidence against iron binding by toposomes is the lack of the canonical ligand amino acids (Figs. 2 and 5). Moreover, one of several iron-less transferrins, the pig carbonic anhydrase inhibitor, has been shown to bind Ca^{2+} in nearly stoichiometric amounts (Wuebbens et al., 1997). In this case, the absence of arginine at the canonical position has been used as argument that the anion-binding arginine is indispensable for mediating the binding of iron. Without stoichiometric binding data, claims of iron-binding and -transport (Brooks and Wessel, 2002) are not convincing.

It is important to point out in this connection that most members of the transferrin family have functions other than iron transport. The growing number of mono- and non-iron transferrins (Rose et al., 1986; Moskaitis et al., 1990; Bartfeld and Law, 1990; Morabito and Moczydlowski, 1994; Kurama et al., 1995; McNagny et al., 1996; Fisher et al., 1997; Liang et al., 1997; Wuebbens et al., 1997; Yoshiga et al., 1999) further illustrates the functional diversity of this gene family and poses interesting evolutionary questions.

An important aspect in these experiments is that in dissociated cells DNA synthesis stops until cell contact is re-established (Sconzo et al., 1970; De Petrocellis and Vittorelli, 1975) or the dissociated cells are exposed to reaggregation-inhibiting Fab (Vittorelli et al., 1980). Apparently, the binding of Fabs to the contact sites mimics cell–cell contact and thus stimulates DNA synthesis in the same way that binding Fab to the receptors of epidermal growth factor or insulin mimics the action of those hormones (Kahn et al., 1978; Schreiber et al., 1981).

Sequence alignments identify contact site and confirm genus-dependent antigenic specificity

Early experiments established that mixtures of cells dissociated from different sea urchin genera sorted themselves out to form only homologous developmentally competent aggregates (McClay et al., 1977). Moreover, Fab fragments of antisera raised against membranes of *Lytechinus* inhibited the aggregation of *Lytechinus* but not of *Tripneustes* cells. Confirmation and extension of these experiments (Noll et al., 1979, 1981) resulted in the isolation of the active component from membranes (Noll et al., 1985). The inhibiting antibodies were specific for the 22S particle of a given species, and any cross-reactivity with heterologous toposomes diminished sharply with the evolutionary distance (Noll et al., 1981; Gratwohl et al., 1991).

The extreme degree of identity of the five sequences in Fig. 5 was used to localize the species- or genus-specific contact site to the limited areas of non-identity. The most divergent segments in the mature complex are between positions 390 and 480, others in the middle of the insertion (730–830) and the tail section (1200–1340). As the segment in the insertion is the target of the major post-fertilization proteolytic cut between residues 808 and 809 and of additional cuts N-terminal to this first cleavage site (Fig. 6B and Table S2), it is unlikely to be part of the contact site. Similarly, the tail section next to the cystine knot, presumably engaged in subunit interaction, seems equally unsuitable as the determinant of immunological specificity. This leaves as most attractive candidate for the genus-specific epitope the most variable residues 390–480, which form the external loop marked green in Fig. 4B. This makes sense because the region connecting the two lobes is particularly accessible to proteolytic enzymes (Baker et al., 1987). Moreover, binding the antibody to this site is expected to induce a less compact form and the release of Ca^{2+}, promoting the dissociation into two trimers and thus breaking the homophilic cell–cell interaction. By contrast, since binding to the cell surface receptor is not species-specific, the sequences containing the receptor-binding site are expected to be in regions of identity. Inasmuch as all Tfs bind to their cell surface receptors, these sequences are likely to be in portions of the toposome molecule shared with or homologous to the Tf family. Moreover, in the case of human serum-Tf, both lobes are required for binding to the receptor (Mason et al., 1997). As evident from Fig. 5, toposomes contain in both lobes regions of identity, which on the basis of their homology with the receptor-binding sequences of serum-Tf (Mason et al., 1997) could be responsible for the same function.

Proteolytic modification in second transferrin domain of toposomes is required for development beyond blastula

Since all sea urchin genera share a similar developmentally regulated post-fertilization proteolytic modification of their toposomes, it must be essential to embryogenesis. The cuts occur in the insertion and when analyzed by reducing SDS-PAGE, the patterns of the fragments vary somewhat, depending on the genus or species. The sequence comparison of five species in Fig. 5 shows a stretch of some 100 amino acids in the middle of the insertion (amino acids 730–830) displaying the expected species variability (Fig. 5), which is largely lost during proteolytic processing (Table S2). Yet, it is important to emphasize that these proteolytic modifications leave the hexameric structure of the particle and its adhesive activity intact, as evident from the fact that Fabs from serum raised against blastula membranes are equally active in promoting the dissociation of blastulae and plutei (Noll et al., 1981; Matranga et al., 1986). Similarly, years ago Giudice et al. (1969) reported that mixtures of dissociated cells from blastulae and plutei of the same species were able to form developmentally competent aggregates.

The post-fertilization processing occurs in the cytoplasmic vesicles, apparently as the result of a cathepsin B-like activity sensitive to thiol-protease inhibitors (Yokota and Kato, 1988; Mallya et al., 1992). Thus, the protease inhibitor benzamidine did not interfere with development from egg to blastula, but arrested further development, unless the inhibitor was washed out. In that case both processing and development to pluteus resumed (Mallya et al., 1992). It should be noted that proteolytic modification of the toposomes is restricted to the insertion in the C-lobe and the region C-terminal to the Cys-knot (Figs. 4A and 6B).

Some of the N-termini of the proteolytic fragments of toposomes of the Japanese sea urchin *H. pulcherrimus* (Yokota et al.,

2003) can be aligned with ours, but are not consistent with our time course. Earlier studies (Lee et al., 1989), which analyzed the fragments from protease treatment followed by paper chromatography, showed that the first cut produced two fragments of unequal size and that all further cuts produced two groups of peptides derived from each of the original fragments. Not knowing the amino acid sequence, they were unable to know which of the original two fragments were N- and C-terminal.

It is interesting that exposure of mesenchyme blastulae to low concentrations of either Con A (Lallier, 1972, 1980; Matsumoto et al., 1984) or anti-toposome Fab (Noll et al., 1981) arrests development, allowing the blastulae to survive for long periods. The same is true for blastulae kept in the presence of the protease inhibitors leupeptin and antipain (Mallya et al., 1992; H.N. and J.A., unpublished results). Development is similarly arrested if the Ca^{2+} concentration is reduced to a subcritical level that is insufficient to cause dissociation into single cells. All of these treatments appear to inhibit cell division, presumably because not enough contact sites are occupied to transduce the signal permitting transition of the cell cycle from G1 into S phase. If some of these signals were supplied by the post-fertilization modifications, their suppression by certain protease inhibitors would explain why development beyond the blastula stage is blocked. A similar function might be exerted by some of the larger pieces, 5–10 kDa, released in the vesicle during the shortening of the earlier fragments N1 and C1 (Fig. 6B) and lost during purification, as they are not bound to the 22S complex by cysteine bridges like their parent fragments. A signal function of these fragments would further emphasize the evolutionary significance of the insertion in the second domain, which is the major distinguishing feature of this unusual transferrin.

The observation that the proteolytic modifications occur without disrupting the hexamer was difficult to understand until the discovery that it occurred in the cysteine-less insertion of a member of the transferrin family. Now it is clear that the canonical S–S bonds, still present in the 22S particle isolated after these proteolytic cuts have occurred, stabilize the subunits. At the same time, this protease-dependent processing event would explain the importance of both the presence of this novel insertion in a Tf-like protein and its modification during development. In addition, these modifications are consistent with earlier results that 7 different monoclonal antibodies generated against butanol extract and specific for 22S particles stained parts and their combinations of the cell surface in thin sections of blastulae (Noll et al., 1985). Clearly, the strategy of post-translational proteolytic modification to produce variants in

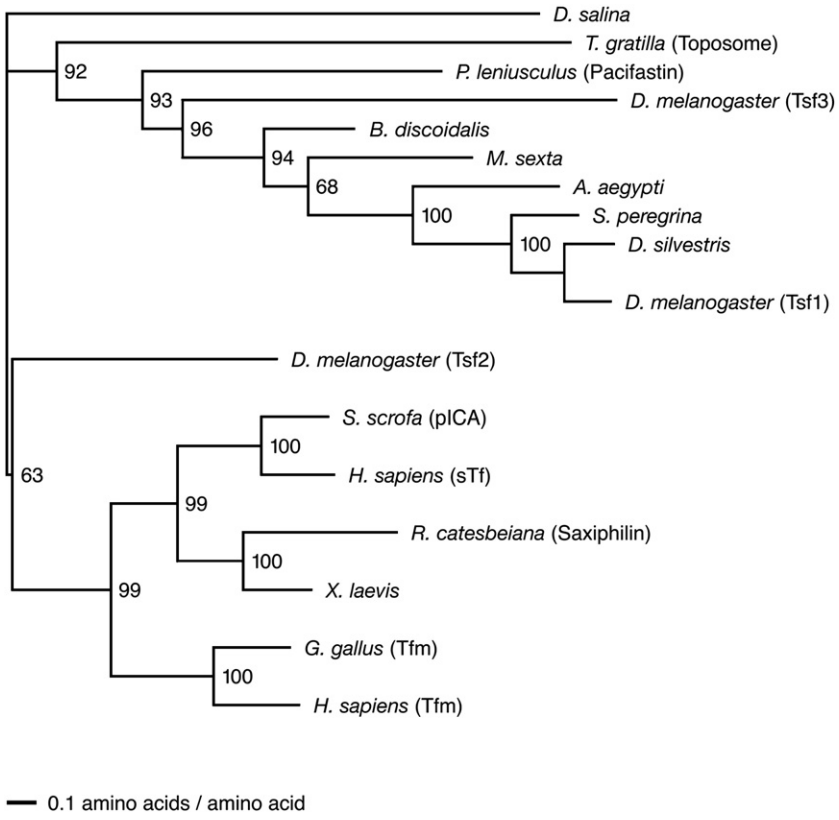

Fig. 7. Phylogenetic tree of the transferrin-like family including di-, mono-, and zeroferric representatives. The tree was derived from the Tf-like sequences of *D. salina* (Fisher et al., 1997), *P. leniusculus* (Pacifastin) (Liang et al., 1997), *D. melanogaster* (Tsf1–3; CG6186, CG10620, and CG3666), *B. discoidalis* (Jamroz et al., 1993), *M. sexta* (Bartfeld and Law, 1990), *A. aegypti* (Yoshiga et al., 1997), *S. peregrina* (Kurama et al., 1995), *D. silvestris* (Yoshiga et al., 1999), *S. scrofa* (pICA) (Wuebbens et al., 1997), *H. sapiens* (sTf) (Yoshiga et al., 1999), *R. catesbeiana* (Saxiphilin) (Morabito and Moczydlowski, 1994), *X. laevis* (Moskaitis et al., 1990), *G. gallus* (Tfm) (McNagny et al., 1996), and *H. sapiens* (Tfm) (Rose et al., 1986). Pacifastin and the product of the green alga *D. salina* have three internal repeats. The sequences were aligned with the aid of Clustal W (Higgins et al., 1996), and the tree was obtained with the quartet puzzling program Tree-Puzzle v5.0 (Strimmer and von Haeseler, 1996). The distances shown in the figure are based on maximum-likelihood estimation, and the numbers at the nodes are the bootstrap percentages for 1000 iterations (Strimmer and von Haeseler, 1996).

a cell surface protein to direct morphogenic cell movements makes sense and would be a logical new tool for generating epigenetic diversity.

The phylogenetic origin of toposomes is probably related to that of other iron-less transferrin-like molecules with multiple cation-activated allosteric functions

The evolutionary tree in Fig. 7 places the toposome of *T. gratilla* as representative of other sea urchin genera at a great distance from the Tfs of other organisms examined, with its node closest to the outgroup *D. salina* from the green algae. The position of the sea urchin proteins falls in a clade with Tfs of other invertebrates that, with the exception of the two *D. melanogaster* proteins Tsf2 and Tsf3, appear to fall in positions expected from the phylogeny of their respective species. From the positions of the three *D. melanogaster* transferrin-like proteins in the tree and the fact that there are three representatives in the *Drosophila* genome, it is apparent that there have been two subsequent duplications in the ancestral gene, which itself duplicated before these duplications. One duplication occurred after the divergence of the crayfish and before the divergence of the cockroach. The other duplication took place at about the time of the divergence of the invertebrates and vertebrates. A further duplication is evident in the vertebrates, which has produced the melanotransferrin and serum transferrin types.

The occurrence of both iron-binding and iron-less transferrin-like forms early in evolution makes it difficult to judge the evolutionary relationship, especially in consideration of the fact that the function of the iron-less and of most monoferric forms is still unclear. By contrast, the 22S glycoprotein complex from sea urchins is the first example of the iron-less transferrin-like proteins whose function as cell adhesion molecules has been exhaustively documented. However, because of the conserved cysteines and disulfide bonds in fixed positions, all forms, irrespective of the presence or absence of iron-binding ligands, would share the same basic structure which, being inherently stable, resists evolutionary change (Williams, 1982). The evolutionary relationship of transferrins from 51 species has recently been examined (Lambert et al., 2005).

In view of the unknown functions associated with these early iron-less and monoferric homologs, it is attractive to argue that the appearance of five particular amino acid residues in the exact positions required for the binding of iron is a later development. In this interpretation, the cation-activated allosteric transitions involving the two lobes are assumed to be in all homologs the key for their general function, of which iron transport in the higher vertebrates is a special case.

Acknowledgments

We thank Melchiorre Cervello for providing Northern blots of *P. lividus*, Salvo Feo for amplification of a cDNA fragment from *P. lividus*, Patrick Spielmann for DNA sequencing, Thomas Gutjahr for help in the search for protein sequences homologous to toposomes, Rolf Knoth for advice, and F. Ochsenbein for art work. This work has been supported by grants from the NIH (6M 38984), the American Cancer Society (RDP-36), The Roche Foundation and personal savings of H.N., who is a career professor of the American Cancer Society, the NIH-NCCR award RR-11823 (to R.A.), the Swiss National Science Foundation grants 31-26652.89, 31-40874.94, and 31-56817.99 (to M.N.), and by the Kanton Zürich, Switzerland.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2007.07.016.

References

Alemany, R., Vilá, M.R., Francí, C., Egea, G., Real, F.X., Thomson, T.M., 1993. Glycosyl phosphatidylinositol membrane anchoring of melanotransferrin (p97): apical compartmentalization in intestinal epithelial cells. J. Cell Sci. 104, 1155–1162.

Anderson, B.F., Baker, H.M., Dodson, E.J., Norris, G.E., Rumball, S.V., Waters, J.M., Baker, E.N., 1987. Structure of human lactoferrin at 3.2-Å resolution. Proc. Natl. Acad. Sci. U. S. A. 84, 1769–1773.

Armant, D.R., Carson, D.D., Decker, G.L., Welply, J.K., Lennarz, W.J., 1986. Characterization of yolk platelets isolated from developing embryos of *Arbacia punctulata*. Dev. Biol. 113, 342–355.

Baker, E.N., Lindley, P.F., 1992. New perspectives on the structure and function of transferrins. J. Inorg. Biochem. 47, 147–160.

Baker, E.N., Rumball, S.V., Anderson, B.F., 1987. Transferrins: insights into structure and function from studies on lactoferrin. Trends Biochem. Sci. 12, 350–353.

Bartfeld, N.S., Law, J.H., 1990. Isolation and molecular cloning of transferrin from the tobacco hornworm, *Manduca sexta*. J. Biol. Chem. 265, 21684–21691.

Bell, S.L., Xu, G., Forstner, J.F., 2001. Role of the cystine-knot motif at the C-terminus of rat mucin protein Muc2 in dimer formation and secretion. Biochem. J. 357, 203–209.

Beug, H., Gerisch, G., Kempff, S., Riedel, V., Cremer, G., 1970. Specific inhibition of cell contact formation in *Dictyostelium* by univalent antibodies. Exp. Cell Res. 63, 147–158.

Brooks, J.M., Wessel, G.M., 2002. The major yolk protein in sea urchins is a transferrin-like, iron binding protein. Dev. Biol. 245, 1–12.

Cervello, M., Di Ferro, D., D'Amelio, L., Zito, F., Matranga, V., 1992. Calcium-dependent self-aggregation of toposome, a sea urchin embryo cell adhesion molecule. Biol. Cell 74, 231–234.

De Petrocellis, B., Vittorelli, M.L., 1975. Role of cell interactions in development and differentiation of the sea urchin *Paracentrotus lividus*. Changes in the activity of some enzymes of DNA biosynthesis after cell dissociation. Exp. Cell Res. 94, 392–400.

Eng, J.K., McCormack, A.L., Yates III, J.R., 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989.

Fisher, M., Gokhman, I., Pick, U., Zamir, A., 1997. A structurally novel transferrin-like protein accumulates in the plasma membrane of the unicellular green alga *Dunaliella salina* grown in high salinities. J. Biol. Chem. 272, 1565–1570.

Giudice, G., Mutolo, V., Donatuti, G., Bosco, M., 1969. Reaggregation of mixtures of cells from different developmental stages of sea urchin embryos. Exp. Cell Res. 54, 279–281.

Gratwohl, E.K.-M., Kellenberger, E., Lorand, L., Noll, H., 1991. Storage, ultrastructural targeting and function of toposomes and hyalin in sea urchin embryogenesis. Mech. Dev. 33, 127–138.

Gum Jr., J.R., Hicks, J.W., Toribara, N.W., Rothe, E.-M., Lagace, R.E., Kim, Y.S., 1992. The human *MUC2* intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region. J. Biol. Chem. 267, 21375–21383.

Harrington, F.E., Easton, D.P., 1982. A putative precursor to the major yolk protein of the sea urchin. Dev. Biol. 94, 505–508.

Hayley, M., Perera, A., Robinson, J.J., 2006. Biochemical analysis of a Ca^{2+}-dependent membrane-membrane interaction mediated by the sea urchin yolk granule protein, toposome. Dev. Growth Differ. 48, 401–409.

Heifetz, A., Lennarz, W.J., 1979. Biosynthesis of *N*-glycosidically linked glycoproteins during gastrulation of sea urchin embryos. J. Biol. Chem. 254, 6119–6127.

Herbst, C., 1900. Über das Auseinandergehen von Furchungs- und Gewebezellen in kalkfreiem Medium. Arch. Entwicklungsmech. Org. 9, 424–463.

Higgins, D.G., Thompson, J.D., Gibson, T.J., 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. 266, 383–402.

Ii, I., Deguchi, K., Kawashima, S., Endo, S., Ueta, N., 1978. Water-soluble lipoproteins from yolk granules in sea urchin eggs: I. Isolation and general properties. J. Biochem. 84, 737–749.

Infante, A.A., Nemer, M., 1968. Heterogeneous ribonucleoprotein particles in the cytoplasm of sea urchin embryos. J. Mol. Biol. 32, 543–565.

Isaacs, N.W., 1995. Cystine knots. Curr. Opin. Struct. Biol. 5, 391–395.

Jamroz, R.C., Gasdaska, J.R., Bradfield, J.Y., Law, J.H., 1993. Transferrin in a cockroach: molecular cloning, characterization, and suppression by juvenile hormone. Proc. Natl. Acad. Sci. U. S. A. 90, 1320–1324.

Kahn, C.R., Baird, K.L., Jarrett, D.B., Flier, J.S., 1978. Direct demonstration that receptor crosslinking or aggregation is important in insulin action. Proc. Natl. Acad. Sci. U. S. A. 75, 4209–4213.

Kari, B.E., Rottmann, W.L., 1985. Analysis of changes in a yolk glycoprotein complex in the developing sea urchin embryo. Dev. Biol. 108, 18–25.

Katsumi, A., Tuley, E.A., Bodó, I., Sadler, J.E., 2000. Localization of disulfide bonds in the cystine knot domain of human von Willebrand factor. J. Biol. Chem. 275, 25585–25594.

Kirchhoff, C., Habben, I., Ivell, R., Krull, N., 1991. A major human epididymis-specific cDNA encodes a protein with sequence homology to extracellular proteinase inhibitors. Biol. Reprod. 45, 350–357.

Kurama, T., Kurata, S., Natori, S., 1995. Molecular characterization of an insect transferrin and its selective incorporation into eggs during oogenesis. Eur. J. Biochem. 228, 229–235.

Lallier, R., 1972. Effects of concanavalin A on the development of sea urchin eggs. Exp. Cell Res. 72, 157–163.

Lallier, R., 1980. Morphogenetic action of lectins on the development of sea urchin eggs. Biol. Cell. 39, 295–300.

Lambert, L.A., Perri, H., Meehan, T.J., 2005. Evolution of duplications in the transferrin family of proteins. Comp. Biochem. Physiol., Part B 140, 11–25.

Lee, G.F., Fanning, E.W., Small, M.P., Hille, M.B., 1989. Developmentally regulated proteolytic processing of a yolk glycoprotein complex in embryos of the sea urchin, *Strongylocentrotus purpuratus*. Cell Differ. Dev. 26, 5–17.

Lee, H., Yi, E.C., Wen, B., Reily, T.P., Pohl, L., Nelson, S., Aebersold, R., Goodlett, D.R., 2004. Optimization of reversed-phase microcapillary liquid chromatography for quantitative proteomics. J. Chromatogr., B 803, 101–110.

Legouis, R., Hardelin, J.-P., Levilliers, J., Claverie, J.-M., Compain, S., Wunderle, V., Millasseau, P., Le Paslier, D., Cohen, D., Caterina, D., Bougueleret, L., Delemarre-Van de Waal, H., Lutfalla, G., Weissenbach, J., Petit, C., 1991. The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. Cell 67, 423–435.

Legrand, D., Mazurier, J., Metz-Boutigue, M.-H., Jollès, J., Jollès, P., Montreuil, J., Spik, G., 1984. Characterization and localization of an iron-binding 18-kDa glycopeptide isolated from the N-terminal half of human lactotransferrin. Biochim. Biophys. Acta 787, 90–96.

Legrand, D., Mazurier, J., Montreuil, J., Spik, G., 1988. Structure and spatial conformation of the iron-binding sites of transferrins. Biochimie 70, 1185–1195.

Li, X.J., Pedrioli, P.G.A., Eng, J., Martin, D., Yi, E.C., Lee, H., Aebersold, R., 2004. A tool to visualize and evaluate data obtained by liquid chromatography-electrospray ionization-mass spectrometry. Anal. Chem. 76, 3856–3860.

Liang, Z., Sottrup-Jensen, L., Aspán, A., Hall, M., Söderhäll, K., 1997. Pacifastin, a novel 155-kDa heterodimeric proteinase inhibitor containing a unique transferrin chain. Proc. Natl. Acad. Sci. U. S. A. 94, 6682–6687.

Low, M.G., 1987. Biochemistry of the glycosyl-phosphatidylinositol membrane protein anchors. Biochem. J. 244, 1–13.

MacGillivray, R.T.A., Mendez, E., Sinha, S.K., Sutton, M.R., Lineback-Zins, J., Brew, K., 1982. The complete amino acid sequence of human serum transferrin. Proc. Natl. Acad. Sci. U. S. A. 79, 2504–2508.

MacGillivray, R.T.A., Mendez, E., Shewale, J.G., Sinha, S.K., Lineback-Zins, J., Brew, K., 1983. The primary structure of human serum transferrin. The structures of seven cyanogen bromide fragments and the assembly of the complete structure. J. Biol. Chem. 258, 3543–3553.

McClay, D.R., Chambers, A.F., Warren, R.H., 1977. Specificity of cell–cell interactions in sea urchin embryos. Appearance of new cell-surface determinants at gastrulation. Dev. Biol. 56, 343–355.

McNagny, K.M., Rossi, F., Smith, G., Graf, T., 1996. The eosinophil-specific cell surface antigen, EOS47, is a chicken homologue of the oncofetal antigen melanotransferrin. Blood 87, 1343–1352.

Malkin, L.I., Mangan, J., Gross, P.R., 1965. A crystalline protein of high molecular weight from cytoplasmic granules in sea urchin eggs and embryos. Dev. Biol. 12, 520–542.

Mallya, S.K., Partin, J.S., Valdizan, M.C., Lennarz, W.J., 1992. Proteolysis of the major yolk glycoproteins is regulated by acidification of the yolk platelets in sea urchin embryos. J. Cell Biol. 117, 1211–1221.

Marti, T., Rösselet, S.J., Titani, K., Walsh, K.A., 1987. Identification of disulfide-bridged substructures within human von Willebrand factor. Biochemistry 26, 8099–8109.

Martinerie, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L., Perbal, B., 1994. Structural analysis of the human *nov* proto-oncogene and expression in Wilms tumors. Oncogene 9, 2729–2732.

Mason, A.B., Tam, B.M., Woodworth, R.C., Oliver, R.W.A., Green, B.N., Lin, L.-N., Brandts, J.F., Savage, K.J., Lineback, J.A., MacGillivray, R.T.A., 1997. Receptor recognition sites reside in both lobes of human serum transferrin. Biochem. J. 326, 77–85.

Matranga, V., Kuwasaki, B., Noll, H., 1986. Functional characterization of toposomes from sea urchin blastula embryos by a morphogenetic cell aggregation assay. EMBO J. 5, 3125–3132.

Matsumoto, E., Tonegawa, Y., Ishihara, K., 1984. Studies on the localization and activities of Concanavalin-A-reactive glycoproteins on the cell surface of sea urchin embryos. J. Exp. Zool. 232, 157–165.

Mazurier, J., Metz-Boutigue, M.-H., Jollès, J., Spik, G., Montreuil, J., Jollès, P., 1983. Human lactotransferrin: molecular, functional and evolutionary comparisons with human serum transferrin and hen ovotransferrin. Experientia 39, 135–141.

Meitinger, T., Meindl, A., Bork, P., Rost, B., Sander, C., Haasemann, M., Murken, J., 1993. Molecular modelling of the Norrie disease protein predicts a cystine knot growth factor tertiary structure. Nat. Genet. 5, 376–380.

Metz-Boutigue, M.-H., Jollès, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montreuil, J., Jollès, P., 1984. Human lactotransferrin: amino acid sequence and structural comparisons with other transferrins. Eur. J. Biochem. 145, 659–665.

Mitsunaga, K., Yasumasu, I., 1984. Stage specific effects of Zn^{2+} on sea urchin embryogenesis. Dev. Growth Differ. 26, 317–327.

Morabito, M.A., Moczydlowski, E., 1994. Molecular cloning of bullfrog saxiphilin: a unique relative of the transferrin family that binds saxitoxin. Proc. Natl. Acad. Sci. U. S. A. 91, 2478–2482.

Moskaitis, J.E., Pastori, R.L., Schoenberg, D.R., 1990. The nucleotide sequence of *Xenopus laevis* transferrin mRNA. Nucleic Acids Res. 18, 6135.

Murray-Rust, J., McDonald, N.Q., Blundell, T.L., Hosang, M., Oefner, C., Winkler, F., Bradshaw, R.A., 1993. Topological similarities in TGF-β2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. Structure 1, 153–159.

Newman, R., Schneider, C., Sutherland, R., Vodinelich, L., Greaves, M., 1982. The transferrin receptor. Trends Biochem. Sci. 7, 397–400.

Noll, H., Matranga, V., Cascino, D., Vittorelli, L., 1979. Reconstitution of membranes and embryonic development in dissociated blastula cells of the sea urchin by reinsertion of aggregation-promoting membrane proteins extracted with butanol. Proc. Natl. Acad. Sci. U. S. A. 76, 288–292.

Noll, H., Matranga, V., Palma, P., Cutrono, F., Vittorelli, L., 1981. Species-specific dissociation into single cells of live sea urchin embryos by Fab

against membrane components of *Paracentrotus lividus* and *Arbacia lixula*. Dev. Biol. 87, 229–241.

Noll, H., Matranga, V., Cervello, M., Humphreys, T., Kuwasaki, B., Adelson, D., 1985. Characterization of toposomes from sea urchin blastula cells: a cell organelle mediating cell adhesion and expressing positional information. Proc. Natl. Acad. Sci. U. S. A. 82, 8062–8066.

Ozaki, H., 1980. Yolk proteins of the sand dollar *Dendraster excentricus*. Dev. Growth Differ. 22, 365–372.

Richardson, D.R., 2000. The role of the membrane-bound tumour antigen, melanotransferrin (p97), in iron uptake by the human malignant melanoma cell. Eur. J. Biochem. 267, 1290–1298.

Rose, T.M., Plowman, G.D., Teplow, D.B., Dreyer, W.J., Hellström, K.E., Brown, J.P., 1986. Primary structure of the human melanoma-associated antigen p97 (melanotransferrin) deduced from the mRNA sequence. Proc. Natl. Acad. Sci. U. S. A. 83, 1261–1265.

Rothberg, J.M., Jacobs, J.R., Goodman, C.S., Artavanis-Tsakonas, S., 1990. Slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains. Genes Dev. 4, 2169–2187.

Saheki, T., Ito, F., Hagiwara, H., Saito, Y., Kuroki, J., Tachibana, S., Hirose, S., 1992. Primary structure of the human elafin precursor preproelafin deduced from the nucleotide sequence of its gene and the presence of unique repetitive sequences in the prosegment. Biochem. Biophys. Res. Commun. 185, 240–245.

Schalkwijk, J., Wiedow, O., Hirose, S., 1999. The trappin gene family: proteins defined by an N-terminal glutaminase substrate domain and a C-terminal four-disulphide core. Biochem. J. 340, 569–577.

Schreiber, A.B., Lax, I., Yarden, Y., Eshhar, Z., Schlessinger, J., 1981. Monoclonal antibodies against receptor for epidermal growth factor induce early and delayed effects of epidermal growth factor. Proc. Natl. Acad. Sci. U. S. A. 78, 7535–7539.

Sconzo, G., Pirrone, A.M., Mutolo, V., Giudice, G., 1970. Synthesis of ribosomal RNA in disaggregated cells of sea urchin embryos. Biochim. Biophys. Acta 199, 441–446.

Scott, L.B., Lennarz, W.J., 1989. Structure of a major yolk glycoprotein and its processing pathway by limited proteolysis are conserved in echinoids. Dev. Biol. 132, 91–102.

Scott, L.B., Leahy, P.S., Decker, G.L., Lennarz, W.J., 1990. Loss of yolk platelets and yolk glycoproteins during larval development of the sea urchin embryo. Dev. Biol. 137, 368–377.

Shyu, A.-B., Raff, R.A., Blumenthal, T., 1986. Expression of the vitellogenin gene in female and male sea urchin. Proc. Natl. Acad. Sci. U. S. A. 83, 3865–3869.

Shyu, A.-B., Blumenthal, T., Raff, R.A., 1987. A single gene encoding vitellogenin in the sea urchin *Strongylocentrotus purpuratus*: sequence at the 5′ end. Nucleic Acids Res. 15, 10405–10417.

Stetler, G., Brewer, M.T., Thompson, R.C., 1986. Isolation and sequence of a human gene encoding a potent inhibitor of leukocyte proteases. Nucleic Acids Res. 14, 7883–7896.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Unuma, T., Okamoto, H., Konishi, K., Ohta, H., Mori, K., 2001. Cloning of cDNA encoding vitellogenin and its expression in red sea urchin, *Pseudocentrotus depressus*. Zool. Sci. 18, 559–565.

Vittorelli, M.L., Matranga, V., Feo, S., Giudice, G., Noll, H., 1980. Inverse effects on thymidine incorporation in dissociated blastula cells of the sea urchin *Paracentrotus lividus* induced by butanol treatment and Fab addition. Cell Differ. 9, 63–70.

Wessel, G.A., 1995. A protein of the sea urchin cortical granules is targeted to the fertilization envelope and contains an LDL-receptor-like motif. Dev. Biol. 167, 388–397.

Wessel, G.M., Zaydfudim, V., Hsu, Y.J., Laidlaw, M., Brooks, J.M., 2000. Direct molecular interaction of a conserved yolk granule protein in sea urchins. Dev. Growth Differ. 42, 507–517.

Williams, J., 1967. Yolk utilization. In: Weber, R. (Ed.), The Biochemistry of Animal Development, Vol. 2. Academic Press, New York, pp. 341–382.

Williams, J., 1974. The formation of iron-binding fragments of hen ovotransferrin by limited proteolysis. Biochem. J. 141, 745–752.

Williams, J., 1982. The evolution of transferrin. Trends Biochem. Sci. 7, 394–397.

Williams, J., Chasteen, N.D., Moreton, K., 1982a. The effect of salt concentration on the iron-binding properties of human transferrin. Biochem. J. 201, 527–532.

Williams, J., Elleman, T.C., Kingston, I.B., Wilkins, A.G., Kuhn, K.A., 1982b. The primary structure of hen ovotransferrin. Eur. J. Biochem. 122, 297–303.

Wolpert, L., Gustafson, T., 1961. Studies on the cellular basis of morphogenesis of the sea urchin embryo. Exp. Cell Res. 25, 374–382.

Wuebbens, M.W., Roush, E.D., Decastro, C.M., Fierke, C.A., 1997. Cloning, sequencing, and recombinant expression of the porcine inhibitor of carbonic anhydrase: a novel member of the transferrin family. Biochemistry 36, 4327–4336.

Yajima, H., Sakajiri, T., Kikuchi, T., Morita, M., Ishii, T., 2000. Molecular modeling of human serum transferrin for rationalizing the changes in its physicochemical properties induced by iron binding. Implication of the mechanism of binding to its receptor. J. Protein. Chem. 19, 215–223.

Yokota, Y., Kato, K.H., 1988. Degradation of yolk granules in sea urchin eggs and embryos. Cell Differ. 23, 191–199.

Yokota, Y., Unuma, T., Moriyama, A., Yamano, K., 2003. Cleavage site of a major yolk protein (MYP) determined by cDNA isolation and amino acid sequencing in sea urchin, *Hemicentrotus pulcherrimus*. Comp. Biochem. Physiol., Pt. B 135, 71–81.

Yoshiga, T., Hernandez, V.P., Fallon, A.M., Law, J.H., 1997. Mosquito transferrin, an acute-phase protein that is up-regulated upon infection. Proc. Natl. Acad. Sci. U. S. A. 94, 12337–12342.

Yoshiga, T., Georgieva, T., Dunkov, B.C., Harizanova, N., Ralchev, K., Law, J.H., 1999. *Drosophila melanogaster* transferrin. Cloning, deduced protein sequence, expression during the life cycle, gene localization and up-regulation on bacterial infection. Eur. J. Biochem. 260, 414–420.

Table S1

Analysis of mature and proteolytically modified toposomes by mass spectrometry

			T		N1		N2		C1		C2	
1	ICVPPVLNNEDQVVQGPKPVETPDQVR	72- 98										
2	VGLYPAPEQELR	111- 122	9613	0.98	16247	0.98	2597	0.99	2885	0.99	432	0.99
3	YPVNPNVIR	123- 131	12928	0.96	18985	0.97	15123	0.98	6204	0.96	4617	0.99
4	FCVSSTCQMTK	132- 142	3425	0.99	3733	0.99	3646	0.98	1190	0.99	na	
5	MVSEFTFNPNMAPR	146- 159	6088	1.00	8192	1.00	7204	1.00	2283	1.00	1658	1.00
6	CIQADSQEQCMFWIEQGWADIMTAR	164- 188	116	1.00	84	0.98	69	1.00	na		na	
7	EEQVYVANTTFNLKPIAYETTINNDLPETLK	189- 219	7721	1.00	13044	1.00	14779	1.00	6133	1.00	1796	1.00
8	HYQNVTFALK	220- 229	15338	0.95	24831	0.97	20816	0.94	7202	0.95	5761	0.98
9	LINPNTFSELR	233- 243	21575	0.99	33686	0.99	26733	0.99	11228	0.98	5007	0.99
10	TTCHAGIDMPASFADPVCNLIK	246- 267	887	1.00	1160	1.00	1758	1.00	778	1.00	792	1.00
11	EGVIPVTGNYIESFADFVQESCLPGVLNK	268- 296	1233	1.00	1569	1.00	1325	1.00	823	1.00	744	0.96
12	TLISLCEDR	307- 315	8002	0.99	9765	0.99	8733	0.99	3365	1.00	2775	1.00
13	QAEYSGIK	316- 323										
14	GQVTFVDQK	334- 342										
15	IMNDENVR	347- 354										
16	DNYMVVCR	355- 362	7878	0.99	8476	0.99	7059	0.99	1249	0.94	ni	
17	EIFDDVTCHVGHTARPTIFINR	371- 392	1470	1.00	519	1.00	370	0.99	ni		2955	0.99
18	NNTQQQER	393- 400										
19	FNLFDSSVYTCDK	422- 434	2856	1.00	3533	1.00	2920	1.00	1168	1.00	685	1.00
20	NLIFLDESNTLEILDDAK	445- 462	14741	1.00	13395	1.00	9688	1.00	4222	1.00	2370	1.00
21	VFAGEVYAAYNTCSQLVPKPR	463- 483	4216	1.00	7803	1.00	7334	1.00	3199	1.00	1579	1.00
22	ICVTNVTEYEACR	486- 498	3706	1.00	4498	1.00	3768	1.00	1624	1.00	1195	1.00
23	GIAENIPQVK	502- 511	12585	0.98	16581	0.53	14835	0.99	ni		na	
24	NVAWGCVLANSSMECMK	512- 528	ni		1688	0.99	1974	0.30	880	1.00	606	1.00
25	AVHDNTADLFK	529- 539	10803	0.99	14842	0.95	15656	0.96	5192	0.99	3152	0.97
26	ANPQETFIAGK	540- 550	13158	0.99	21315	0.99	19208	0.99	6365	0.97	na	
27	EFLLDPLMSVHR	551- 562	1806	1.00	2750	1.00	2623	1.00	699	1.00	760	1.00
28	NDSVTLNHTYTR	563- 574	1058	0.91	5054	0.96	4515	0.96	1802	0.74	na	
29	SSLNQFPDILNVPEGQPK	582- 599	10020	1.00	14537	1.00	14703	1.00	1125	0.99	596	1.00
30	NFSAFANPIGYLLANGTIPR	616- 635	2960	0.99	4010	1.00	3072	1.00	ni		226	0.99
31	IGSVFESVNR	636- 645	21438	0.98	23175	0.99	20609	0.99	ni		747	1.00
32	YFQSTCIPEIEPER	646- 659	7952	1.00	10262	1.00	9650	1.00	596	1.00	334	0.99
33	FDSDLLLGR	662- 670	14259	1.00	21250	0.99	17257	1.00	1492	1.00	564	1.00
34	ELNWGFSTLNMYNFTGQEWLLWNTPATWNFLTYNR	671- 705										
35	VSNGLDIK	707- 714										
36	QNLTSHIFNQNLTSANVELLDDLVGVDGLSDLIK	722- 755	908	1.00	2319	0.98	ni		na		na	
37	GLQDSISPEGK	756- 766	164	0.99	332	0.99	165	0.99	141	0.99	na	
38	LSNSFPNFEGVR	776- 787	16035	1.00	8859	1.00	771	1.00	543	0.99	201	0.98
39	IQNQDTPFADYIQGK	807- 821	4244	1.00	ni		321	0.93	4906	1.00	1796	1.00
40	FGGELMVDIFSK	822- 833	4672	1.00	88	1.00	248	1.00	13274	1.00	4556	1.00
41	IATLEEIISHVK	842- 853	2601	1.00	128	1.00	302	1.00	27335	1.00	13586	1.00
42	SIPYLTDFK	854- 862	627	0.66	na		na		1468	0.72	343	0.98
43	DEEITTVIK	863- 871	151	0.92	na		ni		73	0.93	102	0.79
44	HPAIMSYVEIYFPR	872- 885										
45	LAQTFVEPFDNAELR	886- 900	20604	1.00	689	0.57	1098	0.99	49161	0.99	24002	1.00
46	YTNPLWLSPR	907- 916	15019	1.00	679	1.00	852	0.99	37917	1.00	15229	1.00
47	IDTYLDVIK	917- 925	5733	0.99	100	0.79	258	0.78	16279	1.00	4659	1.00
48	HQNEITQTCNSNLPLK	927- 942	5214	0.67	96	0.66	176	0.97	14793	0.55	6027	1.00
49	FNGYEGSLR	943- 951	9906	0.99	802	0.94	997	0.97	24769	0.99	2627	0.97
50	SGAADLAFFDEQTLR	955- 969	13004	1.00	448	1.00	972	1.00	27989	1.00	9543	1.00
51	SGAADLAFFDEQTLRDQDLLSR	955- 976	1308	1.00	75	ni	313	1.00	19423	1.00	15159	1.00
52	VGFTYNDLR	977- 985	16820	0.97	606	0.97	961	0.99	37526	0.99	18403	0.99
53	LLCPNGQVVEIDASLDVAK	986-1004	734	1.00	na		na		6746	1.00	1504	1.00
54	VCNFGEVMNPVLLTSYNTSGSLR	1005-1027	971	1.00	ni		ni		3977	1.00	1253	1.00
55	ALMIAHQSVALPALFGEGTVFGK	1033-1055	13	0.93	35	0.87	na		1002	0.98	677	0.98
56	LLPIAPLNQSYQAFLGPKPLR	1060-1080	27	0.98	na		na		833	1.00	444	1.00
57	ASSYDWFK	1088-1095	5737	0.99	236	0.99	218	0.97	11569	0.98	3781	1.00
58	DQPGICYGETYTNIVK	1096-1111	3856	1.00	182	0.93	ni		9108	1.00	3058	1.00
59	NETCQAVVK	1114-1122										
60	DVTCVGTPR	1123-1131										
61	MAEFQCDNGYGYLKPVITAIACECMPCEELIEYNASFTQDQMWK	1161-1204										
62	NVSNQYHLTGEQDVYR	1205-1220	13189	0.99	280	0.99	395	0.99	33232	0.99	14481	0.99
63	QIPIWGNNSFFYNHSLNK	1221-1238	ni		na		na		2994	0.99	1528	0.99
64	NFELGNHSVLVEHVQTVVVENPVGVISQLNTEVDPEDR	1239-1276										
65	VLMDTANITK	1277-1286	111	0.99	ni		ni		355	1.00	na	
66	TCESVWTGQSWLPER	1287-1301	3792	1.00	na		196	0.74	14559	1.00	339	1.00
67	TTGSCVVPEYGR	1307-1318	930	0.99	na		na		767	1.00	na	

The measured intensities of the tryptic peptides of the deglycosylated toposome bands from 16 h-old unhatched blastula embryos (Fig. 6A) are indicated for mature toposomes T and their processed products (N1, C1, N2, C2) obtained by proteolytic cleavage in the Cys-less insertion of the second transferrin domain. All tryptic peptides of mature *T. gratilla* toposomes larger than heptapeptides are listed in order from N- to C-terminus on the left with positions of their N- and C-termini indicated to their right. Note that trypsin cleaves after R and K unless these are followed by P. Peptides shorter than octapeptides could not be detected (cf. Materials and methods) except when detected as a partial tryptic digestion product (peptide 51). In mature toposomes T, all peptides (except two) longer than 9 and shorter than 35 amino acids have been identified with high confidence (confidence levels are indicated behind each measured intensity value, except for peptides not chosen for second MS analysis, na, or peptides chosen for second MS analysis but not identified with a tryptic peptide of the toposome, ni). The first tryptic peptide listed, extending between residues 72 and 98, is not detectable because the N-terminus of mature toposomes is located at residue 87 (cf. Table S2). Peptides whose intensities are marked green are considered to be present, as are all peptides of mature toposomes T after position 87 because these are not yet cleaved internally. The intensities of these tryptic peptides served as reference for those of the processed toposomes N1, C1, N2, and C2. Accordingly, intensities of peptides considered to be present are marked green in processed toposomes if they are not significantly lower than those of mature toposomes or if they are flanked by intensities of tryptic peptides that are clearly present. Intensities of peptides considered to be absent are marked red in processed toposomes if they are much lower (at least 2- to 3-fold) than those of mature toposomes or if they are flanked by intensities of tryptic peptides that are clearly absent. Intensities of peptides are marked yellow if the peptides are considered to be present in a significant fraction, but not all, of the processed toposomes on the basis of their measured intensities. Such peptides are found at the N- or C-terminal ends of the processed toposomes and permit an estimate of the limits within which the N- and C-termini are located (cf. Table S2).

Table S2
Summary of toposome bands analyzed by PAGE, Edman degradation, and mass spectrometry

Toposome bands	T	N1	C1	N2	C2	N3	putative C3
Before PNGase F [1]	170	111	80	105/101	71	n.d.[2]	n.d.[2]
After PNGase F [1]	131	75	59	68	53	55	48
N-terminus (Edman) [3]	87	87	809	113	809	123	n.d.[2]
N-terminus (MS/MS) [4]	< 111 [5]	< 111 [5]	777-821 [9]	112-122	777-821 [9]	n.d.[2]	n.d.[2]
C-terminus (MS/MS) [4]	>1318 [5]	767-808 [8]	1302-1344 [10]	671-786	1239-1285	n.d.[2]	n.d.[2]
Length	1258 [6]	700 ± 20	514 ± 22	640 [12]	465 [14]	526 [16]	440 [18]
MW [7]	144	80	59	73 [11]	53 [13]	60 [15]	48 [17]

[1] MW (kDa) determined by calibration with markers shown on gel in Fig. 6A. [2] Could not be determined. [3] N-termini determined by Edman degradation of PNGase F-treated protein bands of toposomes (cf. Materials and methods) from 16 h-old embryos at unhatched blastula stage and from 44 h-old embryos at early prism stage. [4] Range of location of N- and C-termini determined by mass spectrometry the results of which are summarized in Table S1. While Edman degradation determines the most abundant N-terminus, MS/MS spectrometry indicates the range within which the N- or C-terminus is located. [5] Tryptic fragments between amino acid positions 87 and 111 and between 1318 and the C-terminus at 1344 are shorter than heptapeptides and not detectable by our MS/MS analysis (cf. Materials and methods). Note that trypsin cleaves after K and R unless these are followed by P. [6] We assume that the non-dectable tryptic fragments at the C-terminal end of T, which are shorter than hexapeptides, are present. [7] MW (kDa) calculated from average length of toposome and toposome fragments. [8] A significant fraction of C-termini of N1 are located between amino acid positions 767 and 786 (cf. Table S1), while many may extend to position 808, the N-terminus of C1. [9] A significant fraction of N-termini of C1 and C2 are located between amino acid positions 777 and 806 (cf. Table S1). [10] A significant fraction of C-termini of C1 are located between amino acid positions 1302 and 1317 (cf. Table S1). [11] Calculated from N1 whose MW is 7 kDa larger. [12] The length was computed from the MW of N2. As evident from the gel in Fig. 6A, the lengths of most fragments varies by less than 2 kDa or about 20 amino acids. [13] Calculated from C1 whose MW is 6 kDa larger. [14] The length was computed from the MW of C2. As evident from the gel in Fig. 6A, the lengths of most fragments varies by less than 2 kDa or about 20 amino acids. [15] Calculated from N2 whose MW is 13 kDa larger. [16] The length was computed from the MW of N3. As evident from the gel in Fig. 6A, the lengths of most fragments varies by less than 3 kDa or about 25 amino acids. [17] Calculated from C2 whose MW is 5 kDa larger. [18] The N-terminus is assumed at position 809. The length was computed from the MW of the putative C3. As evident from the gel in Fig. 6A, the lengths of most fragments varies by less than 3 kDa or about 25 amino acids.