FOR THE RECORD

Generation of ligand–receptor alliances by "SEA" module-mediated cleavage of membrane-associated mucin proteins

DANIEL H. WRESCHNER,[1] MICHAEL A. MCGUCKIN,[2] STEFANIE J. WILLIAMS,[2] AMOS BARUCH,[1] MERAV YOELI,[1] RAVIT ZIV,[1] LIRON OKUN,[1] JOSEPH ZARETSKY,[1] NECHAMA SMORODINSKY,[1] IAFA KEYDAR,[1] PAVLOS NEOPHYTOU,[3] MARTIN STACEY,[4] HIS-HSIEN LIN,[4] AND SIAMON GORDON[4]

[1]Department of Cell Research and Immunology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel
[2]Mater Medical Research Institute, Mater Misericordiae Hospitals, South Brisbane Q4101, Australia
[3]Mendel Center for Biomedical Sciences, Nicosia, Cyprus
[4]Sir William Dunn School of Pathology, Oxford University, Oxford, UK

(RECEIVED May 2, 2001; FINAL REVISION October 22, 2001; ACCEPTED December 3, 2001)

Abstract

A mechanism is described whereby one and the same gene can encode both a receptor protein as well as its specific ligand. Generation of this receptor–ligand partnership is effected by proteolytic cleavage within a specific module located in a membrane resident protein. It is postulated here that the "SEA" module, found in a number of heavily O-linked glycosylated membrane-associated proteins, serves as a site for proteolytic cleavage. The subunits generated by proteolytic cleavage of the SEA module reassociate, and can subsequently elicit a signaling cascade. We hypothesize that all membrane resident proteins containing such a "SEA" module will undergo cleavage, thereby generating a receptor–ligand alliance. This requires that the protein subunits resulting from the proteolytic cleavage reassociate with each other in a highly specific fashion. The same SEA module that serves as the site for proteolytic cleavage, probably also contains the binding sites for reassociation of the resultant two subunits. More than one type of module can function as a site for proteolytic cleavage; this can occur not only in one-pass membrane proteins but also in 7-transmembrane proteins and other membrane-associated proteins. The proposal presented here is likely to have significant practical consequences. It could well lead to the rational design and identification of molecules that, by binding to one of the cleaved partners, will act either as agonists or antagonists, alter signal transduction and, hence, cellular behavior.

Keywords: SEA module; mucins; receptors; ligands; signaling; proteolytic cleavage

Interactions between ligands and their membrane receptors affect all aspects of cell behavior, be they related to tissue patterning, cell differentiation, cell growth, or cell death. The engagement of a cell surface receptor by its specific ligand(s) initiates a signaling cascade that ultimately culminates in changed patterns of gene expression, thereby altering cellular characteristics. Activation of the membrane-located receptor can be induced by a soluble secreted ligand, a cell–surface-associated ligand resident on either the same cell or a neighboring cell or by an extracellular matrix-embedded ligand. Whatever the location of the ligand, in almost all cases, the gene coding for the ligand is distinct

Reprint requests to: Daniel H. Wreschner, Department of Cell Research and Immunology, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel; e-mail: danielhw@post.tau.ac.il; fax: 972-3-6422046.

Article and publication are at http://www.genesdev.org/cgi/doi/10.1110/ps.16502.

from that coding for the receptor. There are two notable exceptions to this "canon." The first is the family of proteolytically activated receptors (PAR; Cupit et al. 1999), the founding member being the thrombin/thrombin-receptor couple (Coughlin 1999). In this set, thrombin proteolytic cleavage of its receptor exposes a new N-terminus that represents an unmasked "tethered" ligand. This then activates the receptor by intramolecular interactions. The second is illustrated by the *MUC*1 gene, which, by a process of alternative splicing, may generate both a membrane-tethered receptor protein and a soluble secreted ligand (Wreschner et al. 1990; Baruch et al. 1999).

We propose here an additional mechanism whereby one and the same gene can encode both a receptor protein as well as its specific ligand. Generation of this receptor–ligand partnership is effected by proteolytic cleavage within a specific module located in a membrane resident protein. The proteolytically generated subunits reassociate, and can subsequently elicit a signaling cascade. It appears likely that all membrane resident proteins containing such a module may undergo cleavage, thereby generating a receptor–ligand alliance. This hypothesis requires that the protein subunits resulting from the proteolytic cleavage reassociate with each other in a highly specific fashion. Furthermore, although not essential for our proposal, we will demonstrate that it is likely that the same module that serves as the site for proteolytic cleavage also contains the binding sites for reassociation of the resultant two subunits. In addition, we will show that more than one type of module can function as a site for proteolytic cleavage, and that this hypothesis relates not only to one-pass membrane proteins but also to 7-transmembrane proteins and other membrane-associated proteins.

We have derived this concept using as a paradigm the transmembrane protein, MUC1. We then relate to the generality of the phenomenon.

The *MUC*1 gene, which is overexpressed in breast cancer cells, can generate by alternative splicing at least three isoforms (Fig. 1A), two are membrane bound MUC1/TM and MUC1/Y (Gendler et al. 1990; Lan et al. 1990; Ligtenberg et al. 1990; Zrihan-Licht et al. 1994b; Baruch et al. 1999) and one is secreted, MUC1/SEC (Smorodinsky et al. 1996). We have previously demonstrated that MUC1/SEC interacts with the membrane-tethered MUC1/Y protein and induces phosphorylation of its cytoplasmic domain (Baruch et al. 1999).

The MUC1/TM isoform is a type 1 transmembrane protein that during its biosynthesis is modified to a large extent, and a considerable number of O-linked sugar moieties confer mucin-like characteristics on the mature protein (hence, the name MUC1). Shortly after its synthesis, and at any rate within the cytoplasm, the MUC1 protein undergoes proteolytic cleavage in its extracellular domain (Ligtenberg et al. 1992; Parry et al. 2001). The two resulting cleavage products reassociate in a specific manner and present at the cell surface as a heterodimer (Ligtenberg et al. 1992; Fig. 1A). Cleavage occurs at a site that is approximately 65 amino acids N-terminal to the transmembrane domain-deletion of a 20 amino acid stretch (45 to 65 amino acids N-terminal to the transmembrane domain) abolishes all cleavage (Ligtenberg et al. 1992). Clearly, this region must contain a proteolytic cleavage signal, and it is reasonable to assume that this site, if of functional import, will be conserved across species expressing MUC1 orthologs (Spicer et al. 1995).

Inspection of the proposed region of cleavage (45 to 65 amino acids N-terminal to the transmembrane domain), reveals high across-species conservation of the sequence I^{352}KFRPGSVVV (Fig. 2; the initial "I" residue corresponds to amino acid number 352, amino acid numbering throughout the text as in Wreschner et al. 1990), indicating that cleavage takes place within this sequence. Indeed, the exact cleavage site has been recently located just downstream to the glycine residue present in GSVVV (Parry et al. 2001). Intriguingly, this proposed cleavage site resides within a previously identified module (Bork and Patthy 1995) found in a number of highly O-linked glycosylated proteins that are invariably linked in one way or another to the cell membrane. This module, designated the "SEA" module, is of unknown function.

As the segment of the MUC1 molecule protruding from the cell membrane and generated by cleavage must necessarily bind the larger N-terminal fragment (to generate the heterodimer), we investigated, using the gel-overlay assay (Mehlman and Burgess 1990; Baruch et al. 1999), the minimal length of the membrane proximal fragment required to support such binding. A recombinant bacterial protein extending from the conserved G^{357}SVVV sequence to the transmembrane domain ($GSVVV_{down}$ = GSVVVQLTLA FREGTINVHDVETQFNQYKTEAASRYNLTISDVSVSD VPFPFSAQSGAGV) specifically associates with the large N-terminal fragment. No binding is observed when 12 amino acids are deleted from the $GSVVV_{down}$ N-terminus (Baruch et al. 1999). The across-species conservation of the GSVVV sequence and the minimal region required for a subunit association tally well with the identification of the cleavage within the GSVVV sequence (Parry et al. 2001).

An additional sequence, Y^{325}YQEL, is also highly conserved in all MUC1 orthologs, and is located 32 amino acids N-terminal to the GSVVV cleavage site (Fig. 2). A tblastn search was performed using a query sequence spanning both the YYQEL and GSVVV sites. This revealed one-pass membrane proteins in addition to several membrane-associated proteins, some of which are presented in Figure 3. Additional SEA module-containing proteins are not shown because of space constraints, yet they all show the major features of the SEA module as designated below. Remarkable similarity to the query sequence was observed within the YQEL and GSVVV sequences. As with MUC1, the

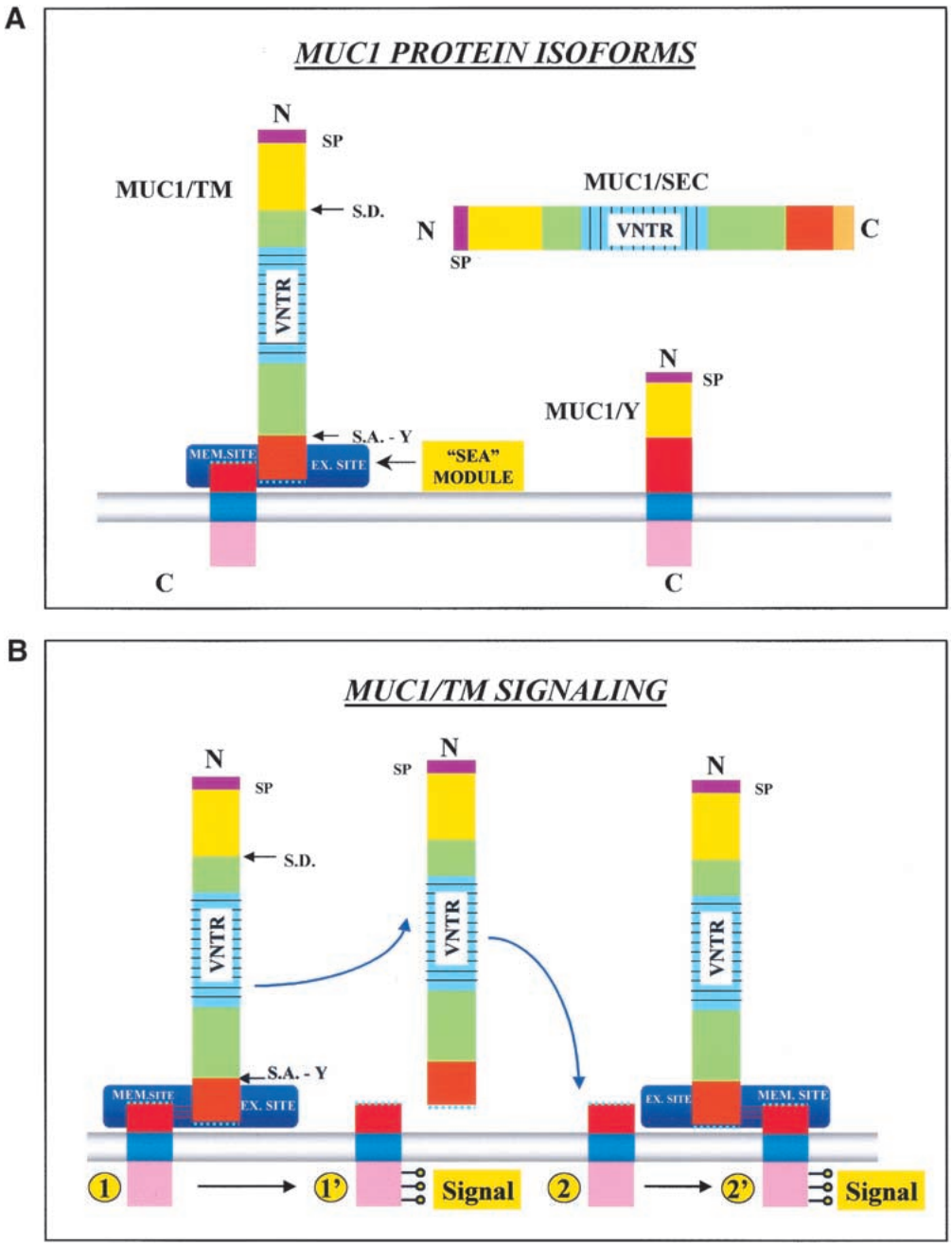

Fig. 1. MUC1 isoforms and their interactions. (*A*) MUC1 isoforms. N and C refer to the amino and carboxyl termini of the proteins. Regions common to the various MUC1 isoforms have the same colours and shadings. SP and VNTR refer to the signal peptide and variable number of tandem repeats, respectively. The splice donor and acceptor sites used to generate the MUC1/Y isoform are indicated by S.D. and S.A., respectively. The membrane proximal binding site is indicated by "MEM. Site," whereas the binding site of the large extracellular domain is designated "EX site." Both are contained within the SEA module. (*B*) Signaling initiated by the release and reassociation of the large extracellular domain.

MUC3, MUC12 mucin-like membrane-associated proteins (Williams et al. 1999a, 1999b), and *Caenorhabditis elegans* proteins identified in this search, are type 1 membrane proteins and carry a YQEL- and GSVVV-containing module within their extracellular domains. The proteins, perlecan (Iozzo et al. 1994) and interphotoreceptor proteoglycan 200 (Kuehn and Hageman 1999), identified in this search, are also membrane associated and are heavily decorated with O-linked glycans. The *C. elegans* type 1 membrane protein is of particular interest as it shows a remarkable similarity to

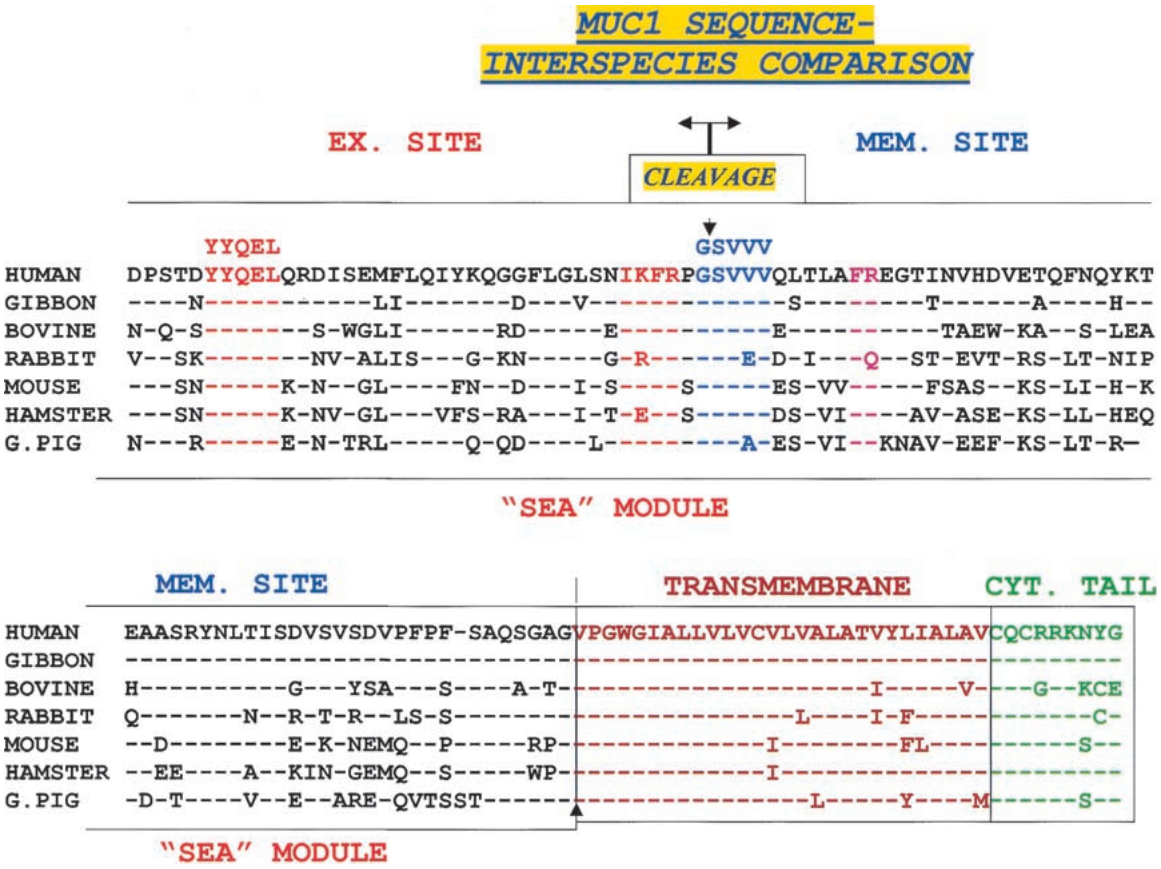

Fig. 2. Across species comparison of the MUC1/TM protein. The sequences were taken from Spicer et al. (1995). The conserved YQEL and GSVVV sites, also noted and discussed in Spicer et al. (1995), are highlighted and the site of cleavage is indicated.

the MUC1 protein. This relates not only to the actual amino acid sequence of the YQEL- and GSVVV-containing module, but also to the location of the module. In both cases it resides just N-terminal to the transmembrane domain (Fig. 4A). The similarity stretches out past the GSVVV sequence and extends into the transmembrane and cytoplasmic domains (Fig. 4B). Furthermore, as for MUC1, the *C. elegans* protein is well endowed with tyrosine residues in its cytoplasmic domain, and harbors in its extracellular domain a considerable number of strings of two or more serine and threonine residues.

Investigation of sequences just N-terminal to the proposed MUC1 GSVVV cleavage site (Fig. 3) among the various proteins reveals that despite the lack of discernible conserved primary amino acid sequence, there is considerable conservation of a 4–8 amino acid unit that comprises interdigitating hydrophilic (K, R, E, and D) and hydrophobic amino acids. It is likely that proteolytic cleavage requires a specific sequence, which includes the conserved GSVVV site (or variations thereof) as well as the four to eight amino acids N-terminal to this site.

As described above, the module portrayed here has been previously identified and designated the "SEA" module. No definite function had been ascribed to the SEA module (Bork and Patthy 1995). Proteins containing this module are consistently highly O-linked glycosylated, adhesive proteins that are associated in one way or another with the cell membrane (Bork and Patthy 1995). These proteins include single-pass transmembrane proteins, GPI-linked membrane proteins (the 63-kD GPI anchored sea urchin sperm protein), and the protein enterokinase, which harbors a hydrophobic anchor sequence near its N-terminus (for a more extensive list of SEA module containing proteins, the reader is referred to the SMART site on the Web, http://smart.embl-heidelberg).

As the SEA module-carrying proteins contain a considerable number of O-glycosidic-linked carbohydrates, it has been proposed that this common module might function in binding to neighboring carbohydrate residues (Bork and Patthy 1995).

The SEA module in the MUC1/TM protein functions as the site for proteolytic cleavage (Ligtenberg et al. 1992; Parry et al. 2001). Intriguingly, an additional type 1 transmembrane mucin-like protein originally designated as 114/A10 (Dougherty et al. 1989), and recently renamed MUC13, also contains a SEA module. It is also proteolytically

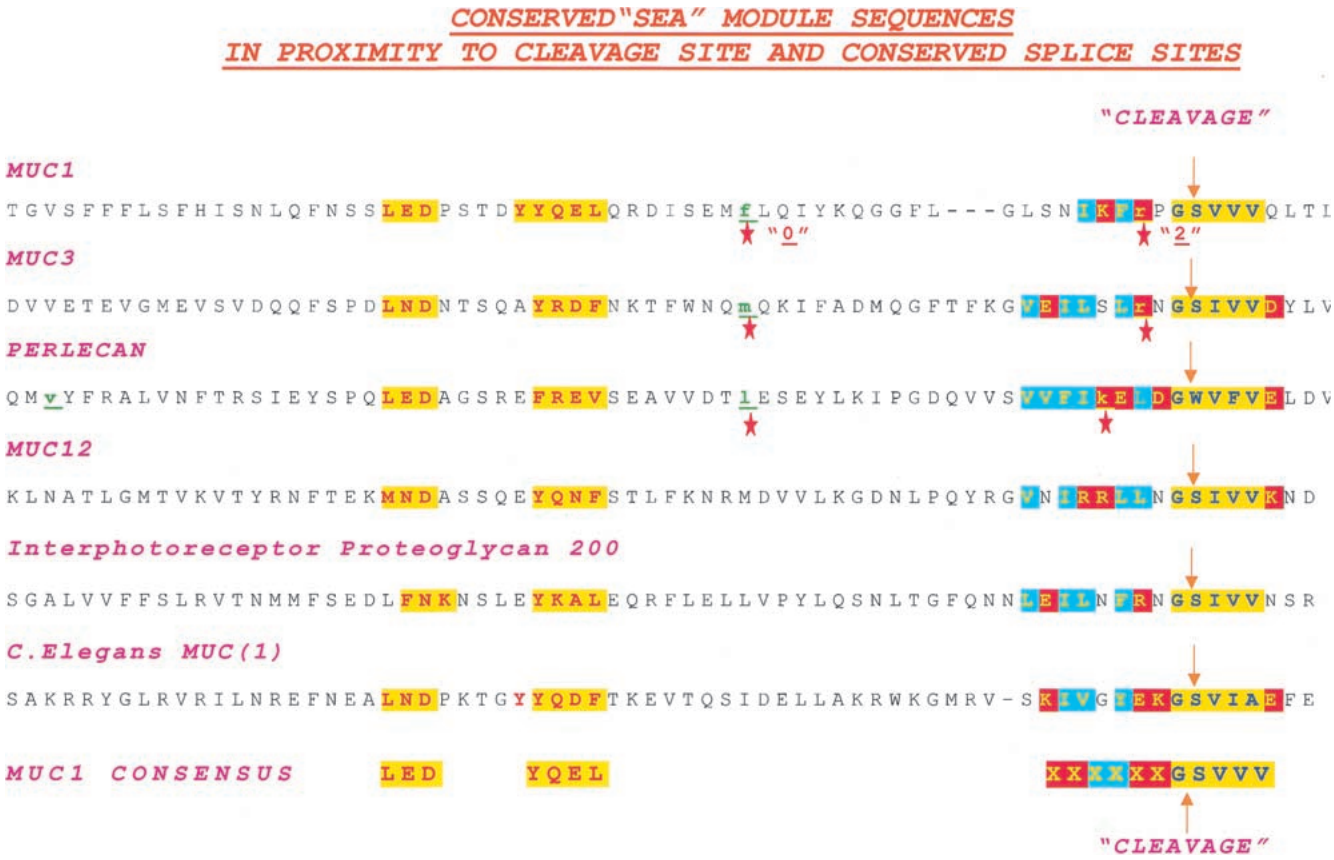

Fig. 3. Conserved SEA module sequences in other proteins upstream of the GSVVV site. The MUC1 (P15941, accession numbers in brackets), MUC3 (AF143373), MUC12 (AF147790), Interphotoreceptor proteoglycan 200 (AF173155), *C. elegans* MUC (CAB03861) are aligned for maximal homology. The conserved LED, YQEL, and GSVVV sites are highlighted and the interdigitating hydrophilic (yellow on red background) and hydrophobic (yellow on blue background) amino acids just upstream of the cleavage site (vertical downward facing arrow) are also indicated. The small print (green) amino acids in the MUC1, MUC3, and perlecan sequences indicate the splice sites. These are also shown by the red stars. Numbers next to the stars indicate the intron phase.

cleaved (Williams et al. 2001). It appears likely that the SEA module may function as a proteolytic cleavage site in all other SEA module-containing proteins. We furthermore propose that the cleaved SEA module crafts in each case, a specific noncovalent partnership between the resultant cleaved subunits (Fig. 5). It cannot be excluded that additional proteins will be found that harbor "degenerate" SEA modules that can no longer support the proteolytic cleavage and will thus not be cleaved.

Despite the fact that MUC1/Y contains part of the SEA module (Zrihan-Licht et al. 1994b), this alternative splice variant is not proteolytically cleaved (Zrihan-Licht et al. 1994b; Fig. 1A). This indicates that sequences spliced out in the MUC1/Y isoform N-terminal to the phenylalanine residue of the sequence F^{314}NSSLED (top line, MUC1 sequence, Fig. 3; note that the LED sequence within FNSSLED is highlighted red on a yellow background), may be required for efficient cleavage, and that besides the GSVVV site, and sequences directly flanking it, additional sequences are required for efficient cleavage. The absence of MUC1/Y cleavage (Zrihan-Licht et al. 1994b) and the homology of the MUC1 and *C. elegans* proteins N-terminal to FNSSLED (Fig. 4B), tally well with the assignment of the N-terminal border of the SEA module to about 55 amino acids N-terminal to the glycine residue of GSVVV.

The two MUC1 subunits reassociate following cleavage, and present on the cell surface as a heterodimer (Fig. 1A,B; see above). What is the nature of the subunit interaction, and does the SEA module encompass the binding sites of the two subunits? We have found (Baruch et al. 1999) that a soluble secreted form of the MUC1 extracellular domain (MUC1/SEC; Figs. 1,2) generated by alternative splicing just C-terminal to the Y^{325}YQELQRDISEM sequence and terminating 11 amino acids C-terminal to this site (Smorodinsky et al. 1996) binds specifically and tightly to the $GSVVV_{down}$ sequence comprising the 60 amino acids extending from GSVVV to the MUC1 transmembrane domain. It is of note that recombinant $GSVVV_{down}$ fragment generated in bacteria (and thus nonglycosylated) functioned admirably in this binding reaction. However, deletion of the

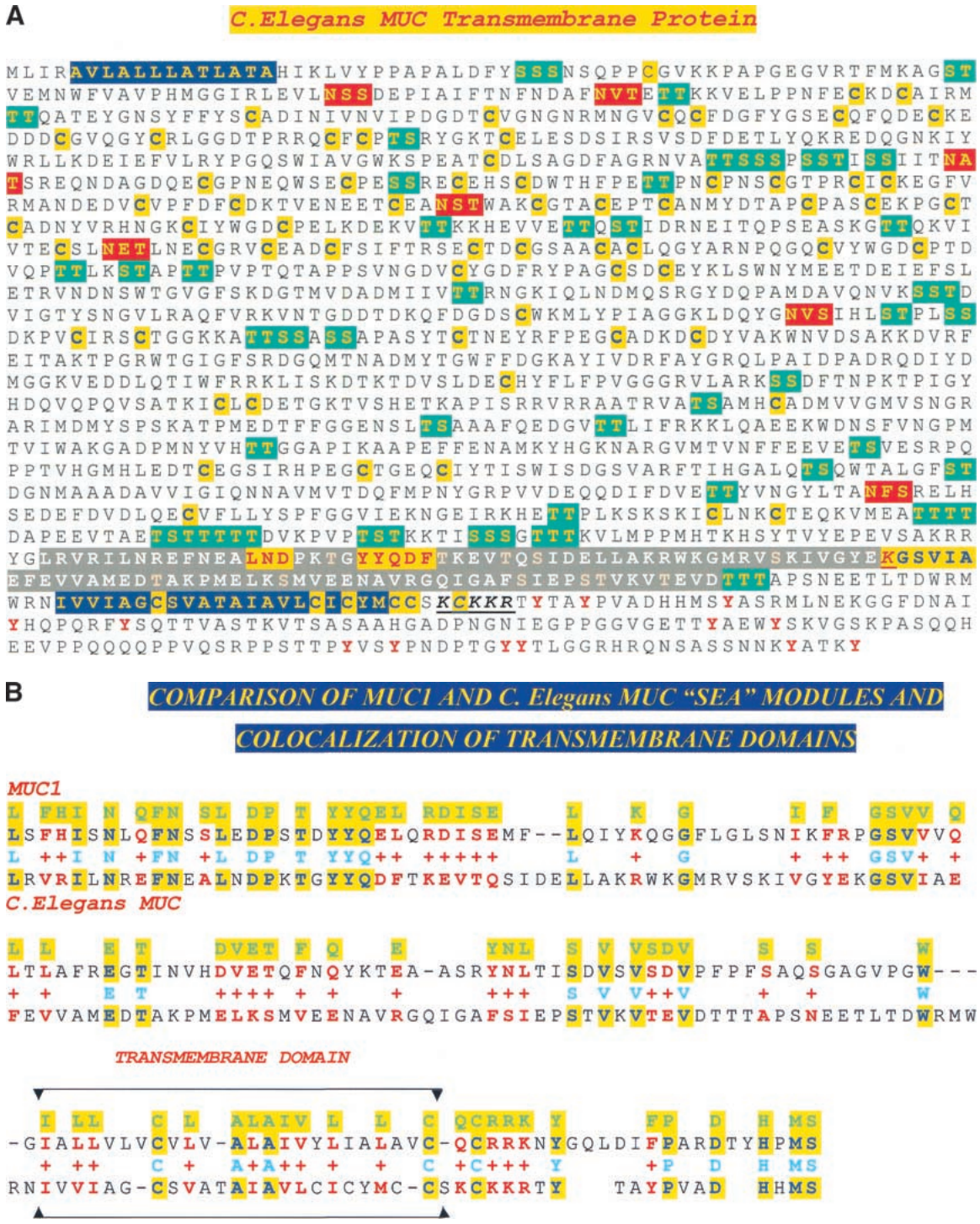

Fig. 4. *C. elegans* SEA module-containing transmembrane protein. (*A*) Sequence of *C. elegans* SEA module-containing transmembrane protein. Yellow letters on a blue background indicate the leader signal peptide and transmembrane domains. The transfer stop signal immediately downstream to the transmembrane domain is written in bold, italic, underlined letters, and the cytoplasmic tyrosine residues are highlighted in red. White letters on a gray background indicate the SEA module juxtaposed upstream of the transmembrane. The conserved motifs within the SEA module are indicated by red letters on a yellow background. Yellow letters on a green background show the strings of threonine and serine residues, and the potential N-glycosylation sites are highlighted yellow on a red background. (*B*) Comparison of *C. elegans* SEA module-containing transmembrane protein with human MUC1. The two sequences are aligned for maximum similarity. Conservatively substituted amino acids are indicated by a plus sign. Note that all hydrophilic amino acids (K, R, D, and E) are considered conservative substitutions for each other.

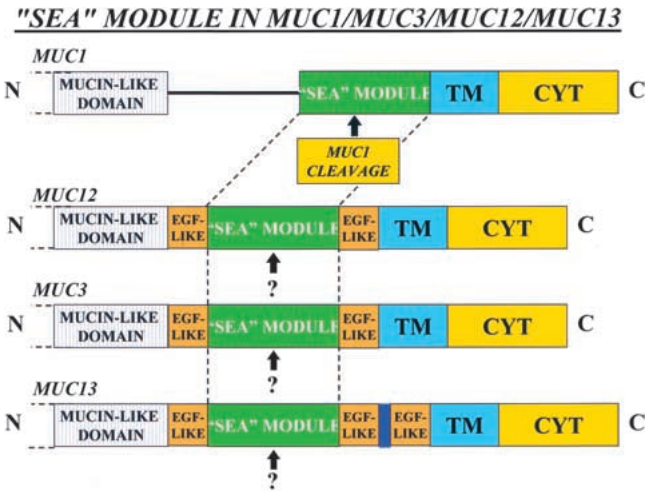

Fig. 5. Proposed cleavage in the MUC12, MUC3, and MUC13 SEA modules. The various proteins are shown schematically. The amino-terminal barred region indicates the tandem repeat mucin domains of these proteins. TM and CYT designate the transmembrane and cytoplasmic domains. The MUC12, MUC3, and MUC13 all have EGF-like domains that bracket the SEA module.

N-terminal 12 amino acids from the $GSVVV_{down}$ fragment abrogated all binding (Baruch et al. 1999). These findings indicate that: (a) at the maximum, 60 amino acids C-terminal to the GSVVV site are required for a productive interaction, (b) the N-terminal part of $GSVVV_{down}$ is essential for subunit association, and (c) no post-translational glycosylation modifications on the $GSVVV_{down}$ fragment are required for subunit binding.

Taken together, one can conclude that the sequences immediately C-terminal to the cleavage site are an integral component of heterodimer formation and, although not related to as such in the original description of the "SEA" module (Bork and Patthy 1995), thus constitute an essential part of the functional SEA module.

With regard to the large, heavily O-linked glycosylated N-terminal MUC1 extracellular domain, deletion of sequences C-terminal to the serine residue of S^{322}TDYYQEL abrogated all subunit interaction (Baruch et al. 1999). This implicates the C-terminal sequences of $GSVVV_{up}$ ($GSVVV_{up}$ refers to sequences N-terminal to the GSVVV cleavage site) as critical elements in subunit association. More surprising was the finding that bacterially generated N-terminal MUC1 extracellular domain had very little binding activity. In contrast, animal cell-generated N-terminal MUC1 extracellular domain performed well in the binding reaction (Baruch et al. 1999). Two conclusions are apparent: (a) sequences immediately N-terminal to the GSVVV cleavage site and contained within the SEA module are an essential component of the subunit association, and (b) post-translational modification of the N-terminal MUC1 extracellular domain may substantially potentiate binding to the $GSVVV_{down}$ subunit.

It is thus clear (Ligtenberg et al. 1992; Parry et al. 2001), that, in the MUC1 protein, the SEA module harbors the proteolytic cleavage site. Moreover, a critical analysis of our results (Baruch et al. 1999) as described above, supports the notion that the SEA module also comprises the binding sites required for the specific association of the two cleaved subunits, and that this interaction is likely to require specific protein structural folds present in the SEA module. Furthermore, the $GSVVV_{up}$ segment is predicted to bear specific post-translational modifications, which may form part of the site recognized by the $GSVVV_{down}$ subunit.

Within all SEA module-containing proteins, the following points are noteworthy: (a) the high conservation of the GSVVV cleavage site, (b) the conservation of the structural folds in the SEA module, (c) the extensive post-translational modifications and especially the presence of a considerable number of O-linked glycans in all of these proteins, and (d) within the extracellular domains, the most conserved region of orthologs from different species shows up in the SEA module.

With these points in mind, we wish to invoke the generality of the function of the SEA module not only as a site for proteolytic cleavage, but also for subsequent reassociation of the subunits. Furthermore, rather than the nonspecific recognition of carbohydrate or heparan sulfate-modified proteins by the SEA module, we contend that variation in the makeup of individual SEA modules allows for a discrete and specific binding between the two subunits formed by bisection of each SEA module-carrying protein.

The theme of a cleaved membrane-associated protein presenting on the cell surface, as a heterodimer is not restricted to the SEA module-cleaved MUC1 protein. An increasing number of 7-transmembrane pass G-protein-coupled receptors (GPCRs) are being discovered that belong to a distinct subfamily of GPCRs, LNB-TM7 (Stacey et al. 2000). The founding member of this family is the macrophage-restricted marker protein F4/80 followed soon thereafter by EMR1 (EGF-like mucin-containing receptor) and EMR2 (Stacey et al. 2000), and references contained therein. Entrance criteria to this select family requires an extended N-terminal extracellular domain that contains, in most cases, one or more EGF-like domains and a mucin-like domain. Invariably, a "cysteine-box" (cys-box) is also located just N-terminal to the transmembrane domain (Stacey et al. 2000). For the latrophilin receptors (CIRLs), which are also LNB-TM7 family members, it has been shown that proteolytic cleavage occurs immediately downstream to the "cys-box" and bisects the protein into two components, which present on the cell surface, in correspondence to the cleaved MUC1, as a heterodimer (Krasnoperov et al. 1997). Although the generality of this phenomenon for all family members is yet to be established, it is not unreasonable to assume this will be so. In support of this, an additional GPCR family member CD97 has been shown to be a

cleaved heterodimer (Gray et al. 1996). Indeed, the proteolytic site has already been designated as GPS (Ponting et al. 1999), *G*PCR *P*roteolytic *S*ite.

For reasons that will become apparent, we would like to direct our attention to the GPCR LNB-TM7 family member, IgHepta (Abe et al. 1999). As for other family members, IgHepta harbors a GPS just N-terminal to the first transmembrane domain and is likely to be cleaved here. More remarkable, is the presence of a perfect SEA module close to the N-terminus of the protein, just C-terminal to an N-terminally located EGF-like module. The IgHepta SEA module bears a perfect GSVVV consensus within the SEA module as well as having the predicted SEA module secondary structure. It would, therefore, appear that IgHepta, a 7-TM GPCR, could be cleaved twice and presents at the cell surface as a heterotrimer.

The assertion that the cell does not perform "anything for nothing" is somewhat glib, but the question does arise, "What is the cell-rationale for cleavage of membrane-associated proteins in their extracellular domains and presentation at the cell surface as heterodi(tri)mers?"

Our proposal is that in so doing the cell generates a receptor–ligand alliance from one and the same gene, and utilizes such a ploy for the purpose of regulating signal transduction. The extracellular domain specifically associates with the membrane-tethered portion. As with all biological noncovalent interactions, such binding has a dissociation constant and the extracellular domain (ligand) can dissociate from its partner (receptor) in what we term the "off" reaction. That this indeed occurs has been demonstrated for the transmembrane mucin, MUC1 (Fig. 1B). The corollary would then be the reassociation of the shed extracellular domain to its unengaged membrane-tethered companion, the "on" reaction (Fig. 1B). It is not unreasonable to assume that both the "off" and "on" reactions perturb the conformation of the membrane-associated partner and thus modulate signal transduction (Fig. 1B). Direct modulation is likely to be the case for cleaved membrane proteins harbouring cytoplasmic tails. Indeed, the MUC1 cytoplasmic domain can be tyrosine phosphorylated (Zrihan-Licht et al. 1994a; Quin and McGuckin 2000). Indirect modulation may be the rule for cleaved membrane-associated proteins lacking a cytoplasmic tail. This general scheme does not preclude the existence of alternative or complementary possibilities that may encompass additional ligands. For example, in the case of the transmembrane mucins MUC1, MUC3, and MUC12, adhesion of bacteria (ligands) to the extracellular mucin domain may induce its shedding ("off" reaction) and thus regulate signal transduction. It has been demonstrated that *Pseudomanas aeruginosa* binds to the extracellular domain of MUC1 inducing tyrosine phosphorylation of its cytoplasmic tail (Lillehoj et al. 2001). It is also likely that additional molecules will either enhance or reduce shedding of the cleaved extracellular domains. These are predicted to be critically important for modulating signal transduction. Whatever the actual case, we contend that cleavage within the extracellular domain is of major functional import vis-à-vis the modulation of signal transduction into the cell.

The hypothesis presented here is likely to have significant practical consequences. It could well lead to the rational design and identification of molecules that, by binding to one of the cleaved partners, will act either as agonists or antagonists, alter signal transduction, and hence, cellular behavior.

Acknowledgments

This work was supported in part by an EMBO short-term fellowship, U.S. Army Medical Research Breast Cancer Grant 17-00-1-0451, Susan G. Komen Breast Cancer Foundation Grant 99-3111, and a Middle East Cancer Consortium Grant 990002, all to D.H.W. and a Federico Grant to I.K.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Abe, J., Suzuki, H., Notoya, M., Yamamoto, T., and Hirose, S. 1999. Ig-hepta, a novel member of the G protein-coupled hepta-helical receptor (GPCR) family that has immunoglobulin-like repeats in a long N-terminal extracellular domain and defines a new subfamily of GPCRs. *J. Biol. Chem.* **274:** 19957–19964.

Baruch, A., Hartmann, M., Yoeli, M., Adereth, Y., Greenstein, S., Stadler, Y., Skornik, Y., Zaretsky, J., Smorodinsky, N.I., Keydar, I., and Wreschner, D.H. 1999. The breast cancer-associated MUC1 gene generates both a receptor and its cognate binding protein. *Cancer Res.* **59:** 1552–1561.

Bork, P. and Patthy, L. 1995. The SEA module: A new extracellular domain associated with O- glycosylation. *Protein Sci.* **4:** 1421–1425.

Coughlin, S.R. 1999. How the protease thrombin talks to cells. *Proc. Natl. Acad. Sci.* **96:** 11023–12037.

Cupit, L.D., Schmidt, V.A., and Bahou, W.F. 1999. Proteolytically activated receptor-3. A member of an emerging gene family of protease receptors expressed on vascular endothelial cells and platelets. *Trends Cardiovasc. Med.* **9:** 42–48.

Dougherty, G.J., Kay, R.J., and Humphries, R.K. 1989. Molecular cloning of 114/A10, a cell surface antigen containing highly conserved repeated elements, which is expressed by murine hemopoietic progenitor cells and interleukin-3-dependent cell lines. *J. Biol. Chem.* **264:** 6509–6514.

Gendler, S.J., Lancaster, C.A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E.N., and Wilson, D. 1990. Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. *J. Biol. Chem.* **265:** 15286–15293.

Gray, J.X., Haino, M., Roth, M.J., Maguire, J.E., Jensen, P.N., Yarme, A., Stetler-Stevenson, M.A., Siebenlist, U., and Kelly, K. 1996. CD97 is a processed, seven-transmembrane, heterodimeric receptor associated with inflammation. *J. Immunol.* **157:** 5438–5447.

Iozzo, R.V., Cohen, I.R., Grassel, S., and Murdoch, A.D. 1994. The biology of perlecan: The multifaceted heparan sulphate proteoglycan of basement membranes and pericellular matrices. *Biochem. J.* **302:** 625–639.

Krasnoperov, V.G., Bittner, M.A., Beavis, R., Kuang, Y., Salnikow, K.V., Chepurny, O.G., Little, A.R., Plotnikov, A.N., Wu, D., Holz, R.W., and Petrenko, A.G. 1997. Alpha-latrotoxin stimulates exocytosis by the interaction with a neuronal G-protein-coupled receptor. *Neuron* **18:** 925–937.

Kuehn, M.H. and Hageman, G.S. 1999. Molecular characterization and genomic mapping of human IPM 200, a second member of a novel family of proteoglycans. *Mol. Cell Biol. Res. Commun.* **2:** 103–110.

Lan, M.S., Batra, S.K., Qi, W.N., Metzgar, R.S., and Hollingsworth, M.A. 1990.

Cloning and sequencing of a human pancreatic tumor mucin cDNA. *J. Biol. Chem.* **265:** 15294–15299.

Ligtenberg, M.J., Kruijshaar, L., Buijs, F., van Meijer, M., Litvinov, S.V., and Hilkens, J. 1992. Cell-associated episialin is a complex containing two proteins derived from a common precursor. *J. Biol. Chem.* **267:** 6171–6177.

Ligtenberg, M.J., Vos, H.L., Gennissen, A.M., and Hilkens, J. 1990. Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene encoding splice variants with alternative amino termini. *J. Biol. Chem.* **265:** 5573–5578.

Lillehoj, E.P., Hyun, S.W., Kim, B.T., Zhang, X.G., Lee, D.I., Rowland, S., and Kim, K.C. 2001. Muc1 mucins on the cell surface are adhesion sites for *Pseudomonas aeruginosa*. *Am. J. Physiol. Lung. Cell. Mol. Physiol.* **280:** L181–L187.

Mehlman, T. and Burgess, W.H. 1990. Detection and characterization of heparin-binding proteins with a gel overlay procedure. *Anal. Biochem.* **188:** 159–163.

Parry, S., Silverman, H.S., McDermott, K., Willis, A., Hollingsworth, M.A., and Harris, A. 2001. Identification of MUC1 proteolytic cleavage sites in vivo. *Biochem. Biophys. Res. Commun.* **11:** 715–720.

Ponting, C.P., Hofmann, K., and Bork, P. 1999. A latrophilin/CL-1-like GPS domain in polycystin-1. *Curr. Biol.* **9:** R585–R588.

Quin, R.J. and McGuckin, M.A. 2000. Phosphorylation of the cytoplasmic domain of the MUC1 mucin correlates with changes in cell–cell adhesion. *Int. J. Cancer* **87:** 499–506.

Smorodinsky, N., Weiss, M., Hartmann, M.L., Baruch, A., Harness, E., Yaakobovitz, M., Keydar, I., and Wreschner, D.H. 1996. Detection of a secreted MUC1/SEC protein by MUC1 isoform specific monoclonal antibodies. *Biochem. Biophys. Res. Commun.* **228:** 115–121.

Spicer, A.P., Duhig, T., Chilton, B.S., and Gendler, S.J. 1995. Analysis of mammalian MUC1 genes reveals potential functionally important domains. *Mamm. Genome* **6:** 885–888.

Stacey, M., Lin, H.H., Gordon, S., and McKnight, A.J. 2000. LNB-TM7, a group of seven-transmembrane proteins related to family-B G-protein-coupled receptors. *Trends Biochem. Sci.* **25:** 284–289.

Williams, S.J., McGuckin, M.A., Gotley, D.C., Eyre, H.J., Sutherland, G.R., and Antalis, T.M. 1999a. Two novel mucin genes down-regulated in colorectal cancer identified by differential display. *Cancer Res.* **59:** 4083–4089.

Williams, S.J., Munster, D.J., Quin, R.J., Gotley, D.C., and McGuckin, M.A. 1999b. The MUC3 gene encodes a transmembrane mucin and is alternatively spliced. *Biochem. Biophys. Res. Commun.* **261:** 83–890.

Williams, S.J., Wreschner, D.H., Tran, M., Eyre, H.J., Sutherland, G.R., and McGuckin, M.A. 2001. MUC13—A novel human cell surface mucin expressed by epithelial and hemopoietic cells. *J. Biol. Chem.* **276:** 18327–18336.

Wreschner, D.H., Hareuveni, M., Tsarfaty, T. Smorodinsky, N., Horev, J., Zaretsky, J., Kotkes, P., Weiss, M., Lathe, R., Dion, A., et al. 1990. Human epithelial tumor antigen cDNA sequences. Differential splicing may generate multiple protein forms. *Eur. J. Biochem.* **189:** 463–473.

Zrihan-Licht, S., Baruch, A., Elroy-Stein, O., Keydar, I., and Wreschner, D.H. 1994a. Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins. Cytokine receptor-like molecules. *FEBS Lett.* **356:** 130–136.

Zrihan-Licht, S., Vos, H.L., Baruch, A., Elroy-Stein, O., Sagiv, D., Keydar, I., Hilkens, J., and Wreschner, D.H. 1994b. Characterization and molecular cloning of a novel MUC1 protein, devoid of tandem repeats, expressed in human breast cancer tissue. *Eur. J. Biochem.* **224:** 787–795.