Evolution of Transmembrane and Gel-forming Mucins Studied with Bioinformatic Methods

AKADEMISK AVHANDLING

som för avläggande av medicine doktorsexamen
vid Sahlgrenska akademin vid Göteborgs universitet
kommer att offentligen försvaras i hörsal Arvid Carlsson,
Academicum, Medicinaregatan 3,
tisdagen den 30 oktober 2007 kl 09.00

av
Tiange Lang

Fakultetsopponent:
Professor Dallas Swallow,
University College London, London

Avhandlingen baseras på följande delarbeten:

I

Lang T., Alexandersson M., Hansson G. C. and Samuelsson T. (2004)

Bioinformatic identification of polymerizing and transmembrane mucins in the puffer fish Fugu rubripes.

Glycobiology 14:521-527.

II

Lang T., Hansson G. C. and Samuelsson T. (2006)

An inventory of mucin genes in the chicken genome shows that the mucin domain of Muc13 is encoded by

multiple exons and that ovomucin is part of a locus of related gel-forming mucins.
BMC Genomics. 7:197-206.

III

Lang, T., Hansson, G. C. and Samuelsson T. (2007)

Gel-forming mucins appeared early in metazoan evolution.

Proc Natl Acad Sci U S A. 104(41):16209-16214.

The Sahlgrenska Academy
AT GÖTEBORG UNIVERSITY
Institutionen för biomedicin

Evolution of Transmembrane and Gel-forming Mucins Studied with Bioinformatic Methods

ABSTRACT

All mucosal membranes of the body are covered by mucus, largely made up of the family of large glycoproteins called mucins. These are instrumental for the protection of the underlying epithelia and involved in the pathogenesis of many diseases in the lungs and the intestine. Several mucins are also involved in the progression of cancer and can often be linked to bad prognosis. The mucins are classified as membrane-bound or secreted. In human there are eight membrane-bound (MUC1, MUC3, MUC4, MUC12, MUC13, MUC15, MUC16, and MUC17) and five secreted and gel-forming mucins (MUC2, MUC5B, MUC5AC, MUC6, and MUC19). Mucins are characterized by domains rich in proline, serine and threonine that are heavily glycosylated (PTS domains) and typically have either von Willebrand D (VWD) or SEA domains.

To aid in understanding this family of proteins we have taken a bioinformatics approach to mine protein and genomic sequence databases for mucins. We have combined different methods to predict mucin proteins. We developed PTSpred, a method to predict PTS domains characteristic of mucins. We also made use of prediction of signal sequences, transmembrane regions, profile based searches and methods to predict genomic regions encoding specific protein domains. We have examined several animals with respect to mucins and other proteins that contain the VWD and SEA domains and have identified numerous novel mucin homologues and mucin-related proteins.

We first made an comprehsnsive inventory of human, mouse and rat mucins including the human chromosome 7q22 region which encodes MUC3, MUC12, and MUC17. During the analysis of the chicken genome we found that the homologues of human chromosome 11p15 gel-forming mucin group (MUC6, MUC2, MUC5AC and MUC5B) are found with the same order as in human, and Muc13 is encoded by a gene where the PTS domain is divided among several exons, where each exon encodes a repeated unit in the protein.

The mucins in *Xenopus tropicalis* are unusual in many respects. The number of gel-forming mucins has been markedly expanded, and the Muc2 homologues contain an unusual PTS domain rich in cysteines. In addition, *Xenopus tropicalis* has a novel family of mucin-like proteins with alternating PTS and SEA domains, a type of protein also identified in the fishes.

The evolution of the MUC4 mucin seems to have occurred by recruitment of a PTS domain to NIDO, AMOP and VWD domains from a sushi-domain containing family of proteins present in lower animals. *Xenopus tropicalis* is the most deeply branching animal where a protein similar to the mammalian MUC4 was identified.

In the gel-forming mucins, a VWD domain typically occurs together with a TIL domain and a domain we have named VWE. We also demonstrated that the gel-forming mucins, von Willebrand factor (VWF), otogelin and insect hemolectin are evolutionary related. Proteins related to these are found in a range of animals, including a mucin in the deeply branching metazoan *Nematostella vectensis* (sea anemone). This demonstrates an early origin of this group of mucins. In contrast, all the transmembrane mucins do not seem to have evolved until the appearance of the vertebrate lineage.

ISBN: 978-91-628-7320-2 **Göteborg 2007**