**THE SEQUENCE PATTERN FOR THE GLYCOSYLPHOSPHATIDYL-ANCHOR POSTTRANSLATIONAL MODIFICATION AND ITS RECOGNITION IN PROPROTEIN SEQUENCES.**

EISENHABER BIRGIT 1,2

1European Molecular Biology Laboratory, Meyerhofstrasse1, Postfach 10.2209, D-69012 Heidelberg, Fed.Rep.Germany ;

2Max-Delbruck-Centrum fur Molekulare Medizin, Robert-Rossle-Strabe 10, D-13122 Berlin-Buch, Fed.Rep.Germany

Keywords: extracellular proteins, glycosylphosphatidylinositol anchoring, posttranslational modification, subcellular localization

Glycosylphosphatidylinositol (GPI) anchoring is a common posttranslational modification of extracellular eukaryotic proteins. Attachment of the GPI moiety to the carboxyl terminus ( -site) of the polypeptide occurs by a transamidation reaction within the endoplasmatic reticulum following proteolytic cleavage of a C-terminal propeptide from the proprotein. Little is known about the putative transamidase but its substrate, the C-terminal sequence segment carrying the GPI-modification signal, has been investigated in detail by site-directed mutations in a few model proteins. Unfortunately, this signal is relatively weakly defined by the available experimental data and appears composed of three sequence regions: (i) the GPI-modification site ( -site), (ii) a moderately polar spacer region with a length of about 8-12 residues (region +1 to about +10), and (iii) an additional hydrophobic C-terminal segment with a length of 10-20 residues (beginning with about +11).

As a result of the experimental work of many groups, sequence databases such as SWISS-Prot have accumulated more and more sequences of GPI-anchored proteins, a resource that has not been exploited yet for the detailed analysis of the GPI-modification motif. In this work, an analysis of the available sequence data is presented which is aimed at a more complete description of this sequence signal in terms of physical properties of amino acid residues that are probably conserved at the motif positions. In addition to a refinement of previously described sequence signals, conserved sequence properties in the regions -11...-1 and +4...+5 (see Table) are reported. There is statistical evidence for volume-compensating residue exchanges with respect to the positions -1... +2. Differences between protozoan and metazoan GPI-modification motifs consist mainly in variations of preferences to amino acid types at the positions near the -site and in the overall motif length. The diversity of polypeptide substrates is exploited to suggest a model of the polypeptide binding site of the putative transamidase, the enzyme catalyzing the GPI-modification. The volume of the active site cleft accommodating the four residues -1...+2 appears about 540A3.

The successful progress of many large-scale genome projects and the subsequent efforts for functional characterization of proteins known by sequence only stimulate analyses of amino acid sequence patterns responsible for posttranslational modifications and prediction techniques for modifications sites. It should be noted that the annotation of a protein as being GPI-anchored is a very valuable functional information since it both defines the subcellular localization and limits the range of possible cellular determinations. Especially for the analysis of Expressed Sequence Tags (ESTs) which very often contain information on C-terminal protein sequences, a GPI-modification prediction tool would be extremely valuable. Such a tool is currently in development.

**Sequence properties of the C-terminal propeptide near the -site**

This table summarizes the results of the correlation analysis with amino acid property indices. The percentage data of amino acid occurrence are related to the alignment of sequences in the largest subset with pairwise sequence identity below 30%.

Table.

|  |  |  |
| --- | --- | --- |
| position relative to the http://www.bionet.nsc.ru/meeting/bgrs/thesis/112/omega.gif-site | Protozoa | Metazoa |
| -11 ... -1 | Unstructured region | unstructured region |
| 0 | Ser ( 44%), Asp, Asn, Ala, Gly | Ser (48%), Gly, Asn, Asp, Cys |
| 1 | similar to position 0 | tiny (Gly, Ala, Ser) |
| 2 | Ser + Ala (94%) | Ala + Gly (70%) |
| 4 ... 5 | Hydrophobic | Hydrophobic |
| 9 | onset of hydrophobic region up to about w+25,mostly Leu | onset of hydrophobic region up to about http://www.bionet.nsc.ru/meeting/bgrs/thesis/112/omega.gif+31,mostly Leu |
| 15 | --- | Ser + Thr + Ala (60%) |

**Acknowledgements**

The author is thankful to F. Eisenhaber, P. Bork, and J.G. Reich for constant support of this work.