Transamidase subunit GAA1/GPAA1 is a M28 family metallo-peptide-synthetase that catalyzes the peptide bond formation between the substrate protein’s omega-site and the GPI lipid anchor’s phosphoethanolamine

Birgit Eisenhaber¹, Stephan Eisenhaber², Toh Yew Kwang¹, Gerhard Grüber¹,³, and Frank Eisenhaber¹,⁴,⁵,∗

¹Bioinformatics Institute (BII); A*STAR; Singapore, Republic of Singapore; ²Department of Physical Chemistry; University of Vienna; Wien/Vienna, Republic of Austria; ³Nanyang Technological University; School of Biological Sciences; Singapore, Republic of Singapore; ⁴Department of Biological Sciences (DBS); National University of Singapore (NUS); Singapore, Republic of Singapore; ⁵School of Computer Engineering (SCE); Nanyang Technological University (NTU); Singapore, Republic of Singapore

The transamidase subunit GAA1/GPAA1 is predicted to be the enzyme that catalyzes the attachment of the glycosylphosphatidylinositol (GPI) lipid anchor to the carbonyl intermediate of the substrate protein at the ω-site. Its ~300-amino acid residue luminal domain is a M28 family metallo-peptide-synthetase with an α/β hydrolase fold, including a central 8-strand β-sheet and a single metal (most likely zinc) ion coordinated by 3 conserved polar residues. Phosphoethanolamine is used as an adaptor to make the non-peptide GPI lipid anchor look chemically similar to the N terminus of a peptide.

Introduction

The transamidase complex catalyzes the glycosylphosphatidylinositol (GPI) lipid anchor attachment to substrate proteins of eukaryotes in the lumen of the endoplasmic reticulum (ER). It remains one of the poorly understood macromolecular machines, both with regard to the molecular function of its many subunits, as well as their 3D structure, despite more than 25 years of research in the vertebrate, yeast, and trypanosomal model systems.¹⁻⁵ The reaction consists of 2 steps. First, a C-terminal propeptide is cleaved from the substrate protein. Then in the next step, a peptide bond is formed between the newly established C-terminal residue (called ω-site) of the substrate protein and a phosphoethanolamine group of the GPI lipid anchor. The C-terminal, 4-partite sequence pattern for GPI lipid anchoring in substrate proteins is well established and can recognize substrate proteins with high sensitivity and low false-positive prediction rate.⁶⁻¹⁰ The GPI lipid anchor pathway has a role in multiple human pathologies¹ including cancer.¹¹

In human, the known subunits of the GPI lipid anchor transamidase complex are PIG-K (Gpi8p in yeast), PIG-S (Gpi17p), PIG-T (Gpi16p), GAA1 (GAA1), and the subunit PIG-U (CDC91/GAB1) was found most recently.¹² Subunits PIG-K and PIG-T were discovered to form a covalent complex via a disulfide bridge.¹³ PIG-K is a C13-clade cysteine protease with a predicted 3D structure similar to that of gingipain R and caspases.¹ It is known to cleave the C-terminal propeptide from the substrate protein even in the absence of a GPI lipid anchor.¹⁴⁻¹⁷ PIG-K’s low-resolution structure was determined recently.¹⁸ The 3D structure of PIG-T is predicted to be a C-terminal β-propeller complemented with an N-terminal α-helical hook that embraces the protease PIG-K.¹ It is thought that PIG-T shields the active site of PIG-K from attacking unrelated proteins.

So far, the molecular functions and structures of the remaining 3 subunits remain in the dark. Here, we report sequence-analytic evidence that the luminal domain of GAA1/GPAA1 has a 3D structure similar to that of an M28-type aminopeptidase. We suggest that GAA1/GPAA1 is the prime and only candidate for the missing enzyme that catalyzes the formation of the peptide bond between the ω-site and a phosphoethanolamine group of the GPI lipid anchor.

Results and Discussion

The sequence architecture¹⁹,²⁰ of GAA1/GPAA1 provides for an N-terminal transmembrane (TM) region followed by a segment of ~300 residues located in the ER lumen and further 6 TM helices.¹ If the luminal GAA1/GPAA1 segments from a wide variety of taxa are queried with HHPRED against the HMM database derived from sequences with known structures (pdb_6Feb14),²¹ a sub-structure of ~290 residues generates
Table 1. Sequence similarity searches with the lumenal domain segment of GAA1/GPAA1 with HHpRED

| Query sequence: ID | 4fu_A(309): E-value Sequence segment | 3gux_A(314): E-value Sequence segment | 4fu_A(312): E-value Sequence segment | 3tc8_A(309): E-value Sequence segment | 1tkj_A(284): E-value Sequence segment |
|------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Hs_O43292 66–348 | 3.9e-28 23–305 4.6e-28 25–312 4.1e-28 8–294 1.2e-24 23–306 1.4e-20 5–277 |
| Mm_Q9WTK3 66–348 | 1.7e-28 23–305 1.8e-28 25–312 6.4e-28 8–294 5.1e-25 23–266 1.1e-20 5–277 |
| Dm_NP_572273 70–365 | 3.5e-30 25–305 3.1e-30 28–312 5.9e-32 14–294 1.4e-28 25–306 3.9e-25 7–277 |
| Ce_NP_491700 83–370 | 2.3e-26 22–266 9.3e-27 24–310 1.8e-27 31–294 4e-23 22–265 8.7e-20 8–277 |
| Sc_P39012 57–331 | 7.6e-17 21–266 9.7e-18 23–310 1.1e-17 31–294 4.3e-16 9–265 4.7e-13 5–277 |
| Pf_XP_002809111 59–337 | 4.4e-08 37–266 1.1e-08 50–310 3.9e-11 32–294 7.6e-09 48–265 2.3e-08 35–277 |

The table presents hits found with HHpRED when using the lumenal domain segments of the GAA1/GPAA1 protein sequences of various taxa. Ce, Ceanorhabditis elegans; Dm, Drosophila melanogaster; Hs, Homo sapiens; Mm, Mus musculus; Pf, Plasmodium falciparum; Sc, Saccharomyces cerevisiae. The first column shows the accession number and the alignable sequence segment. The following 5 columns, separately for each structure, present the PDB structure code, the sequence length in the first row, and the E-value of the hit and the aligned segment in each following row.
Inspection of the chemistry linking the substrate protein’s ω-site with the GPI lipid anchor shows that the naturally used adaptor moiety, a phosphoethanolamine, pre-attached to the anchor actually forms a peptide bond with the C-terminal amino acid (Fig. 2). Since catalyzers facilitate reactions in both directions, with the net result depending on the circumstances, we conclude that the luminal domain of GAA1/GPAA1 is the enzyme still missing that catalyzes the formation of the peptide bond between the ω-site and the respective phosphoethanolamine moiety.

To emphasize, GAA1/GPAA1 is the most plausible candidate for this function among the remaining 3 transamidase units (including PIG-S and PIG-U), as previous indirect hints from literature and sequence studies indicate. Most importantly,
PIG-K (Gpi8p) and GPAA1 (GAA1) were the first transamidase subunits discovered.\textsuperscript{12,36-39} The respective mutations led to the accumulation of completely synthesized, free GPI lipid anchors. With hindsight, these 2 transamidase subunits are the enzymes, and they would provide the easiest measurable (all-or-none) effect in a mutation screen. The sub-complex of Gpi8p (PIG-K), Gp16p (PIG-T), and GAA1 (GPAA1), the catalytic core of the transamidase, is most tolerant to purification conditions.\textsuperscript{13,40} As a side note, previous work has shown that preparations of purified GAA1/GPAA1 suffer from slow degradation.\textsuperscript{29,40,41} It cannot be excluded that GAA1/GPAA1 might have some exopeptidase activity when isolated, and this activity could be responsible for the observed instability.

The addition of phosphoethanolamine to the tetrasaccharide during synthesis of the GPI lipid anchor was experimentally proven to be absolutely instrumental before attachment of the anchor to the substrate protein can occur.\textsuperscript{5,42,43} To note, nature uses phosphoethanolamine as adaptor in this case to make the non-peptide GPI lipid anchor appear as the N terminus of a peptide or amino acid, so that a peptidase module can be evolutionarily repurposed for catalyzing the lipid anchor attachment. In this context, it is intriguing that sortase A (SrtA), a completely different, C60 family (trans-) peptidase\textsuperscript{31} from gram-positive bacteria can be used for chemoenzymatic coupling of peptides and proteins to GPI lipid anchors in an artificial system.\textsuperscript{44}

In the 2003 review,\textsuperscript{1} it was hypothesized that GAA1/GPAA1 binds the free GPI lipid anchor for consumption by the transamidase complex. The concept of simple/complex TM regions can be used to distinguish between mere hydrophobic anchors in the membrane (simple TMs) in contrast to complex TMs that fulfill also other structural and/or functional roles.\textsuperscript{45-47} With the exception of the \textit{Plasmodium falciparum} case, all other GAA1/GPAA1 sequences studied (human, fly, worm, yeast, \textit{Arabidopsis thaliana}, \textit{Leishmania}, \textit{Trypanosoma}) have a least 6 complex TMs (as reported by the TMSOC server).\textsuperscript{45} It was experimentally shown that the GAA1/GPAA1 TM regions (especially the C-terminal

\textbf{Figure 2.} The peptide bond linking the ω-site of the substrate protein with the phosphoethanolamine of the GPI lipid anchor. The typical chemical structure of the GPI lipid anchor\textsuperscript{4} and its linkage via the ω-site to the substrate protein for the transamidase reaction are schematically illustrated (drawn with the software suite ChemBioDraw/Perkin Elmers). The GPI lipid anchor itself is shown in black. Only its terminal phosphoethanolamine unit is presented in green color. The substrate protein is colored red, with “R” designating the side chain of the ω-site residue. Only residues Ala, Asn, Asp, Cys, Gly, and Ser are possible in this position.\textsuperscript{1} The peptide bond between the phosphoethanolamine unit and the ω-site residue (in blue) is marked with an arrow. It is thought that this bond is established with catalytic support from the luminal domain of GAA1/GPAA1.
one with a conserved proline) are important for binding the GPI lipid anchor in a functionally productive manner to the transamidase complex.14,48 With regard to the other 2, more loosely bound, transamidase components, PIG-U would have too small a luminal domain for exhibiting protease activity, and PIG-S appears to carry too few TMs (just 2) to hold the GPI lipid anchor moiety.

Notably, the recently published genome of the fungus *Glarea lozoyensis* ATCC 2086849 includes the gene coding for the protein EPE25974, annotated just as “Zn-dependent exopeptidase”, obviously, by an automated annotation pipeline. Actually, this is the GAA1 for this fungal organism. Apparently, the density of sequences has become large enough toward late 2012 that automated annotation pipelines have recognized the aminopeptidase-like luminal domain, though the more obvious function as GAA1 became obscured in the process.

To summarize, the transamidase subunit GAA1/GPAA1 is the long sought for enzyme that catalyzes the attachment of the GPI lipid anchor to the carbonyl intermediate of the substrate protein at the ω-site. Its luminal domain is a metallo–peptidase synthetase with an α/β hydrolase fold and a central 8-strand β-sheet and a single metal (most likely zinc) ion coordinated by 3 conserved polar residues. Phosphoethanolamine is used as an adaptor to make the non-peptide GPI lipid anchor look chemically like the N terminus of a peptide.

Functional characterization of non-understood genome regions, especially of protein-coding genes, is certainly the most pressing task in life sciences today.6,50 This discovery of GAA1/GPAA1’s molecular function will help to understand the biochemical mechanisms of GPI lipid anchoring and help to interfere into the process pharmacologically, for example in battling parasites.6

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
Sebastian Maurer-Stroh is acknowledged for advice. We are grateful to the Agency for Science, Research and Technology (A*Star) Singapore. This work was partially supported by the grants A*STAR-NKTH 10/1/06/24635, IAF311010, A*Star IMAGIN, and IAF311011. Westley Sherman is thanked for language editing advice.

References
1. Eisenhaber B, Maurer-Stroh S, Novatchkova M, Schneider G, Eisenhaber F. Enzymes and auxiliary factors for GPI lipid anchor biosynthesis and post-transferal translation to proteins. Bioessays 2003; 25:367-85; PMID:12655644; http://dx.doi.org/10.1002/bies.10254
2. Ferguson MA. The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. J Cell Sci 1999; 112:2799-809; PMID:10444375
3. Kinoshita T, Inoue N. Dissecting and manipulating the pathway for glycosylphosphatidylinositol-anchor biosynthesis.Curr Opin Chem Biol 2000; 4:632-8; PMID:11102867; http://dx.doi.org/10.1016/s1367-5931(00)00151-4
4. Eisenhaber B, Eisenhaber F. Posttranslational modifications and subcellular localization signals: indicators of sequence regions without inherent 3D structure? Curr Protein Pept Sci 2007; 8:197-203; PMID:17430201; http://dx.doi.org/10.2174/138920307780363424
5. Paulick MG, Bertozzi CR. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. Biochemistry 2008; 47:6991-7000; PMID:18557633; http://dx.doi.org/10.1021/bi8006324
6. Eisenhaber B, Bork P, Eisenhaber F. Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. Protein Eng 1998; 11:1155-61; PMID:9930665; http://dx.doi.org/10.1093/protein.11.12.1155
7. Eisenhaber B, Bork P, Eisenhaber F. Prediction of potential GPI-modification sites in proprotein sequences. J Mol Biol 1999; 292:741-58; PMID:10497036; http://dx.doi.org/10.1006/jmbi.1999.3069
8. Eisenhaber B, Wildpaner M, Schultz CJ, Borner GH, Dupree P, Eisenhaber F. Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for Arabidopsis and rice. Plant Physiol 2003; 133:1061-701; PMID:14681532; http://dx.doi.org/10.1104/pp.103.013580
9. Eisenhaber B, Schneider G, Wildpaner M, Eisenhaber F. A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for Aspergillus nidulans, Candida albicans, Neurospora crassa, Saccharomyces cerevisiae and Schizosaccharomyces pombe. J Mol Biol 2004; 337:242-53; PMID:15003443; http://dx.doi.org/10.1016/j.jmb.2004.01.025
10. Eisenhaber B, Eisenhaber F. Prediction of post-translational modification of proteins from their amino acid sequence. Methods Mol Biol 2010; 609:365-84; PMID:20221930; http://dx.doi.org/10.1007/978-1-60327-241-4_21
11. Gamage DG, Hendrickson TL. GPI transamidase and GPI anchored proteins: oncogenes and biomarkers for cancer. Crit Rev Biochem Mol Biol 2013; 48:446-64; PMID:23978072; http://dx.doi.org/10.1080/10409238.2013.831024
12. Hong Y, Ohishi K, Kang JY, Tanaka S, Inoue N, Nishimura J, Maeda Y, Kinoshita T, Human PIG-U and yeast Cde9p are the fifth subunit of GPI transamidase that attaches GPI-anchors to proteins. Mol Biol Cell 2003; 14:1780-9; PMID:12802054; http://dx.doi.org/10.1091/mbc.E02-12-0794
13. Ohishi K, Nagamura K, Maeda Y, Kinoshita T. Two subunits of glycosylphosphatidylinositol transamidase, PIGI8 and PIG-T, form a functionally important intermolecular disulfide bridge. J Biol Chem 2003; 278:13959-67; PMID:12921867; http://dx.doi.org/10.1074/jbc.M300586200
14. Chen R, Udenfriend S. An active carbonyl formed during gpi anchoring of nascent polypeptides by the glycosylphosphatidylinositol addition. Proc Natl Acad Sci U S A 1995; 92:1550-4; PMID:7878018; http://dx.doi.org/10.1073/pnas.92.5.1550
15. Maxwell SE, Ramalingam S, Gerber LD, Udenfriend S. Cleavage without anchor addition accompanies the processing of a nascent protein to its glycosylphosphatidylinositol-anchored form. Proc Natl Acad Sci U S A 1995; 92:1550-4; PMID:7878018; http://dx.doi.org/10.1073/pnas.92.5.1550
16. Ramalingam S, Maxwell SE, Medof ME, Chen R, Gerber LD, Udenfriend S. COOH-terminal processing of nascent polypeptides by the glycosylphosphatidylinositol transamidase in the presence of hydrazine is governed by the same parameters as glycosylphosphatidylinositol addition. Proc Natl Acad Sci U S A 1996; 93:7528-33; PMID:8755508; http://dx.doi.org/10.1073/pnas.93.15.7528
17. Tok YK, Kamarial N, Maurer-Stroh S, Roessle M, Eisenhaber F, Adhikari S, Eisenhaber B, Grübler G. Structural insight into the glycosylphosphatidylinositol transamidase subunits PIG-K and PIG-S from yeast. J Struct Biol 2011; 173:221-81; PMID:21314462; http://dx.doi.org/10.1016/j.jsb.2010.11.026
18. Ooi HS, Kwo CY, Wildpaner M, Sirota FL, Eisenhaber B, Maurer-Stroh S, Wong WC, Schleifer A, Eisenhaber F, Schneider G. ANNIIE: integrated de novo protein sequence annotation. Nucleic Acids Res 2009; 37:W435-40; PMID:19389726; http://dx.doi.org/10.1093/nar/gkp254
19. Schneider G, Sherman W, Kuchibhatla D, Ooi HS, Sirota FL, Maurer-Stroh S, Eisenhaber B, Eisenhaber F. Protein sequence–structure–function–network links discovered with the ANNOTATOR software suite: application to Elys/Mel-28. In: Trajanoski Z, ed. Computational Medicine. 2012; Vienna: Springer.
20. Hilderbrand A, Remmert M, Biegert A, Soding J. Fast and accurate automatic structure prediction with H Pred. Proteins 2009; 77(Suppl 9):S28-32; PMID:19626712; http://dx.doi.org/10.1002/prot.22409
21. Koch B, Kolenco P, Buchholz M, Carrillo DR, Pantzier C, Wermann M, Rahfeld JU, Reuter G, Schilling S, Stubbs MT, et al. Crystal structures of glutaminyl cyclases (QC) from Drosophila melanogaster reveal active site conservation between insect and mammalian QCs. Biochemistry 2012; 51:7383-92; PMID:22897232; http://dx.doi.org/10.1021/ bi3006087
