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 Eisenhaber[1], Stephan Eisenhaber[2], Toh Yew Kwang[1], Gerhard Grüber[1,3], and Frank Eisenhaber[1,4,5,]*

[1]Bioinformatics Institute (BII); A*STAR; Singapore, Republic of Singapore; [2]Department of Physical Chemistry; University of Vienna; Wien/Vienna, Republic of Austria; [3]Nanyang Technological University; School of Biological Sciences; Singapore, Republic of Singapore; [4]Department of Biological Sciences (DBS); National University of Singapore (NUS); Singapore, Republic of Singapore; [5]School 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 glyco-sylphosphatidyl (GPI) lipid anchor to the carbonyl intermediate of the substrate protein at the ω-site. Its ~300-amino acid residue lumenal 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 glycosylphosphati-dylinositol (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 sub-units, 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 estab-lished 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 sub-strate proteins with high sensitivity and low false-positive predic-tion rate.[6-10] The GPI lipid anchor pathway has a role in multiple human pathologies[1] including cancer.[11]

In human, the known subunits of the GPI lipid anchor trans-amidase complex are PIG-K (Gpi8p in yeast), PIG-S (Gpi17p), PIG-T (Gpi16p), GPAA1 (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 disul-phide bridge.[13] PIG-K is a C13-clade cysteine protease with a pre-dicted 3D structure similar to that of gingipain R and caspases.[1]

It is known to cleave the C-terminal propeptide from the sub-strate 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 remain-ing 3 subunits remain in the dark. Here, we report sequence-analytic evidence that the lumenal 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 pep-tide bond between the ω-site and a phosphoethanolamine group of the GPI lipid anchor.

Results and Discussion

The sequence architecture[19,20] 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 lumenal GAA1/GPAA1 segments from a wide variety of taxa are queried with HHPRED against the HMM database derived from sequences with known struc-tures (pdb_6Feb14),[21] a sub-stretch of ~290 residues generates

significant, full-length hits into numerous M28-type peptidase sequences with structures such as 4fuu_A, 3gux_A, 4f9u_A,[22] 3tc8_A, and 1tkj_A[23,24] (see **Table 1**). With the identical queries, the structure prediction tool PHYRE2[25] delivers hits to the same set of proteins with confidence scores in the range 97–100%. Remarkably, this is despite the fact that the sequence identities of the alignments generated by PHYRE2 are very low (9–17% except for the one match between 1tkj_A and the fly GPAA1 NP572273, where it is 21%; to note, sequence identity loses its predictive role for tertiary structure similarity at the threshold ~30%[26]). Whereas PSI-BLAST[27,28] did collect only GAA1/GPAA1 sequences from various eukaryote organisms in 2003,[1] a decade later with much larger sequence databases, it runs into microbial sequences with similarity to M28-type peptidases beginning with round 2 (e.g., the first hit CBH37168 from an unidentified archebacterial species with significance 1E-05).

Thus, there is no doubt that the lumenal domain of GAA1/GPAA1 has a structure very similar to that of M28 peptidases. As shown in the alignment in **Figure 1**, this is an α/β-hydrolase fold with a central β-sheet consisting of 8 strands surrounded by 7 α-helices and a GAA1/GPAA1-specific elaboration closer to the C terminus (N-terminal to α-helix 6) with a possible additional α-helix. It was serendipitously found with small angle X-ray scattering (SAXS) that the GAA1 sub-segment 70–247 from baker's yeast forms a compact structure (~two-thirds of the whole domain, corresponding to the stretch with secondary structural elements from α2 up to α5 in **Fig. 1**),[29] and its secondary structural content measured with circular dichroism (28% α-helix and 27%β-sheet) is quite well approximated by this prediction (25% α-helix and 20% β-sheet as lower estimates in **Fig. 1**; i.e., about equal amounts of both types of secondary structure and each type covering close to a quarter of the total sequence).

M28 family peptidases are metalloenzymes that usually carry 2, sometimes (e.g., in the case of the glutaminyl cyclase[22,30]) 1 metal ion (most often zinc) in a tetrahedral coordination. Three of the 4 ligands required are provided by amino acid residues from the protein structure. These metal ions have catalytic function (see the peptidase database MEROPS[31]). In the case of the 2-ion *Streptomyces griseus* aminopeptidase 1tkj_A, these are residues His85, Asp97, and Asp 160 for Zn1 (sites 1, 2, and 4 in **Fig. 1**) and Asp97 (shared among the 2 zincs), Glu132, and His247 for Zn2 (sites 2, 3, and 5 in **Fig. 1**). Although all 5 sequence positions are conserved in the mammalian cyclase sequences, only the second metal ion site is occupied.[22,30,32] In GAA1/GPAA1, 4 of the 5 sequence positions (i.e., sites 2, 3, 4, and 5) in the respective 4 loops (C-terminal of β3 and β3-α3, β4-α4, C-terminal of β5, β8-α7; see **Fig. 1**) can be aligned with residues suitable for zinc binding, such as Asp, Glu, His, Tyr.[33,34] Given the local sequence conservation pattern, we suggest Asp153, Asp188, Glu226, and Tyr328 in human GPAA1 (O43292) as the equivalents of Asp97, Glu132, Asp160, and His 247 from *S. griseus*. To note, the residues are determined to the accuracy of the loop between secondary structural elements. Sometimes there are several candidate residues within the respective loops, where, generally, alignments are quite uncertain; thus, our assignment might not be the final word in every case. To conclude, we think that GAA1s/GPAA1s carry one metal binding site corresponding to Zn2, similar to the glutaminyl cyclase case (sites 2, 3, and 5 in **Fig. 1**).

The M28 family of metalloproteases is unusual in that it contains mostly aminopeptidases (cleaving N-terminal amino acids) but also carboxypeptidases (cleaving C-terminal residues; see the peptidase database MEROPS[31]). It includes also enzymes catalyzing some more exotic chemistry, such as the cyclization of N-terminal glutamine (with the formation of pyroglutamine for the loss of N-terminal basicity).[22,30] Neither of these catalytic options appears of direct relevance in the transamidase context. In particular, the protease function needed for cleaving the C-terminal propeptide from the substrate protein has already been solidly associated with PIG-K/Gpi8p.[15-17,35]

Table 1. Sequence similarity searches with the lumenal domain segment of GAA1/GPAA1 with HHPRED

Query sequence: ID Sequence segment	4fuu_A(309): E-value Sequence segment	3gux_A(314): E-value Sequence segment	4f9u_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**).[4] Since catalyzers facilitate reactions in both directions, with the net result depending on the circumstances, we conclude that the lumenal 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.[1] Most importantly,
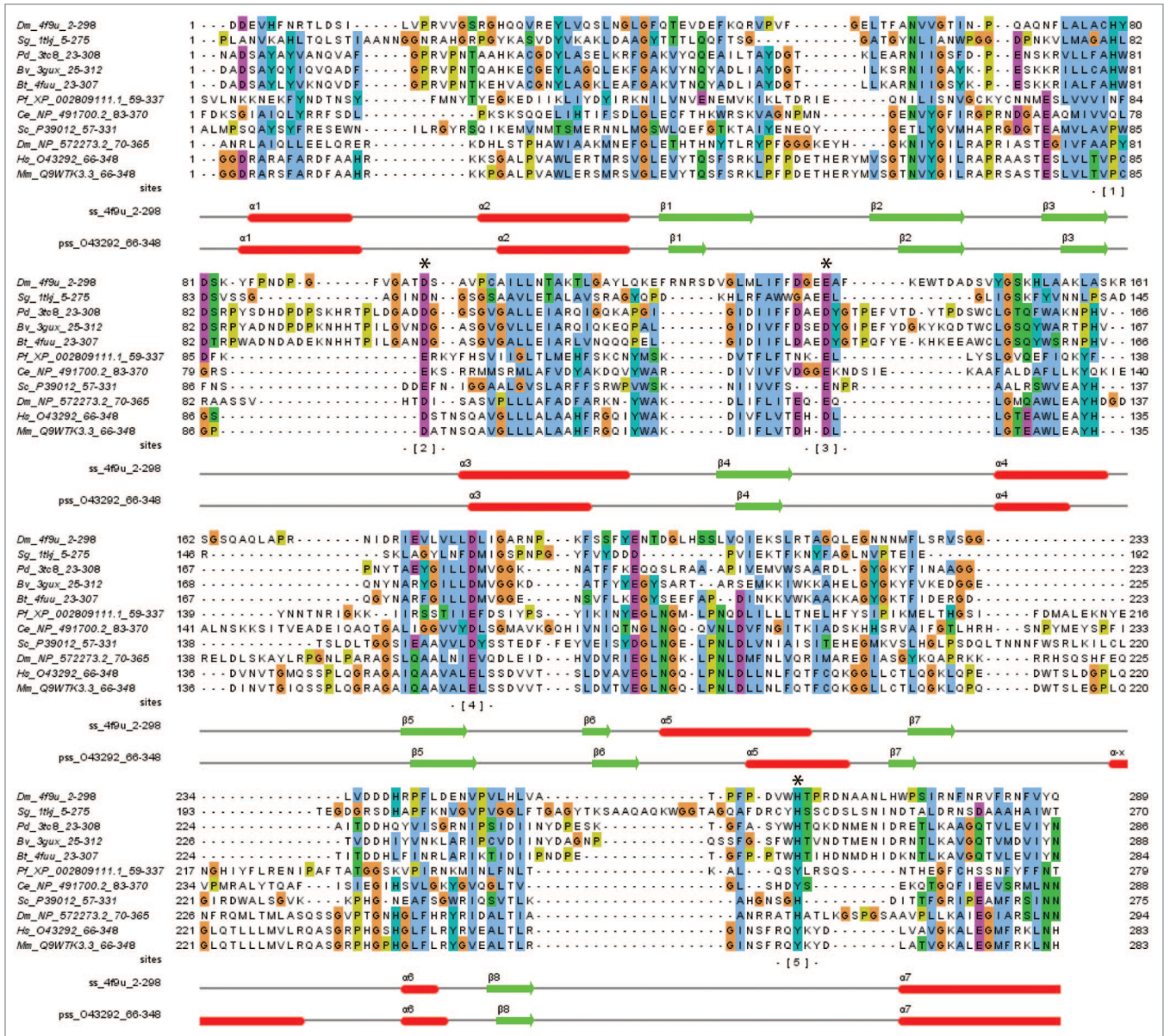

Figure 1. Alignment of GAA1/GPAA1 sequences with structures of M28 peptidases. A representative set of lumenal domain segments of GAA1/GPAA1 protein sequences is shown together with sequences from the metalloprotease M28 family of proteins (with Protein Structure Data Bank [PDB] and chain identifiers 4fuu_A, 3gux_A, 4f9u_A,[22] 3tc8_A and 1tkj_A[23,24]). Five sites (numbered 1, 2, 3, 4, and 5; for better visibility, surrounded by brackets) indicate conserved polar residue positions that are known to play in role for metal ion binding for various members of this family. The Zn2 site conserved among the GAA1/GPAA1 sequences is formed by residues at site positions 2, 3, and 5 (all marked by a star [*] on top of the alignment column). We also show the experimentally determined secondary structure of 4f9u_A (α-helices as red and β-strands as green bars) and the predicted secondary structure for the human GPAA1 (O43292) derived with HHPRED.[21] There is one GAA1/GPAA1-specific helix denoted as "α$_x$" between β7 and α6. The multiple alignments were created with input from HHPRED[21] and MUSCLE[52], and manually adjusted subsequently. The final figure was generated with JalView[53] using CLUSTALX[54] coloring. The 2-letter prefix in front of the accession numbers and the sequence ranges denotes the species. Bt, *Bacteroides thetaiotaomicron*; Bv, *Bacteroides vulgatus* ATCC 8482; Ce, *Ceanorhabditis elegans*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Pd, *Parabacteroides distasonis*; Pf, *Plasmodium falciparum*; Sc, *Saccharomyces cerevisiae*; Sg, *Streptomyces griseus*.
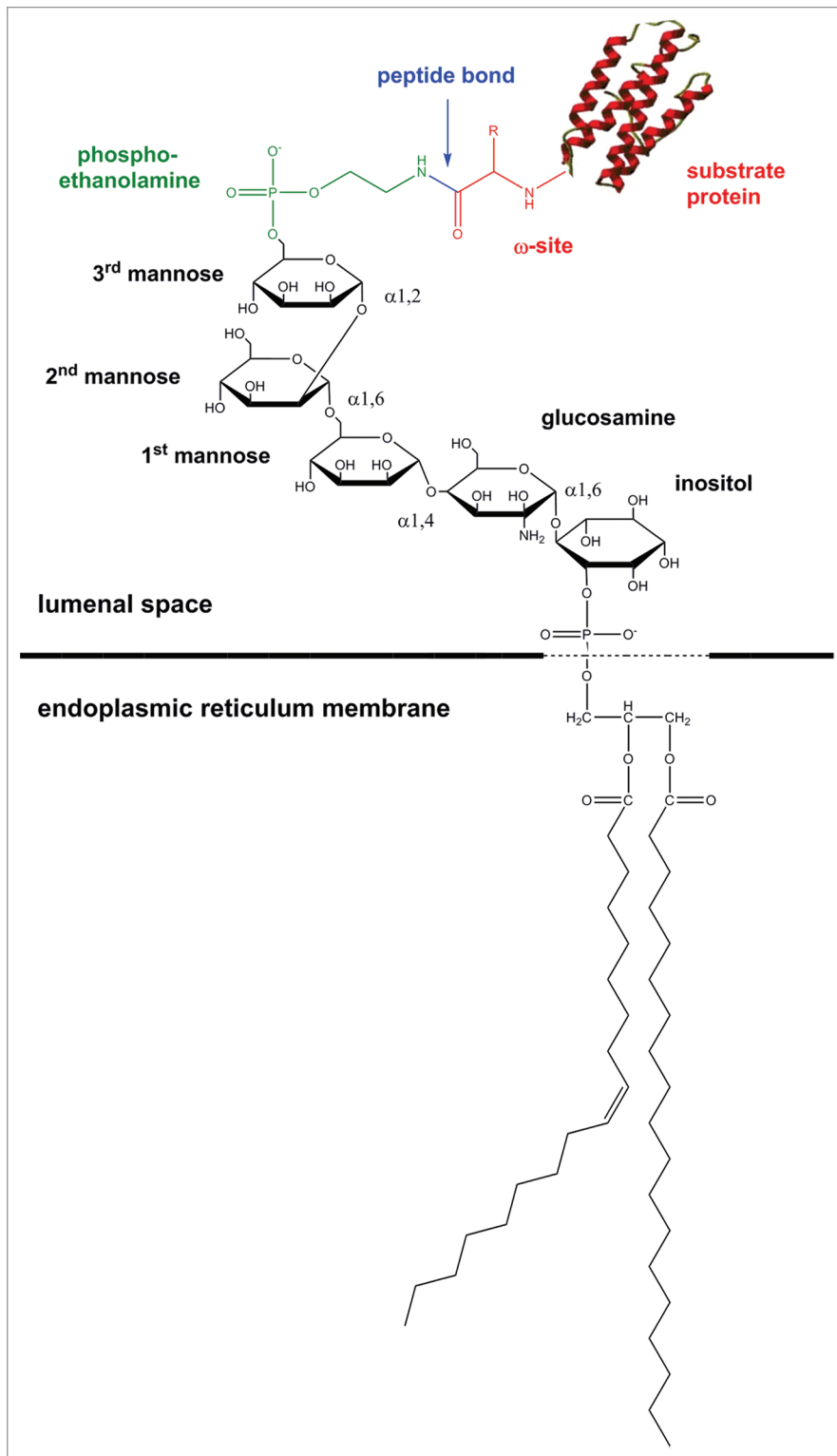
peptide bond

phospho-ethanolamine

substrate protein

ω-site

3rd mannose

α1,2

2nd mannose

α1,6

1st mannose

glucosamine

α1,6

inositol

α1,4

lumenal space

endoplasmic reticulum membrane

Figure 2. The peptide bond linking the ω-site of the substrate protein with the phospho-ethanolamine of the GPI lipid anchor. The typical chemical structure of the GPI lipid anchor[4,5] 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.[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 lumenal domain of GAA1/GPAA1.

transamidase, is most tolerant to purification conditions.[13,40] As a side note, previous work has shown that preparations of purified GAA1/GPAA1 suffer from slow degradation.[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.[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[31] from gram-positive bacteria can be used for chemoenzymatic coupling of peptides and proteins to GPI lipid anchors in an artificial system.[44]

In the 2003 review,[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

PIG-K (Gpi8p) and GPAA1 (GAA1) were the first transamidase subunits discovered.[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

membrane (simple TMs) in contrast to complex TMs that fulfil also other structural and/or functional roles.[45-47] With the exception of the *Plasmodium falciparum* case, all other GAA1/GPAA1 sequences studied (human, fly, worm, yeast, *Arabidopsis thaliana*, *Leishmania*, *Trypanosoma*) have a least 6 complex TMs (as reported by the TMSOC server[45]). It was experimentally shown that the GAA1/GPAA1 TM regions (especially the C-terminal

one with a conserved proline) are important for binding the GPI lipid anchor in a functionally productive manner to the transamidase complex.[41,48] With regard to the other 2, more loosely bound, transamidase components, PIG-U would have too small a lumenal 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 20868[49] 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 lumenal 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 lumenal domain is a metallo–peptide 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.[50,51] 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.[2]

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Eisenhaber B, Maurer-Stroh S, Novatchkova M, Schneider G, Eisenhaber F. Enzymes and auxiliary factors for GPI lipid anchor biosynthesis and post-translational transfer 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 glycosylphos-phatidylinositol-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; 133:1691-701; PMID:14681532; http://dx.doi.org/10.1104/pp.103.023580

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; 337:243-53; 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/10.3109/10409238.2013.831024

12. Hong Y, Ohishi K, Kang JY, Tanaka S, Inoue N, Nishimura J, Maeda Y, Kinoshita T. Human PIG-U and yeast Cdc91p 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, Nagamune K, Maeda Y, Kinoshita T. Two subunits of glycosylphosphatidylinositol transamidase, GPI8 and PIG-T, form a functionally important intermolecular disulfide bridge. J Biol Chem 2003; 278:13959-67; PMID:12582175; http://dx.doi.org/10.1074/jbc.M300586200

14. Chen R, Udenfriend S, Prince GM, Maxwell SE, Ramalingam S, Gerber LD, Knez J, Medof ME. A defect in glycosylphosphatidylinositol (GPI) transamidase activity in mutant K cells is responsible for their inability to display GPI surface proteins. Proc Natl Acad Sci U S A 1996; 93:2280-4; PMID:8637863; http://dx.doi.org/10.1073/pnas.93.6.2280

15. Maxwell SE, Ramalingam S, Gerber LD, Brink L, Udenfriend S. An active carbonyl formed during glycosylphosphatidylinositol addition to a protein is evidence of catalysis by a transamidase. J Biol Chem 1995; 270:19576-82; PMID:7642644; http://dx.doi.org/10.1074/jbc.270.33.19576

16. 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

17. Ramalingam S, Maxwell SE, Medof ME, Chen R, Gerber LD, Udenfriend S. COOH-terminal processing of nascent polypeptides by the glycosylphosphatidylinositol transamidase 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

18. Toh YK, Kamariah N, Maurer-Stroh S, Roessle M, Eisenhaber F, Adhikari S, Eisenhaber B, Grüber G. Structural insight into the glycosylphosphatidylinositol transamidase subunits PIG-K and PIG-S from yeast. J Struct Biol 2011; 173:271-81; PMID:21134462; http://dx.doi.org/10.1016/j.jsb.2010.11.026

19. Ooi HS, Kwo CY, Wildpaner M, Sirota FL, Eisenhaber B, Maurer-Stroh S, Wong WC, Schleiffer A, Eisenhaber F, Schneider G. ANNIE: integrated de novo protein sequence annotation. Nucleic Acids Res 2009; 37:W435-40; PMID:19389726; http://dx.doi.org/10.1093/nar/gkp254

20. 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.

21. Hildebrand A, Remmert M, Biegert A, Söding J. Fast and accurate automatic structure prediction with HHpred. Proteins 2009; 77(Suppl 9):128-32; PMID:19626712; http://dx.doi.org/10.1002/prot.22499

22. Koch B, Kolenko P, Buchholz M, Carrillo DR, Parthier C, Wermann M, Rahfeld JU, Reuter G, Schilling S, Stubbs MT, et al. Crystal structures of glutaminyl cyclases (QCs) 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/bi300687g

23. Greenblatt HM, Almog O, Maras B, Spungin-Bialik A, Barra D, Blumberg S, Shoham G. Streptomyces griseus aminopeptidase: X-ray crystallographic structure at 1.75 A resolution. J Mol Biol 1997; 265:620-36; PMID:9048953; http://dx.doi.org/10.1006/jmbi.1996.0729

24. Reiland V, Gilboa R, Spungin-Bialik A, Schomburg D, Shoham Y, Blumberg S, Shoham G. Binding of inhibitory aromatic amino acids to Streptomyces griseus aminopeptidase. Acta Crystallogr D Biol Crystallogr 2004; 60:1738-46; PMID:15388919; http://dx.doi.org/10.1107/S0907444904018281

25. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009; 4:363-71; PMID:19247286; http://dx.doi.org/10.1038/nprot.2009.2

26. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 1991; 9:56-68; PMID:2017436; http://dx.doi.org/10.1002/prot.340090107

27. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389-402; PMID:9254694; http://dx.doi.org/10.1093/nar/25.17.3389

28. Altschul SF, Gertz EM, Agarwala R, Schäffer AA, Yu YK. PSI-BLAST pseudocounts and the minimum description length principle. Nucleic Acids Res 2009; 37:815-24; PMID:19088134; http://dx.doi.org/10.1093/nar/gkn981

29. Saw WG, Eisenhaber B, Eisenhaber F, Grüber G. Low-resolution structure of the soluble domain GPAA1 (yGPAA170-247) of the glycosylphosphatidylinositol transamidase subunit GPAA1 from Saccharomyces cerevisiae. Biosci Rep 2013; 33:e00033; PMID:23458223; http://dx.doi.org/10.1042/BSR20120107

30. Ruiz-Carrillo D, Koch B, Parthier C, Wermann M, Dambe T, Buchholz M, Ludwig HH, Heiser U, Rahfeld JU, Stubbs MT, et al. Structures of glycosylated mammalian glutaminyl cyclases reveal conformational variability near the active center. Biochemistry 2011; 50:6280-8; PMID:21671571; http://dx.doi.org/10.1021/bi200249h

31. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 2014; 42:D503-9; PMID:24157837; http://dx.doi.org/10.1093/nar/gkt953

32. Koch B, Buchholz M, Wermann M, Heiser U, Schilling S, Demuth HU. Probing secondary glutaminyl cyclase (QC) inhibitor interactions applying an in silico-modeling/site-directed mutagenesis approach: implications for drug development. Chem Biol Drug Des 2012; 80:937-46; PMID:22967026; http://dx.doi.org/10.1111/cbdd.12046

33. Dudev T, Lin YL, Dudev M, Lim C. First-second shell interactions in metal binding sites in proteins: a PDB survey and DFT/CDM calculations. J Am Chem Soc 2003; 125:3168-80; PMID:12617685; http://dx.doi.org/10.1021/ja0209722

34. Yamashita MM, Wesson L, Eisenman G, Eisenberg D. Where metal ions bind in proteins. Proc Natl Acad Sci U S A 1990; 87:5648-52; PMID:2377604; http://dx.doi.org/10.1073/pnas.87.15.5648

35. Meyer U, Benghezal M, Imhof I, Conzelmann A. Active site determination of Gpi8p, a caspase-related enzyme required for glycosylphosphatidylinositol anchor addition to proteins. Biochemistry 2000; 39:3461-71; PMID:10727241; http://dx.doi.org/10.1021/bi992186o

36. Hamburger D, Egerton M, Riezman H. Yeast Gaa1p is required for attachment of a completed GPI anchor onto proteins. J Cell Biol 1995; 129:629-39; PMID:7730400; http://dx.doi.org/10.1083/jcb.129.3.629

37. Kodukula K, Maxwell SE, Udenfriend S. Processing of nascent proteins to glycosylphosphatidylinositol-anchored forms in cell-free systems. Methods Enzymol 1995; 250:536-47; PMID:7651176; http://dx.doi.org/10.1016/0076-6879(95)50095-2

38. Ohishi K, Inoue N, Kinoshita T. PIG-S and PIG-T, essential for GPI anchor attachment to proteins, form a complex with GAA1 and GPI8. EMBO J 2001; 20:4088-98; PMID:11483512; http://dx.doi.org/10.1093/emboj/20.15.4088

39. Ohishi K, Inoue N, Maeda Y, Takeda J, Riezman H, Kinoshita T. Gaa1p and gpi8p are components of a glycosylphosphatidylinositol (GPI) transamidase that mediates attachment of GPI to proteins. Mol Biol Cell 2000; 11:1523-33; PMID:10793132; http://dx.doi.org/10.1091/mbc.11.5.1523

40. Fraering P, Imhof I, Meyer U, Strub JM, van Dorsselaer A, Vionnet C, Conzelmann A. The GPI transamidase complex of Saccharomyces cerevisiae contains Gaa1p, Gpi8p, and Gpi16p. Mol Biol Cell 2001; 12:3295-306; PMID:11598210; http://dx.doi.org/10.1091/mbc.12.10.3295

41. Vainauskas S, Maeda Y, Kurniawan H, Kinoshita T, Menon AK. Structural requirements for the recruitment of Gaa1 into a functional glycosylphosphatidylinositol transamidase complex. J Biol Chem 2002; 277:30535-42; PMID:12052837; http://dx.doi.org/10.1074/jbc.M205402200

42. Hong Y, Maeda Y, Watanabe R, Inoue N, Ohishi K, Kinoshita T. Requirement of PIG-F and PIG-O for transferring phosphoethanolamine to the third mannose in glycosylphosphatidylinositol. J Biol Chem 2000; 275:20911-9; PMID:10781593; http://dx.doi.org/10.1074/jbc.M001913200

43. Shishioh N, Hong Y, Ohishi K, Ashida H, Maeda Y, Kinoshita T. GPI7 is the second partner of PIG-F and involved in modification of glycosylphosphatidylinositol. J Biol Chem 2005; 280:9728-34; PMID:15632136; http://dx.doi.org/10.1074/jbc.M413755200

44. Wu Z, Guo X, Gao J, Guo Z. Sortase A-mediated chemoenzymatic synthesis of complex glycosylphosphatidylinositol-anchored protein. Chem Commun (Camb) 2013; 49:11689-91; PMID:24195111; http://dx.doi.org/10.1039/c3cc47229a

45. Wong WC, Maurer-Stroh S, Schneider G, Eisenhaber F. Transmembrane helix: simple or complex. Nucleic Acids Res 2012; 40:W370-5; PMID:22564899; http://dx.doi.org/10.1093/nar/gks379

46. Wong WC, Maurer-Stroh S, Eisenhaber F. Not all transmembrane helices are born equal: Towards the extension of the sequence homology concept to membrane proteins. Biol Direct 2011; 6:57; PMID:22024092; http://dx.doi.org/10.1186/1745-6150-6-57

47. Wong WC, Maurer-Stroh S, Eisenhaber F. More than 1,001 problems with protein domain databases: transmembrane regions, signal peptides and the issue of sequence homology. PLoS Comput Biol 2010; 6:e1000867; PMID:20686689; http://dx.doi.org/10.1371/journal.pcbi.1000867

48. Vainauskas S, Menon AK. A conserved proline in the last transmembrane segment of Gaa1 is required for glycosylphosphatidylinositol (GPI) recognition by GPI transamidase. J Biol Chem 2004; 279:6540-5; PMID:14660601; http://dx.doi.org/10.1074/jbc.M312191200

49. Chen L, Yue Q, Zhang X, Xiang M, Wang C, Li S, Che Y, Ortiz-López FJ, Bills GF, Liu X, et al. Genomics-driven discovery of the pneumocandin biosynthetic gene cluster in the fungus Glarea lozoyensis. BMC Genomics 2013; 14:339; PMID:23688303; http://dx.doi.org/10.1186/1471-2164-14-339

50. Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, Yuan Y. Predicting function: from genes to genomes and back. J Mol Biol 1998; 283:707-25; PMID:9790834; http://dx.doi.org/10.1006/jmbi.1998.2144

51. Eisenhaber F. A decade after the first full human genome sequencing: when will we understand our own genome? J Bioinform Comput Biol 2012; 10:1271001; PMID:22849370; http://dx.doi.org/10.1142/S0219720012710011

52. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004; 32:1792-7; PMID:15034147; http://dx.doi.org/10.1093/nar/gkh340

53. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009; 25:1189-91; PMID:19151095; http://dx.doi.org/10.1093/bioinformatics/btp033

54. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. Multiple sequence alignment with Clustal X. Trends Biochem Sci 1998; 23:403-5; PMID:9810230; http://dx.doi.org/10.1016/S0968-0004(98)01285-7