Study of molecular recognition mechanism in protein GPI modification: a bioinformatics analysis of interaction between GPI-anchored proteins and modification enzyme

Yuri MUKAI*, Daiki TAKAHASHI*, Tsubasa OGAWA*, Kota HAMADA* and Kenji ETCHUYA*

* Department of Electronics, Graduate School of Science and Technology, Meiji University
1-1-1 Higashi-mita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan
E-mail: yuri@meiji.ac.jp

Received 29 June 2015

Abstract
The interaction for protein glycosylphosphatidylinositol (GPI) modification between the premature GPI-anchored proteins and the active sites in PIG-K, one of the GPI transamidase proteins, is discussed in this study by the homology modeling method and amino acid propensity analysis in the space around the ω-sites in premature GPI-anchored proteins. In particular, the direct interaction between ω-sites of GPI-anchored proteins and PIG-K was focused on, the root-mean-square deviation (RMSD) and three-dimensional amino acid propensity around the ω-sites were analyzed in the present study. As the results, PIG-K was considered to recognize the specific structure around the ω-sites of the GPI-anchored proteins, the positively-charged Lys (K) and Arg (R) residues around the ω-sites have the possibility to interact with the negatively-charged Asp (D) and Glu (E) residues around an active site of PIG-K, and Tyr (Y) and Ala (A) residues around the ω-sites are thought to be essential for molecular recognition by PIG-K. The findings in this study that structural recognition around the ω-site in the mature GPI-anchored proteins by PIG-K are useful for understanding the local mechanism of the GPI modification enzyme and can be applied for the development of cell-surface-engineering.

Key words: Bioinformatics, Post-translational modification, GPI-anchored protein, PIG-K, Mammalian protein

1. Introduction

Many membrane proteins exist on the surface of Eukaryotic cells. Glycosylphosphatidylinositol (GPI)-anchored proteins are known as membrane proteins which have no transmembrane region. GPI is a glycolipid molecule closely related to protein post-translational modification which anchors soluble proteins to the plasma membrane (Udenfriends et al., 1995; Udenfriends et al., 1995). GPI-anchored proteins exist widely throughout Eukaryotic cells as enzymes, receptors and immune proteins. Plant GPI-anchor proteins also play essential roles as elicitor receptors (Fliegmann et al., 2011; Kaku et al., 2006). The malfunction of GPI modification in human cells is well known to have correlations to serious disorders, including bovine spongiform encephalopathy (BSE) (Donne et al., 1997; James et al., 1997; Lopez-Garcia et al., 2000; Riek et al., 1996; Riek et al., 1997; Zahn et al., 1997), Creutzfeld-Jacob disease (CJD) (Cashman et al., 1990), paroxysmal nocturnal hemoglobinuria (PNH) (Nagarajan et al., 1995), and deep vein thrombosis (DVT) (Davies et al., 1989; Holguin et al., 1990; Davitz et al., 1986; Hansch et al., 1988). Thus, the understanding of the GPI modification mechanism is believed to be crucial for the medical treatment of those disorders and the accurate development of protein engineering.

GPI-anchored proteins have a hydrophobic signal peptide at the N-terminus which is essential for the localization of the protein to the endoplasmic reticulum (ER). First, premature GPI-anchored proteins are localized into the ER lumen after completing biosynthesis by translocon. In the ER lumen, the GPI attachment signal, located at the C-terminus of the premature GPI-anchored protein, is recognized by the GPI modification enzyme, GPI transamidase.
Simultaneously, GPI transamidase cleaves the GPI attachment signal and modifies GPI molecules at the new C-terminus referred to as the ω-site. Once this is complete, mature GPI-anchored proteins with GPIs are translocated to the Golgi apparatus by transport vesicles and finally reach the raft region in the plasma membrane. As the result, GPI-anchored proteins become peripheral membrane proteins with the functional domain toward the extracellular side.

GPI transamidase, the GPI modification enzyme, is a complex consisting of five proteins (Hong et al., 2003). PIG-K, one of the proteins included in GPI transamidase, is known to cleave premature GPI proteins (Eisenhaber et al., 2003). However, the way in which premature GPI-anchored proteins are recognized by the molecular mechanism of PIG-K has not been clarified yet, because PIG-K is an intrinsic membrane protein which has a helix spanning the ER membrane and is difficult to resolve the crystal structure by X-ray diffraction or NMR. Though the computational methods for detecting GPI-anchored protein sequences and the fact that GPI transamidase recognizes common features around the ω-sites of the premature GPI-anchored proteins were pointed out in previous studies (Eisenhaber et al., 1998, Eisenhaber et al., 2004, Mukai et al., 2012, Mukai et al., 2013), the interaction between PIG-K and premature GPI-anchored proteins was not discussed.

In this study, the amino acid residues of GPI-anchored proteins interacting with PIG-K were identified by amino acid propensity analysis in the space around the ω-sites in premature GPI-anchored proteins. The reaction sub-site and the specificity sub-sites of PIG-K for recognizing premature GPI-anchored proteins were discussed using the functional domain structure of PIG-K created by the homology modeling method. The structure recognition around the ω-site in GPI-anchored proteins by PIG-K was pointed out through the comparison analysis of the structures around the ω-site.

2. Materials and methods

Fifty-three entries of mammalian GPI-anchored proteins were obtained from the UniProt Knowledgebase/Swiss-Prot protein sequence database release 2014_01 (Bairoch and Apweiler, 1999) by searching with the keywords “mammalia” in the OC lines and “GPI” in the FT LIPID lines. The entries that had “fragment” in the DE lines and “potential” or “probable” annotations in the FT LIPID lines were excluded from the datasets. Nine of the GPI-anchored protein entries which had crystal structural information in the DR PDB lines, the linkage of the Protein Data Bank (PDB) release 2014_01 (Berman et al., 2000; Berman et al., 2007; Bourne et al, 2004), were extracted from the above sequence dataset. Four of the GPI-anchored protein entries had no PDB structures but had a model structure of which sequence identity was more than 60% defined by the HHpred (Söding et al., 2005), a homology modelling tool.

Table 1   Structural information of GPI-anchored proteins from the Protein Data Bank or the homology modelling method by HHpred

| Swiss-Prot ID | protein name                  | PDB-ID   | Structure data type            |
|---------------|--------------------------------|----------|-------------------------------|
| CAH4_HUMAN    | human carbonic anhydrase IV    | 1ZNC     | X-ray diffraction (2.8 Å)     |
| 5NTD_HUMAN    | human 5'-nucleotidase          | 4H1S     | X-ray diffraction (2.2 Å)     |
| CAH4_MOUSE    | murine carbonic anhydrase IV   | 2ZNC     | X-ray diffraction (2.8 Å)     |
| DPEP1_HUMAN   | human renal dipeptidase        | 1ITQ     | X-ray diffraction (2.3 Å)     |
| FOLR1_HUMAN   | human folate receptor alpha    | 4LRH     | X-ray diffraction (2.8 Å)     |
| CBPM_HUMAN    | human carboxypeptidase M       | 1UWY     | X-ray diffraction (3.0 Å)     |
| FOLR2_HUMAN   | human folate receptor beta     | 4KMY     | X-ray diffraction (1.8 Å)     |
| ALL1_HORSE    | horse major allergen EQU C 1   | 1EW3     | X-ray diffraction (2.3 Å)     |
| PRO_MOUSE     | mouse prion protein            | 4MA7     | X-ray diffraction (2.0 Å)     |
| 5NTD_MOUSE    | mouse 5'-nucleotidase          | template: 4H2I | Modelling                  |
| CD59_RAT      | rat glycoprotein               | template: 2J8B  | Modelling                  |
| CD59_PONAB    | orangutan glycoprotein         | template: 1CDQ  | Modelling                  |
| DPEP1_MOUSE   | mouse dipeptidase 1            | template: 1ITU  | Modelling                  |
| XPP2_PIG      | pig Xaa-Pro aminopeptidase 2   | template: 3CTZ  | Modelling                  |

Human PIG-K sequence data (Swiss-Prot ID: GPI8_HUMAN) was also extracted from the UniProt Knowledgebase/Swiss-Prot, and its model structure was also created by the HHpred homology modelling tool using the template structure (PDB-ID: 4AW9). Structural images of the PIG-K model structure and the GPI-anchored proteins were displayed by the PyMOL Molecular Graphics System.
The root-mean-square deviation (RMSD) was calculated for the quantitative evaluation of the deviation between partial structures of GPI-anchored proteins around the ω-sites and thirteen other partial structures in Table 1. The calculation areas were within a unit ball in a 10, 20 and 30 Å radiuses from the Cα atom in the ω-site using Equation 1. In this study, the average distance of two Cα atoms \((a_i, b_i; i=2)\) between two structures \((N=2)\), protein A and protein B, was estimated.

\[
RMSD(A, B) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (a_i - b_i)^2}
\]  

(1)

The three-dimensional amino acid propensities \((f_j)\) of 14 proteins in Table 1 were calculated using Equation 2, based on general calculation for the position-specific scoring matrix (PSSM) which can discriminate the features of the structures and functions of proteins (Mukai et al., 2011; Mukai et al., 2013; Nanni et al., 2014; Etchuya et al., 2015).

\[
f_j = \frac{n_j}{\sum_{j=1}^{20} n_j}
\]  

(2)

The frequency \((n_j)\) signifies the total occurrences of the amino acid \(j\) of 14 proteins. The amino acid propensity around the ω-site in GPI-anchored proteins was estimated using the coordinates of Cα atoms of each amino acid within each unit ball. By choosing points at a 5, 10, 15, 20, 25 and 30 Å distance from the ω-site, the calculation areas were determined.

3. Results and Discussion

To understand the interaction between PIG-K and GPI-anchored proteins, each structure was compared around the ω-site of GPI-anchored proteins using RMSD calculation. In the RMSD for protein structural comparison, the deviation between two protein structures was able to be estimated quantitatively. The RMSD was calculated by the partial structures of two GPI-anchored proteins within a unit ball in a 10 Å radius from the Cα atom in the ω-site using Cα atoms of all amino acids in this region by the averaging of all possible regressions. The RMSDs within a unit ball in a 20 and 30 Å radius were also calculated (Table 2). The structure deviation increased depending on the distance from the ω-site, it can be hypothesized that the specific structure of GPI-anchored proteins are recognized by PIG-K based on the results.

| radius of unit ball | average RMSD |
|---------------------|--------------|
| 10 Å                | 1.38         |
| 20 Å                | 5.11         |
| 30 Å                | 7.45         |

Though the features of the sequences around ω-sites in previous studies for the sequence analysis of GPI-anchored proteins was pointed out (Eisenhaber et al., 1998, Eisenhaber et al., 2004, Mukai et al., 2012, Mukai et al., 2013), it was not clarified how the structure is recognized by PIG-K. Therefore in this study, to clarify the factors around the ω-sites of GPI-anchored proteins which are recognized by PIG-K, the amino acid propensities within a unit ball in a 5 and 20 Å radius from the ω-sites were compared (Fig. 1). The intermolecular forces such as the disulfide bond, salt bridge, hydrogen bond and the Van der Waals force usually give the interaction within 5 Å. The long-range electrostatic interaction influencing around 15 Å is thought to not be reached to 20 Å.

In Fig. 1, the propensities of Tyr (Y), Ala (A) and Leu (L) residues are higher than other amino acids within a unit ball in a 5 Å radius from the ω-site. Though the propensities of Leu within a 5 Å radius are not different from those of within 20 Å radius, the propensities of Tyr and Ala within 5 Å increased two times than those of within a 20 Å radius. Tyr, Trp (W) and Phe (F) are all known as aromatic residues. Phe residues were found in a 5 and 20 Å radius, though Trp appeared only within a 5 Å radius. Also Tyr, Ser (S) and Thr (T) are all known as phosphorylable residues, however, the propensities of Ser and Thr were higher within a 20 Å radius than those of within a 5 Å radius. These results indicate that Tyr is a specific aromatic/ phosphorylable residue in the region close to the ω-sites. Residues such as Ala and Gly (G) are both considered to be small and hydrophobic. Ala and Gly appear in the case that enzyme reaction
should be protected from the sterific hindrances, and the hydrophobic environment is required around the target site. Ala and Gly are often found in the enzyme which binding should be protected from the steric hindrance and the hydrophobic environment is required around the target site. The propensity of Ala as higher at a 5 Å radius than at a 20 Å radius whereas the propensity of Gly at a 5 Å radius was lower than at a 20 Å radius. Because the propensity of Ala is higher near the ω-site, Ala is thought to create a hydrophobic environment which has no steric hindrance in this area. In contrast, Gly is not adequate to stabilize conformation in this region due to its lower propensity.

![Fig. 1. Three-dimensional amino acid propensity of the partial structures in the GPI-anchored proteins within a unit ball in a 5 (black bar) and 20 (white bar) Å radius from the Cα atom in the ω-site](image)

Among these residues, Tyr and Ala have the highest levels of propensity followed by Asp (D), a positively charged residue, and Lys (K) a negatively charged residue. The propensities of other charged residues such as Glu (E) and Arg (R) are not higher at a 5 Å radius than at a 20 Å radius. Though the propensity of His (H) was lower than those of other residues, His existed in the region close to the ω-sites. Asp, Lys and His are thought to interact with the specificity sub-site of PIG-K in the region close to the ω-site. In conclusion, these findings indicate that residues including Tyr, Ala, Asp, Glu and His are related to binding with the active-site of PIG-K.

![Fig. 2 Change of the three-dimensional amino acid propensity around the ω-site of the GPI-anchored proteins using calculation ranges within a unit ball from 5 to 30 Å along the radius from the Cα atom in the ω-site](image)

In Fig. 2, the propensity changes, which depend on the distance from the ω-sites of the main residues related to the short range force including hydrophobic and electrostatic interaction, were observed. The propensity was higher at a 10 and 15 Å radius in the positively-charged residues (K and R) which are thought to be related to the interaction with PIG-K. In contrast, the propensity of the negatively-charged residues (D and E) did not vary in relation to the distance from the ω-sites. According to the PIG-K model structure created by the HHpred, the negatively-charged residues (D and E) exist within 5 Å from an active site, Cys-189. The positively-charged residues (K and R) near the ω-sites of GPI-anchored proteins are considered to contact to these Asp and Glu residues around the active site of PIG-K. As the
propensity of Tyr decreased in inverse proportion to the distance from the $\omega$-sites, Tyr is considered to exist extremely close to the $\omega$-sites of the GPI-anchored proteins. The $\omega$-site of GPI-anchored proteins interacts with the reaction sub-site of PIG-K, and the position and angle of the interaction are able to be controlled by Tyr, one of the specificity sub-sites of PIG-K.

Fig. 3 (A) shows the three dimensional structure around the $\omega$-site of human carbonic anhydrase protein IV (PDB-ID: 1ZNC), a typical GPI-anchored protein. Ser-284 ($\omega$-site), which interacts with a reaction sub-site and Tyr-217, which is thought to interact with a specificity sub-site of PIG-K, are close to each other. The model three-dimensional structure around the active site of PIG-K created by the HHpred is shown in Fig. 3 (B), and Cys-189 (reaction sub-site), Tyr-217 are also close to each other in the structure. As the specificity sub-site which controls the position and angle of GPI-anchored proteins, the Tyr-217 of PIG-K is thought to play an important role in the cleaving of the GPI-protein’s $\omega$-site by the Cys-189 of PIG-K. The Tyr of PIG-K could interact with the Tyr of GPI-anchored proteins by hydrophobic force. In conclusion, the Tyr residues exist extremely close to both the reaction sub-site of PIG-K and the $\omega$-site of GPI-anchored proteins according to the structural images determined by X-ray diffraction.

The findings of this study are summarized as follows: PIG-K recognizes the tertiary structures of GPI-anchored proteins, and the characteristics of the amino acids existing around the $\omega$-site are similar to each GPI-anchored protein. Based on these findings, a three-dimensional PSSM calculated by spatial position-specific amino acid propensity can be applied for the development of GPI-anchored protein prediction methods and be useful to improve prediction accuracy. If the structural analysis of premature GPI-anchored proteins progresses in the future, the molecular interaction between GPI-anchored proteins and PIG-K can be understood more exactly.

Fig. 3 (A) Structure around the $\omega$-site of GPI-anchored protein (PDB-ID: 1ZNC powered by PyMOL), (B) PIG-K model structure created using 4AW9 template structure around the active site

Acknowledgement

The authors are grateful to Dr. Adam Seth Magdalena for providing useful comments and proofreading. This work was financially supported in part by Grant-in-Aid for Young Scientists (B), Grant-in-Aid for Scientific Research (C) (KAKENHI) and the Collaborative Research Project of the Institute of Fluid Science, Tohoku University.

References

Bairoch, A., Apweiler, R., The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999, Nucl. Acid. Res., Vol. 27, No.1 (1999), pp. 49-54.
Berman, H., Henrick, K., Nakamura, H., Markley, J. L., The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data, Nucl. Acid. Res., Vol. 35, (2007), pp. D301-D303.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E., The Protein Data Bank, Nucl. Acid. Res., Vol. 28, No. 1 (2000), pp. 235-242.
Bourne P. E., Addess K. J., Bluhm W. F., Chen L., Deshpande N., Feng Z., Fleri W., Green R., Merino-Ott J. C.,
Townsend-Merino W., et al., The distribution and query systems of the RCSB protein data bank, Nucleic Acid Res., Vol. 32, (2004), pp. D223-D225.

Cashman, N. R., Loertscher, R., Nalbantoglu, J., Shaw, I., Kascak, R. J., Bolton, D. C., Bendheim, P. E., Cellular isoform of the scrapie agent protein participates in lymphocyte activation, Cell, Vol. 61 (1990), pp. 185-192.

Davies, A., Simmons, D. L., Hale, G., Harrison, R. A., Tighe, H., Lachmann, P. J., Waldmann, H., CD59, an LY-6-like protein expressed in human lymphoid cells, regulates the action of the complement membrane attack complex on homologous cells, J. Exp. Med., Vol. 170 (1989), pp. 637-654.

Davitz, M. A., Low, M. G., Nussenzweig, V., Decay-accelerating factor (DAF) shares a common carbohydrate determinant with the variant surface glycoprotein (VSG) of the African Trypanosoma brucei, J. Exp. Med., Vol. 163 (1986), pp. 1150-1161.

Donne, D., Viles, J. H., Groth, D., Mehlhorn, I., James, T. L., Cohen, F. E., Prusiner, S. B., Wright, P. E., Dyson, H. J., Structure of the recombinant full-length hamster prion protein PrP(29-231): the N terminus is highly flexible, Proc. Natl. Acad. Sci. USA, Vol. 94 (1997), pp. 13452-13457.

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., Vol. 12 (1998), pp. 1155-1161.

Eisenhaber, B., Eisenhaber, F., Maurer-Stroh, S., Neuberger, G., Prediction of sequence signals for lipid post-translational modifications: insights from case studies, Proteomics, Vol. 4, No. 6 (2004), pp. 1614-1625.

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, Vol. 25, No. 4 (2003) pp. 367-385.

Etchuya, K., Mukai, Y., Structural characteristics around O-glycosylation sites in mammalian proteins, J. Biomech. Sci. Eng., Vol. 10, No. 1 (2015), pp. 14-00249: 1-6.

Fliegmann, J., Uhlenbroich, S., Shinya, T., Martinez, Y., Lefebvre, B., Shibuya, N., Bono, J. J., Biochemical and phylogenetic analysis of CEBP-like LysM domain-containing extracellular proteins in higher plants, Plant Physiol. Biochem., Vol. 49, No. 7 (2011), pp. 709-720.

Hansch, G. M., Weller, P. F., Nicholson-Weller, A., Release of C8 binding protein (C8bp) from the cell membrane by phosphatidylinositol-specific phospholipase C, Blood, Vol. 72 (1988), pp. 1089-1092.

Hong, Y., Ohishi, K., Kang, J. Y., 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., Vol. 14, No. 5 (2003), pp. 1780-1789.

Holguin, M. H., Wilcox, L. A., Bernshaw, N. J., Rosse, W. F., Parker, C. J., Erythrocyte membrane inhibitor of reactive lysis: effects of phosphatidylinositol-specific phospholipase C on the isolated and cell-associated protein, Blood, Vol. 75 (1990), pp. 284-289.

James, T. L., Liu, H., Ulyanov, N. B., Farr-Jones, S., Zhang, H., Donne, D. G., Kaneko, K., Groth, D., Mehlhorn, I., Prusiner, S. B., Cohen, F. B., Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform, Proc. Natl. Acad. Sci. USA, Vol. 94 (1997), pp. 10086-10091.

Kaku, H., Nishizawa, Y., Ishii-Minami, N., Akimoto-Tomiyama, C., Dohmae, N., Takio, K., Minami, E., Shibuya, N., Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor, Proc. Natl. Acad. Sci. USA., Vol. 103, No. 29 (2006), pp. 11086-11091.

Lopez-Garcia, F., Zahn, R., Rick, R., Wuthrich, K., NMR structure of the bovine prion protein, Proc. Natl. Acad. Sci. USA, Vol. 97, (2000), pp. 8334-8339.

Mukai Y., Ikeda, M., Tanaka, H., Konishi, T., Oura, O., Sasaki, T., Discrimination of mammalian GPI-anchored by hydrophathy and amino acid propensities, Biosci. Biotechnol. Biochem., Vol. 77, No. 3 (2013), pp. 526-533.

Mukai, Y., Tanaka, H., Yoshizawa, M., Oura, O., Sasaki, T., Ikeda, M., A computational identification method for GPI-anchored proteins by artificial neural network, Current Bioinformatics, Vol. 7, No. 2 (2012), pp. 125-131.

Nagarajan, S., Brodsky, R. A., Young, N. S., Medof, M. E., Genetic defects underlying paroxysmal nocturnal hemoglobinuria that arises out of aplastic anemia, Blood, Vol. 86 (1995), pp. 4656-4661.
into general Chou's pseudo amino acid composition, J. Theor. Biol., Vol. 360, (2014), pp. 109-116.
Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R., Wuthrich, K., NMR structure of the mouse prion protein domain PrP (121-231), Nature, Vol. 382 (1996), pp. 180-183.
Riek, R., Hornemann, S., Wider, G., Glockshuber, R., Wuthrich, K., NMR characterization of the full-length recombinant murine prion protein, mPrP(23-231), FEBS Lett., Vol. 413 (1997), ppp. 282-288.
Söding, J., Biegert, A., Lupas, A. N., The HHpred interactive server for protein homology detection and structure prediction, Nucleic Acids Res., Vol. 33 (2005), pp. W244-W248.
Udenfriend, S., Kodukula, K., How glycosylphosphatidylinositol-anchored membrane proteins are made, Annu. Rev. Biochem., Vol. 64 (1995), pp. 405-420.
Udenfriend, S., Kodukula, K., Prediction of omega site in nascent precursor of glycosylphosphatidylinositol protein, Methods Enzymol., Vol. 250 (1995), pp. 571-572.
Zahn, R., Liu, A., Luhrs, T., Riek, R., von Schroetter, C., Lopez-Garcia, F., Billeter, M., Calzolai, L., Wider, G., Wuthrich, K., NMR solution structure of the human prion protein, Proc. Natl. Acad. Sci. USA, Vol. 97 (2000), pp. 145-150.