Defining the boundaries of species specificity for the *Saccharomyces cerevisiae* glycosylphosphatidylinositol transamidase using a quantitative *in vivo* assay

Rachel MORISSETTE*1, Yug VARMA*†2 and Tamara L. HENDRICKSON*†3

*Department of Chemistry, Remsen Hall, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, U.S.A., and †Department of Chemistry, Wayne State University, 5101 Cass Avenue, Detroit, MI 48202, U.S.A.

Synopsis

In eukaryotes, GPI (glycosylphosphatidylinositol) lipid anchoring of proteins is an abundant post-translational modification. The attachment of the GPI anchor is mediated by GPI-T (GPI transamidase), a multimeric, membrane-bound enzyme located in the ER (endoplasmic reticulum). Upon modification, GPI-anchored proteins enter the secretory pathway and ultimately become tethered to the cell surface by association with the plasma membrane and, in yeast, by covalent attachment to the outer glucan layer. This work demonstrates a novel *in vivo* assay for GPI-T. *Saccharomyces cerevisiae* INV (invertase), a soluble secreted protein, was converted into a substrate for GPI-T by appending the C-terminal 21 amino acid GPI-T signal sequence from the *S. cerevisiae* Yapsin 2 [Mkc7p (Y21)] on to the C-terminus of INV. Using a colorimetric assay and biochemical partitioning, extracellular presentation of GPI-anchored INV was shown. Two human GPI-T signal sequences were also tested and each showed diminished extracellular INV activity, consistent with lower levels of GPI anchoring and species specificity. Human/fungal chimaeric signal sequences identified a small region of five amino acids that was predominantly responsible for this species specificity.

Key words: glycosylphosphatidylinositol anchoring, glycosylphosphatidylinositol transamidase, invertase, *Saccharomyces cerevisiae*, species specificity


INTRODUCTION

Lipidation of proteins is often used to properly target specific proteins and enzymes to their correct locales *in vivo* [1,2]. In eukaryotes, GPI (glycosylphosphatidylinositol) membrane anchoring of proteins is a common example of protein lipidation. In fact, computational algorithms predict that 0.5% of all eukaryotic proteins are C-terminally modified to contain a GPI anchor [3,4]. The fate of GPI-anchored proteins is to enter the secretory pathway to ultimately become tethered to the plasma membrane via the lipid portion of the GPI anchor [2,5]. In yeast, further modifications can occur where specific proteins lose most of their GPI anchor and become covalently attached to the outer glucan layer of the fungal cell wall [6–8]. Since GPI-anchored proteins are specifically localized to the outer membrane of eukaryotic cells, these proteins often serve important extracellular functions, ranging from cell wall biosynthesis to cell adhesion and morphogenesis [2,8,9].

The conserved GPI anchor core is comprised of a diacyl lipid, a *myo*-inositol, a single glucosamine, three mannose residues and a terminal ethanolamine phosphate. Diversification of this core structure is common and has been reviewed previously [10]. Attachment of this anchor to proteins is catalysed by GPI-T (GPI transamidase), a multimeric, membrane-bound enzyme in the ER (endoplasmic reticulum) [10–26]. All substrates for GPI-T...
contain an N-terminal signal sequence to direct them into the ER prior to modification by GPI-T. GPI-T identifies its protein substrates via a C-terminal signal sequence; features of this sequence target anchor attachment to an internal amide bond (near the C-terminus) at a position that has been termed the \( \omega \)-site. In the net reaction, the GPI-T signal sequence is released and transamidatively replaced by the GPI anchor, with an amide bond formed between the protein’s new C-terminus (the \( \omega \)-site residue) and the terminal amine of the GPI anchor [16, 27–30].

The GPI-T signal sequence lacks a specific consensus sequence and is simply comprised of small amino acids at the \( \omega \) and \( \omega + 2 \) positions (with numbering increasing towards the C-terminus), followed by a short hydrophilic sequence and a longer hydrophobic sequence (reviewed by Orlean and Menon [10]). Figure 1 represents exemplary examples of three different signal sequences for GPI-T: the first (abbreviated Y21) is from the \textit{Saccharomyces cerevisiae} Yapsin 2 protease (Mkc7p) [31], the second (abbreviated UP30) is from the human uPAR (urokinase-type plasminogen-activated receptor) [32], and the third (abbreviated CA25) is from the human Campath-1 antigen, the smallest known GPI-anchored protein [33].

Despite these straightforward signal sequence features, there is clear evidence that GPI-T more favourably modifies some substrate proteins over others. For example, the wild-type GPI-T signal sequence in a miniaturized version of placental alkaline phosphatase (miniPLAP, a well-characterized GPI protein construct) was replaced with nine different human GPI-T signal sequences (one from mammals, one from yeast and a putative plant sequence). The plant and yeast signal sequences were cleaved and replaced by GPI anchors to similar extents, indicative of plant sequence. The plant and yeast signal sequences were cleaved and replaced by GPI anchors to similar extents, indicative of plant sequence.

GPI-T participates in governing the cell surface concentration of anchor-attached proteins in yeast. INV catalyses the conversion of sucrose into fructose and glucose and its activity is readily and rapidly quantifiable using commercially available colorimetric assays for glucose [39]. In wild-type yeast strains, INV is expressed in both \( \alpha \) and \( \beta \) fructofuranoside fructohydrolase), encoded by the \textit{SUC2} gene, offers an interesting starting point for the development of high-throughput methods to quantify surface presentation of GPI-anchored proteins in yeast. INV catalyses the conversion of sucrose into fructose and glucose and its activity is readily and rapidly quantifiable using commercially available colorimetric assays for glucose [39]. In wild-type yeast strains, INV is translated in a soluble, cytoplasmic form and as a highly secreted enzyme. Using strains with a \textit{SUC2} \( ^{-} \) background, INV has been used extensively as a tool to study protein translocation pathways, including secretion and targeting to different cellular compartments [40, 41]. In the present work, the \textit{SUC2} gene was modified to append different yeast and human GPI-T signal sequences on in GPI signal sequence recognition arose from experiments that showed that the artificial signal sequence Ser\(^{-}\)Thr\(^{-}\)Leu\(^{+} \) converted CD46, a type I membrane protein that is not naturally GPI-anchored, into a robust substrate for GPI-T (\( \sim 80 \% \) conversion) [35].

Some evidence of species specificity among different GPI-T orthologues has also been reported. For example, early work demonstrated that expression of the \textit{Trypanosoma brucei} VSG (variant surface glycoprotein) in COS cells led to protein expression but only low levels of GPI anchor attachment; this defect was rescued when the VSG C-terminal GPI-T signal sequence was replaced with the human decay accelerating factor signal sequence [36]. Later, the \textit{Clostridium thermocellum} endoglucanase E was expressed in \textit{Nicotiana tabacum} with three different C-terminal signal sequences (one from mammals, one from yeast and a putative plant sequence). The plant and yeast signal sequences were cleaved and replaced by GPI anchors to similar extents, indicating that \textit{N. tabacum} GPI-T recognized both of these peptide sequences as substrates. In contrast, the use of a human GPI-T signal sequence significantly reduced the level of GPI anchoring [37]. Calculations and extensive comparisons within and between taxons have also pointed to subtle signal sequence variations in different organisms [3, 4].

The number of GPI signal sequences that have been evaluated between heterologous systems is still small, therefore further experiments are necessary to determine whether or not species-specific recognition of GPI-T protein substrates is a universal phenomenon and to what degree variations amongst species can be expected. One of the challenges faced in evaluating something like the species specificity of GPI-T is that the methods to observe and confirm GPI-anchoring of a protein are cumbersome and most commonly involve separation of soluble and membrane fractions, immunoprecipitation and analysis by SDS/PAGE [10]. Although not directly addressing species specificity, other techniques have been used to quantify the extent of GPI anchoring of a given protein. One experiment took advantage of a chromogenic assay for \( \alpha \)Gal (\( \alpha \)-galactosidase). A series of putative yeast GPI signal sequences were appended on to the C-terminus of \( \alpha \)Gal and whole cells were assayed for cell surface \( \alpha \)Gal activity, leading to the identification of new fungal GPI-T substrates [38]. Another report used flow cytometry to quantify cell surface expression of CD46 upon addition of synthetic GPI-T signal sequences [35].
to INV. Analysis of these substrates for GPI-T demonstrates that INV can be used as a reporter enzyme in in vivo assays to quantify GPI-T activity and species specificity.

**EXPERIMENTAL**

**Materials**

Most materials were purchased from Sigma and used without further purification, unless indicated otherwise. Oligonucleotides were purchased from Invitrogen. Plasmids were purified using plasmid isolation kits from Promega. PCR products and plasmids were gel purified using QIAGEN gel extraction kits. DNA sequencing was performed at the Johns Hopkins University School of Medicine Synthesis and Sequencing Facility. The entire gene insert was completely sequenced for each plasmid construct. The plasmid pCYI-20 [42] and S. cerevisiae strain SEY6210 were provided by Professor Beverly Wendland (Johns Hopkins University, Baltimore, MD, U.S.A.).

**Plasmid construction and yeast transformation**

Three PCR reactions were used to ultimately introduce a SphI site into pCYI-20 immediately upstream of the stop codon. First, primers RM-A-F3 and RM-A-R1 were used with pCYI-20 as the template to amplify an internal XbaI site while introducing the SphI site at the desired position. Similarly, primers RM-B-F4 and RM-B-R5 amplified a 3’ PvuII site while inserting the same SphI site. These two PCR products were reamplified in a self-templating reaction to generate the desired insert. The insert was gel purified and inserted into pCR2.1 TOPO (Invitrogen), according to the manufacturer’s instructions. The fragment was then sub-cloned into the XbaI and PvuII sites of pCYI-20, to yield pJLM021.

Primers RM-FLAG-F1 and RM-FLAG-R1 (purchased with 5’-phosphorylation) were directly ligated into the SphI site of pJLM021 to introduce a FLAG tag at the C-terminus of the SUC2 gene product (pRM028); these primers were designed to ablate the first SphI site upon ligation. Two different strategies were used to introduce different GPI signal sequences into pRM028. For the insert encoding the UP30 C-terminal GPI signal sequence (Figure 1), primers RM-30mer-F5 and RM-30mer-R5 were amplified without added template using Pfu Turbo polymerase. The resultant PCR product was gel purified, digested with SphI and ligated into the same site in pRM028. The remaining constructs, encoding for shorter GPI signal sequences, were generated by ligating complementary, phosphorylated primers directly into the SphI site of pRM028. All primers were designed for optimal S. cerevisiae codon usage to facilitate expression (see Supplementary Table S1 at http://www.bioscirep.org/bsr/032/bsr0320577add.htm for the primer sequences).

All INV-encoding plasmids were transformed into the strain SEY6210 using standard lithium acetate transformation protocols (see Supplementary Table S2 at http://www.bioscirep.org/bsr/032/bsr0320577add.htm for plasmid names and insert sequences).

**INV assays**

Each S. cerevisiae strain, containing a different INV-encoding plasmid, was grown at 30 °C in minimal medium (3.35 g yeast nitrogen base, 1 g of dropout mix minus uracil (US Biological) and 1 % fructose (w/v) in 500 ml of water) for approx. 36 h. Secreted INV activity (extracellular or plasma membrane-associated) was quantified by a variation of a previously reported method [39]. Briefly, each assay was conducted using 1.0 absorbance unit of cells. In a typical assay, the appropriate volume of cell culture was combined with 20 μl of 1 M NaOAc, pH 4.9 and diluted to a final volume of 200 μl with water. INV assays were conducted at 30 °C and were initiated with the addition of 50 μl of 0.5 M sucrose (Sigma) that had been prepared in 0.1 M NaOAc, pH 4.9 and pre-incubated at 30 °C for at least 30 min. (This pre-treatment was essential to eliminate anomalous assay results that arose, even in the absence of cells, apparently from a contamination in the commercial sucrose. Results not shown.)

Timepoints (50 μl), ranging from 1 to 16 min incubation times, were quenched in 0.2 M K2HPO4, pH 10.0 (75 μl) and immediately boiled for 3 min.

The amount of liberated glucose was quantified using a reaction mixture containing glucose oxidase (500 units), HRP (horseradish peroxidase; 100 purpargalin units), NEM (N-ethylmaleimide; 100 μM) and o-dianisidine (0.1 mg/ml) in 0.1 M K2HPO4, pH 7.0 [39]. Aliquots of this solution (0.5 ml) were added to each quenched timepoint and the reaction mixtures were incubated at 30 °C for exactly 5 min. Each timepoint was quenched by the addition of 0.5 ml of 6 M HCl. The amount of oxidized o-dianisidine was quantified by measuring the absorbance of these solutions at 530 nm. All assay data were corrected for no sucrose controls and normalized according to absorbance at 700 nm. All experiments were conducted in triplicate from separate cell cultures and error bars represent standard error.

**Localization of INV activity to the plasma membrane**

In assays where the amount of secreted versus membrane-associated INV activity was compared, 1.0 absorbance unit of cells was combined with 20 μl of 1 M NaOAc, pH 4.9 and diluted to a final volume of 200 μl in water. Cells were gently pelleted and the medium collected and spun. The cell pellets were washed twice with 0.1 M NaOAc, pH 4.9 and resuspended in 200 μl of 0.1 M NaOAc, pH 4.9 buffer. The cell pellet suspension and medium were both assayed as above to quantify the amount of membrane-bound compared with secreted INV.

The protein expressed by the Y21-modified INV construct (pRM031) was further examined to confirm that it was modified by GPI-T to contain a C-terminal GPI anchor. Cells (200 ml) were grown as described above to near saturation and pelleted. The pellet was resuspended in 100 mM Hepes/OH, pH 7.4 with PMSF and cells were lysed by vortexing with glass beads and sonication. Cell debris was removed by spinning at 100 g for 5 min. The membrane fraction was isolated by centrifugation at 13,400 g for 30 min. The membrane pellet was washed in the same buffer and repelleted. The pellet was then resuspended
in 0.5 % (w/v) CHAPS, boiled for 10 min to solubilize membrane-associated proteins, and repelled at 13 400 g for 10 min. This pellet was saved for cell-wall association analysis (see below). Triton X-114 [0.5 ml of a 10 % solution (v/v)] and Tris-buffered saline (0.5 ml) were added to the CHAPS supernatant and incubated at 37 °C for 10 min to cause Triton X-114 partitioning. The aequous top layer was discarded. Triton X-114 partitioning was repeated one more time to remove all remnants of soluble proteins. Half of this solution was precipitated with TCA (trichloroacetic acid) (6 % v/v final concentration) to remove SDS, which denatures anti-FLAG resin. The precipitated proteins were washed with cold acetone and loaded on to EZview Red ANTI-FLAG M2 Affinity Gel (Sigma). After overnight incubation at 4 °C, bound proteins were eluted with FLAG peptide and treated with Endo H (New England Biolabs) according to the manufacturer’s instructions. The presence of INV was confirmed by SDS/PAGE and Western blotting using an anti-FLAG M2 monoclonal antibody-peroxidase conjugate (1:2000 dilution).

The possibility that the Y21-tagged form of INV would be associated with the outer glucan layer of the S. cerevisiae cell wall was also examined. The pellet described above was resuspended in 200 μl of 50 mM Hepes/OH, pH 7.4, with 0.1 μl of 2-mercaptoethanol. QUANTAZYME β-glucosidase (10 units, QBIogene) was added and the solution was incubated at 30 °C for 2 h to release any glucan-bound proteins [7]. The supernatant was immunoprecipitated on anti-FLAG resin, eluted, treated with Endo H, and evaluated by SDS/PAGE and Western blot as described above.

RESULTS

Conversion of the secreted form of invertase into a substrate for GPI-T

The SUC2 gene in pCYI-20, encoding a secreted form of INV, was modified at its 3’ end to append a FLAG tag (Figure 1, INV) to facilitate immunoprecipitation and detection and to serve as a control construct for initial experiments. This plasmid was then modified to introduce the 21 amino acid C-terminal signal sequence from Y21, a fungal acid protease (Figure 1, Y21 [31]): the putative ω-site was mutated to alanine in this and all subsequent constructs to optimize continuity following anchor attachment (Figure 1). (Predictive algorithms point to the cysteine at the ω-1 position as a potential alternative to alanine as the ω-site (see the Discussion section); the results presented herein assume that measured INV activity will be unaffected by modification at cysteine compared with alanine. These two plasmids were transformed into the yeast strain SEY6210α. Cell cultures were grown and evaluated for extracellular INV activity using a colorimetric assay for glucose that is commonly used to assay this enzyme [39]. Intact cells, containing either construct, displayed robust INV activity (Figure 2A), demonstrating that INV was expressed and translocated through the secretory pathway as predicted. Total INV expression levels were highly similar between these two and other constructs evaluated in the present study (see Supplementary Figure S1 at http://www.bioscirep.org/bsr/032/bsr0320577add.htm).

Several experiments were used to further confirm that the Y21 signal sequence had converted INV into a GPI-anchored protein, leading to its association with the plasma membrane or cell wall rather than secretion as a soluble protein. First, the media and cells were separated from a yeast cell culture and each fraction was assayed for INV activity. More than 90 % of the detectable INV activity remained associated with the yeast cell (Figure 2B). In contrast, when the INV construct was examined, >30 % of the INV activity was secreted (results not shown). The observation that 70 % of the secreted INV activity was still associated with cells was assumed to arise from most of this protein having been trapped in the periplasmic space. Therefore, to rigorously confirm that the Y21 INV construct had been modified with a GPI anchor rather than secreted into the periplasm, the membrane and cell wall fractions of the Y21-encoding strain were separated by detergent solubilization. Several different detergents were tested for optimal partitioning of the soluble form of INV into the aqueous layer (e.g. SDS, CHAPS, Triton X-100 and igepeal); CHAPS was selected because it yielded the most reproducible results (results not shown). CHAPS–solubilized membranes and purified cell wall fractions were analysed for the presence of INV, via Western blot detection of the FLAG tag (Figure 2B, inset). INV was detected in both fractions, with most of the protein being retained in the detergent-soluble fraction (lane 1), indicative of GPI anchoring. Low levels of INV could be released from the cell wall material upon treatment with QUANTAZYME, a β-1-3 glucanase (lane 2); covalent attachment to the cell wall glucans is another hallmark of proteins that have been modified by a GPI anchor. The localization of INV to the plasma membrane and the outer cell wall confirms that the Y21 signal sequence is sufficient to convert INV into a substrate for GPI-T.

Finally, because the Y21 construct only encoded the C-terminal GPI signal sequence and none of the ω minus region from Y21, an additional Y21-based construct, containing the wild-type ω-site and ω-1 to ω-8, followed by the C-terminal signal sequence found in Y21, was also evaluated (Y21-Ext). Expression of INV from this construct led to similar extracellular presentation and plasma-membrane localization of GPI-anchored INV as compared with the Y21 construct (Figures 2C and 2D), demonstrating that Y21 is a robust substrate for GPI-T and that the FLAG tag is not interfering with GPI-T’s ability to recognize Y21 as a substrate. In order to focus solely on the impact of the C-terminal GPI signal sequence, all further experiments were conducted using the original Y21 construct.

GPI-T optimally recognizes the yeast Y21 signal sequence over two different human GPI-T signal sequences

The Y21 GPI-T signal sequence was replaced by two different human sequences in order to determine if this INV assay is sensitive enough to identify differences in anchoring efficiency.
Species specificity within the GPI signal sequence

Figure 2  Comparison between the secreted and GPI-anchored forms of INV
(A) A colorimetric extracellular assay of INV activity. The rates of glucose formation from intact cells secreting INV (circles) versus those expressing Y21 (squares), a designed substrate for GPI-T, were determined. These data represent the averages of experiments run in triplicate; error bars represent standard error. (B) Y21 is attached to the yeast cell via both the plasma membrane and the outer glucan layer. More than 90% of the total INV activity was associated with the yeast cells rather than secreted. Inset: The Y21 construct led to expression of INV in a form that was largely associated with the plasma membrane (lane 1) and partially associated with the outer glucan layer (lane 2). Comparison between the Y21 and Y21-Ext INV constructs. (C) The extended Y21 sequence (Y21-Ext) has greater surface presentation than Y21. (D) Y21 and Y21-Ext are each associated more with the outer cell wall than secreted (74±6 and 96±4%, respectively).

and to probe for species specific recognition of different signal sequences. The first construct incorporated the C-terminal 25 amino acids from the human CA25 antigen on to the C-terminus of INV (Figure 1, CA25 [33]). The 30 amino acid GPI-T signal sequence of the uPAR was chosen as the second human sequence to be examined (Figure 1, UP30 [32]). Plasmids expressing INV with the two different human signal sequences were evaluated as described above for the Y21 construct. Both CA25 and UP30 signal sequences resulted in less presentation of INV on the cell surface than Y21 (Figure 3A), consistent with optimal recognition of the yeast Y21 sequence over either of the two human signal sequences. UP30 resulted in substantially more surface presented INV activity than CA25, which was low enough to be only slightly above background. Like Y21, for both of these constructs, nearly all INV activity was associated with cells rather than secreted into the medium (Figure 3B).

Determining the boundaries that dictate species specificity
Based on the Y21 and CA25 results shown in Figure 3(A), a series of CA25/Y21 chimaeras were constructed and evaluated in order to identify which parts of the sequence contribute to the high activity observed with Y21 (Figure 4A). Both CA25 (Figure 4A, light grey) and Y21 (Figure 4A, black) were divided into four sections of equal length (±1 amino acid) and these sections were used as the basis for chimaera design. Each chimaera was constructed to contain one section of Y21, inserted into the analogous position in CA25. In this way, each first generation CH (chimaera), denoted CH-A through CH-D, contained approx. 75% of CA25 human sequence, with the remaining 25% replaced by 5–6 residues of the fungal Y21 sequence.

Analysis of cells expressing these chimaeric proteins revealed that CH-C, containing the third quarter of the Y21 sequence,
The amount of surface-associated INV activity was quantified for three different INV constructs, containing the Y21 signal sequence (circles), the UP30 signal sequence (squares), and the CA25 signal sequence (diamonds). The Y21 construct results in the highest level of measurable INV activity. CA25 and UP30 are each predominantly associated with the yeast cell rather than secreted (90±4% and 84±9%, respectively).

led to greater extracellular INV activity than CA25 or any of the other chimaeras, however this activity was only about 30% that of the original Y21 sequence (Figure 4B). CH-A, containing the Y21 ω+1 to ω+5 residues, did not produce any detectible cell surface activity, indicating unexpected, deleterious synergy between this region of the Y21 sequence and the rest of the CA25 peptide. The remaining two chimaeras (CH-B and CH-D) were nearly identical to the wild-type CA25 peptide.

Next, a second generation of chimaeras, arising from different combinations of CH-A through CH-D, was constructed and evaluated (Figure 5A). In this case, each chimaera contained 50% human CA25 character and 50% fungal Y21 character, based on primary sequence. Analysis of these chimaeras reconfirmed the importance of the five Y21 amino acids inserted in CH-C (Ala–Arg–Phe–Ile–Thr), as the two chimaeras (CH-CD and CH-BC), both of which included this pentapeptide region, had enriched activities (70 and 45% compared with Y21, respectively, Figure 5B). Furthermore, the heightened INV activity observed with CH-CD suggests favourable synergy between the C and D regions of Y21 because this chimaera is significantly more active than the sum of the activities from CH-C and CH-D. Even CH-AC showed INV activity levels equal to that of CH-C, indicating that the importance of the five amino acids in CH-C are sufficient to overcome the deleterious effects observed when only the Y21 CH-A region was inserted [compare CH-A in Figure 4(B) with CH-AC in Figure 5(B)]. These results define the peptide Ala–Arg–Phe–Ile–Thr (ω+11 to ω+15) as a critical region for recognition by S. cerevisiae GPI-T.
Species specificity within the GPI signal sequence

Figure 5  Second generation of human/yeast chimaeric GPI-T signal sequences

(A) Based on the results shown in Figure 4, a second generation of CA25 chimaeras were constructed and tested. These chimaeras were comprised of 50% human (light grey) and 50% yeast (black) signal sequence character, as shown. Complete signal sequences are given in Supplementary Table S2 at http://www.bioscirep.org/bsr/032/bsr0320577add.htm.

(B) Each chimaera was assayed for cell surface INV activity and compared with the activity of the Y21 and CA25 constructs and to chimaera C, the best of the first generation of chimaeras. (C) Rosetta models of Y21, CA25 and three chimaeric GPI-T signal sequences. Models are lined up in order of increasing expression of cell surface INV activity and are colour coded according to the location of the human CA25 sequence (teal) and yeast Y21 sequence (orange). The ω-site (left of each model) is shown as a stick figure. For the four models that contained the ARFIT sequence (ω + 11 to ω + 15), the arginine is also shown as a stick figure. This residue was used to align the four helices in the same orientation. Cartoon schematics (bottom) are coloured as in (A).

DISCUSSION

Conversion of INV into a substrate for GPI-T

In the present study, we have demonstrated that the addition of the 21 amino acid C-terminal GPI-T signal sequence of Y21 is sufficient to convert the soluble form of INV into a substrate for GPI-T (Figure 2A). GPI-anchored INV is translocated to the outer membrane bilayer where it remains associated with the plasma membrane (Figure 2B). Furthermore, a small amount of INV became covalently associated with the outer glucan layer of the S. cerevisiae cell wall (Figure 2B) [6]. In combination, these experiments clearly indicate the presence (and loss) of a GPI anchor. Likewise, the extended construct (Y21-Ext) behaved in a similar fashion to Y21 (Figures 2C and 2D) confirming that the presence of the FLAG tag had no deleterious effects on expression and cell surface presentation of GPI-anchored INV. This assay offers a new, facile method for quantifying the surface presentation of different GPI-T signal sequences.

This assay was used to compare two human sequences (CA25 and UP30) with Y21, demonstrating that it is sensitive to variations that arise from species-specific perturbations in the GPI-T signal sequence (Figure 3A). Importantly, because changes were only introduced in the signal sequence, once INV has been modified by GPI-T to contain a GPI anchor, the impact of the signal sequence is negated and translocation to the surface will be independent of the original signal sequence. Since all constructs were expressed at similar levels (Figure S1), the observed differences in INV activity could only arise from differences in how GPI-T recognizes each signal sequence or, less probably, in different rates of translocation into the ER for the different C-terminal GPI-T signal sequences. This last possibility seems unlikely because the two human signal sequences were optimized for S. cerevisiae codon usage, an approach known to increase acceptability of heterologous sequences in yeast [43], and because the signal sequence is C-terminal, following expression of INV, a wild-type S. cerevisiae protein. Consequently, we can expect that these alterations would prevent dramatic differences in translation and translocation. At this point, it seems most probable that the differences observed among Y21, UP30 and CA25 are a result of differences in recognition by GPI-T, although the possibility that translocation plays some small role cannot be entirely ruled out.

Detailed computational analyses have previously been used to predict GPI-anchored proteins in the S. cerevisiae proteome and in other eukaryotes, based on trends observed in known...
anchored proteins [3,4,9,44]. Sequences can be analysed via an online server (http://mendel.imp.ac.at/gpi/fungi_server.html) in order to estimate whether a given signal sequence will be a substrate for GPI-T. Interestingly, only CA25 and several of the chimaeras had positive scores, leading to the prediction that these proteins, but not Y21 and UP30, would be substrates for the *S. cerevisiae* GPI-T. This prediction is in contrast with the results presented herein which demonstrate that Y21 and UP30 are indeed substrates for GPI-T. One possible explanation for this dichotomy is that the big-PI predictor was designed based on a limited number of known fungal GPI-T signal sequences. Perhaps the use of the FLAG tag in the ω-minus region of our constructs is too unnatural to be accommodated by the algorithm. Consistent with this hypothesis, the Y21-Ext sequence, which includes eight amino acids N-terminal to the ω-site, receives a positive score when analysed by the big-PI algorithm and is also clearly a substrate for the *S. cerevisiae* GPI-T (see Figure 2C). The practical limitations of a predictive algorithm must be considered here, since it predicts that Y21-Ext (and not Y21) is a good GPI-T substrate, whereas our assay clearly demonstrates that both constructs are good substrates. This highlights the importance of developing a facile assay with an easy readout such as this one, to confirm the veracity of this prediction. Further experiments are needed to understand whether or not this arginine causes favourable or unfavourable interactions; mutation of this residue and characterisation using the INV assay described herein will offer some insight. One intriguing possibility is that this arginine is included to limit the extent of Y21 cell surface presentation in wild-type yeast cells.

### Preferential recognition of an *S. cerevisiae* signal sequence over two human sequences

By comparing Y21 with the two human signal sequences CA25 and UP30, quantitative evidence for species specificity was obtained, as Y21 clearly leads to the highest levels of extracellular presentation of INV (Figure 3A). As only a few examples of species specificity have been reported, further comparisons need to be conducted and our INV assay offers an ideal platform for such an analysis. Questions we can ask include where does the Y21 sequence fall in the spectrum of other fungal GPI-T signal sequences and how do other human signal sequences respond? To our knowledge, the impact of the Y21 signal sequence has not been quantitatively evaluated, although the signal sequences of other Yapsin proteases led to comparable, mid to high levels of surface presentation when fused to αGal, suggesting that Y21 should also be about average amongst yeast signal sequences [38]. Efforts are underway to expand the present analysis to include a wider range of fungal and mammalian signal sequences.

### A hydrophobic peptide that is interrupted by an arginine is critical for GPI-T recognition

The dramatic differences observed between CA25 and Y21 (Figure 3A) offered an ideal starting point to begin to dissect how GPI-T distinguishes between optimal and non-optimal sequences, particularly with implications for species specificity. The chimaeras analysed herein point to a pentapeptide in the hydrophobic tail of the fungal Y21 sequence (ω+11 to ω+15, Ala–Arg–Phe–Ile–Thr) as critical for cell surface presentation of GPI-anchored INV (Figures 4 and 5). All chimaeras containing this peptide were sufficient to confer activity, to varying extents, on the human CA25 construct that by itself was a very weak substrate for GPI-T. Lengthening of this peptide towards either the N- or C-terminus further enhanced activity. These experiments also demonstrated the importance of the CH-C region as an anti-determinant for recognition of CA25 by the fungal GPI-T; truncated peptide substrates, wherein this region is deleted, are more active GPI-T substrates than the full-length CA25 peptide (R. Morissette, Y. Varma and T. Hendrickson, unpublished work).

This pentapeptide includes a peculiar feature of the Y21 sequence: the ω+11 to ω+15 pentapeptide falls in the hydrophobic region of this signal sequence, but it is disrupted by a highly hydrophilic, charged arginine (ω+12). The presence of charged residues is not a common feature of yeast GPI-T signal sequences, but there are other examples, including Gas5, a β1-3 glucanosyltransferase that has an arginine at ω+12 and Utr2, a glycosidase involved in cell wall maintenance, which has an arginine at ω+11 [45]. (To our knowledge neither of these ω-sites has been experimentally validated.) Significant advances in our understanding of how GPI-T recognizes its signal sequences are needed to understand whether or not this arginine causes favourable or unfavourable interactions; mutation of this residue and characterisation using the INV assay described herein will offer some insight. One intriguing possibility is that this arginine is included to limit the extent of Y21 cell surface presentation in wild-type yeast cells.

### Does GPI-T recognize structural features in its C-terminal signal sequences?

Finally, the fact that the GPI-T signal sequence lacks a consensus sequence but appears to be recognized by various GPI-T orthologues differently raises the possibility that structural variations could participate in dictating sequence specificity. Prior analyses have predicted that the region around the ω-site of different signal sequences would be disordered [44]. Furthermore, the possibility that the hydrophobic tail would have α-helical character has also been considered but is still a point of debate [3,46]. We used the Rosetta structure prediction algorithm (www.rosetta.org) [47] to model structures for the signal sequences Y21, CA25 and the three chimaeras that showed surface INV activity (CH-C, CH-BC and CH-CD). Rosetta uses a variety of methods to predict secondary structure and, in the case of these models this program builds tertiary models de novo [48]. These models are shown in Figure 5(C) in order of increasing INV activity, and are colour-coded so that CA25 sequences are shown in teal and Y21 sequences are in orange. What is immediately striking is that the predicted helical character correlates with observed INV activity, with Y21 and CH-CD having the longest helices and the most activity and CA25, having the shortest helix and almost no detectible activity. Additionally, the α-helices in the Y21 and CH-CD structures are predicted to extend almost completely through to the C-terminus. These models suggest that the fungal GPI-T...
recognizes the α-helical structure, at least in the hydrophobic region of substrate signal sequences. Perhaps this structural requirement forms the basis for species-specific recognition of fungal GPI-T signal sequences. Further studies are underway to test this hypothesis.

**AUTHOR CONTRIBUTION**
Rachel Morissette designed the study and the experiments, and carried out the experiments. Yug Varma carried out the experiments. Tamara Hendrickson designed the study and the experiments, and supervised the research. All authors participated in the interpretation of the data and in writing the paper.

**ACKNOWLEDGEMENTS**
We thank Professor Beverly Wendland for supplies (pCYI-20 and SEY6210u) and for helpful discussions. We also thank Dr Jennifer Meitzler for technical assistance in the construction of pJLM021 and Dr Rebecca Toroney for helpful suggestions and editing prior to submission.

**FUNDING**
This work was supported by the Elsa C. Pardee Foundation and the American Cancer Society [grant number RSG-07-217-01-TBE].

**REFERENCES**


SUPPLEMENTARY ONLINE DATA

Defining the boundaries of species specificity for the *Saccharomyces cerevisiae* glycosylphosphatidylinositol transamidase using a quantitative *in vivo* assay

Rachel MORISSETTE*1, Yug VARMA*†2 and Tamara L. HENDRICKSON*†3

*Department of Chemistry, Remsen Hall, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, U.S.A., and †Department of Chemistry, Wayne State University, 5101 Cass Avenue, Detroit, MI 48202, U.S.A.

![Anti-FLAG blot](image)

**Figure S1** Anti-FLAG blot showing similar expression levels for all invertase constructs investigated in the present study

Samples were not treated with Endo H to reflect the natural state of the constructs. GAPDH (glyceraldehyde-3-phosphate dehydrogenase) was used for loading control.

1 Present address: National Institutes of Health, National Institute on Aging, Baltimore, MD 21224, U.S.A.
2 Present address: Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA 94158, U.S.A.
3 To whom correspondence should be addressed (email tamara.hendrickson@chem.wayne.edu).

© 2012 The Author(s) This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial Licence (http://creativecommons.org/licenses/by-nc/2.5/) which permits unrestricted non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.
<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5′ → 3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM-A-F3</td>
<td>GGATTCTTCAATGGTACTCATTGGAAGC</td>
</tr>
<tr>
<td>RM-A-R1</td>
<td>CCTCTAGCATGCTTTTACTTCCCTTACTTGG</td>
</tr>
<tr>
<td>RM-B-F4</td>
<td>GTAAAAGCATGCTAGAGGTTATAAACTTATTGTC</td>
</tr>
<tr>
<td>RM-B-R5</td>
<td>GGGTTATTATTATTTTTGACACCATCAACTGG</td>
</tr>
<tr>
<td>RM-FLAG-F1</td>
<td>TGATTTACAAGGATGACGACGATAAGGCATG</td>
</tr>
<tr>
<td>RM-FLAG-R1</td>
<td>CCTTATCGTCGTCATCCTTGTAATCACATG</td>
</tr>
<tr>
<td>RM-21mer-F4</td>
<td>CGCTGGTGGTCATAATTTGAATCCACCATTTTTTGCTAGATTTATTACTGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-21mer-R4</td>
<td>CCTAAATATGAGATGAAAAATAGCAGTAATAAATCTAGCAAAAAATGGTGGATTCAAATTATGACCACCACGCGCATG</td>
</tr>
<tr>
<td>RM-WT25mer-F3</td>
<td>CGCTGCATCCAGCAATATCAGCGGAGGCATTTTCCTTTTCTTCGTGGCCAATGCTCAATCACCACCTCCTCTCTGCTGGATGGCAGCATG</td>
</tr>
<tr>
<td>RM-WT25mer-R3</td>
<td>CCTAACTGAAACAGAAGAGGTGGATGATGGCATTGGCCACGAAGAAAATGGTGGATTGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-30mer-F5</td>
<td>ACATGCAATGCGTCTGCTGCTCCACACACACAGGTCCAGCTCATTTGTCTTTGACTATTACTTTGATG</td>
</tr>
<tr>
<td>RM-30mer-R5</td>
<td>ATGGTCACTGCTATGCTGCAACACACTGCTAATAGCATCAACACACACACACACTGCTAATAGCATCAACACACAC</td>
</tr>
<tr>
<td>RM-Chim25-A-F1</td>
<td>CGCTGGTGGTCATAATTTGGGAGGCATTTTCCTTTTCTTCGTGGCCAATGCTCATGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-Chim25-A-R1</td>
<td>CCTAACTGAAACAGAAGAGGTGGATGATGGCATTGGCCACGAAGAAAATGGTGGATTGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-Chim25-B-F1</td>
<td>CGCTGCATCCAGCAATATCAGCAATCCACCATTTTTTTTCGTGGCCAATGCTCATGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-Chim25-B-R1</td>
<td>CCTAACTGAAACAGAAGAGGTGGATAGTAATAAATCTAGCGAAAAGGAAAATGCCTCCGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-Chim25-C-F1</td>
<td>CGCTGCATCCAGCAATATCAGCAATCCACCATTTTTTTTCGTGGCCAATGCTCATGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-Chim25-C-R1</td>
<td>CCTAACTGAAACAGAAGAGGTGGATAGTAATAAATCTAGCGAAAAGGAAAATGCCTCCGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-Chim25-D-F1</td>
<td>CGCTGCATCCAGCAATATCAGCAATCCACCATTTTTTTTCGTGGCCAATGCTCATGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-Chim25-D-R1</td>
<td>CCTAAATATGAGATGAAAAATAGCGATGGCATTGGCCACGAAGAAAATGGTGGATTGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-Chim25-E-F1</td>
<td>CGCTGGTGGTCATAATTTGGGAGGCATTTTCCTTTTCTTCGTGGCCAATGCTCATGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-Chim25-E-R1</td>
<td>CCTAAATATGAGATGAAAAATAGCGATGGCATTGGCCACGAAGAAAATGGTGGATTGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-Chim25-F-F1</td>
<td>CGCTGCATCCAGCAATATCAGCAATCCACCATTTTTTTTCGTGGCCAATGCTCATGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-Chim25-F-R1</td>
<td>CCTAAATATGAGATGAAAAATAGCGATGGCATTGGCCACGAAGAAAATGGTGGATTGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-Chim25-G-F1</td>
<td>CGCTGCATCCAGCAATATCAGCAATCCACCATTTTTTTTCGTGGCCAATGCTCATGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-Chim25-G-R1</td>
<td>CCTAAATATGAGATGAAAAATAGCGATGGCATTGGCCACGAAGAAAATGGTGGATTGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-Chim25-H-F1</td>
<td>CGCTGGTGGTCATAATTTGGGAGGCATTTTCCTTTTCTTCGTGGCCAATGCTCATGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-Chim25-H-R1</td>
<td>CCTAAATATGAGATGAAAAATAGCGATGGCATTGGCCACGAAGAAAATGGTGGATTGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-Chim25-I-F1</td>
<td>CGCTGCATCCAGCAATATCAGCAATCCACCATTTTTTTTCGTGGCCAATGCTCATGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-Chim25-I-R1</td>
<td>CCTAAATATGAGATGAAAAATAGCGATGGCATTGGCCACGAAGAAAATGGTGGATTGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-Chim25-J-F1</td>
<td>CGCTGGTGGTCATAATTTGGGAGGCATTTTCCTTTTCTTCGTGGCCAATGCTCATGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-Chim25-J-R1</td>
<td>CCTAAATATGAGATGAAAAATAGCGATGGCATTGGCCACGAAGAAAATGGTGGATTGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
<tr>
<td>RM-21merExt-F</td>
<td>CTCTTCATCTTCAACTAGAAAAGAAAATGGTGGTCATAATTTGAATCCACCATTTTTTGCTAGATTTATTACTGCTATTTTTCATCATATTTAGGCATG</td>
</tr>
<tr>
<td>RM-21merExt-R</td>
<td>CCTAAATATGAGATGAAAAATAGCGATGGCATTGGCCACGAAGAAAATGGTGGATTGCTGATATTGCTGGATGCAGCGCATG</td>
</tr>
</tbody>
</table>
### Table S2 Plasmids and GPI-T signal sequences used in the present study

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Insert</th>
<th>Signal sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCYI-20</td>
<td>Original plasmid</td>
<td></td>
</tr>
<tr>
<td>pJLM021</td>
<td>pCYI-20 with SphI site inserted</td>
<td></td>
</tr>
<tr>
<td>pRM028</td>
<td>Full-length secreted invertase with C-terminal FLAG tag (INV)</td>
<td></td>
</tr>
<tr>
<td>pRM031</td>
<td>INV with Y21 signal sequence</td>
<td>A_GGHNLNPPFFARFITAIFHHI</td>
</tr>
<tr>
<td>pRM032</td>
<td>INV with CA25 signal sequence</td>
<td>A_ASSNISGGFIFLFFVANAIHLFCFS</td>
</tr>
<tr>
<td>pRM048</td>
<td>INV with UP30 signal sequence</td>
<td>A_AAPQP GPAHLSLTITLMTARLWGGTLW</td>
</tr>
<tr>
<td>pRM064*</td>
<td>CH-A</td>
<td>A_GGHNLNPPFFVANAIHLFCFS</td>
</tr>
<tr>
<td>pRM065*</td>
<td>CH-B</td>
<td>A_ASSNISGGGIFLFAIFTHLFCFS</td>
</tr>
<tr>
<td>pRM066*</td>
<td>CH-C</td>
<td>A_ASSNISGGIFLFAIFTHLFCFS</td>
</tr>
<tr>
<td>pRM067*</td>
<td>CH-D</td>
<td>A_ASSNISGGGIFLFAIFTHLFCFS</td>
</tr>
<tr>
<td>pRM069*</td>
<td>CH-AB</td>
<td>A_GGHNLNPPFFVANAIHLFCFS</td>
</tr>
<tr>
<td>pRM070*</td>
<td>CH-BD</td>
<td>A_ASSNISGGFFVANAIHLFCFS</td>
</tr>
<tr>
<td>pRM071*</td>
<td>CH-AD</td>
<td>A_GGHNLNPPFFVANAIHLFCFS</td>
</tr>
<tr>
<td>pRM073*</td>
<td>CH-AC</td>
<td>A_GGHNLNPPFFVANAIHLFCFS</td>
</tr>
<tr>
<td>pRM074*</td>
<td>CH-CD</td>
<td>A_ASSNISGGIFLFAIFTHLFCFS</td>
</tr>
<tr>
<td>pRM076*</td>
<td>CH-BC</td>
<td>A_ASSNISGGIFLFAIFTHLFCFS</td>
</tr>
<tr>
<td>pRM077</td>
<td>INV with Y21 extended sequence</td>
<td>SSSSTRKEN_GGHNLNPPFFARFITAIFHHI</td>
</tr>
</tbody>
</table>

*Bold residues indicate Y21 signal sequence portions. Non-bolded residues indicate CA25 signal sequence portions.*