

RESEARCH

Open Access



Streamlining N-terminally anchored yeast surface display via structural insights into *S. cerevisiae* Pir proteins

Tea Martinić Cezar¹, Mateja Lozančić¹, Ana Novačić¹, Ana Matičević¹, Dominik Matijević¹, Béatrice Vallée², Vladimir Mrša¹, Renata Teparić¹ and Bojan Žunar^{1*}

Abstract

Surface display co-opts yeast's innate ability to embellish its cell wall with mannoproteins, thus converting the yeast's outer surface into a growing and self-sustaining catalyst. However, the efficient toolbox for converting the enzyme of interest into its surface-displayed isoform is currently lacking, especially if the isoform needs to be anchored to the cell wall near the isoform's N-terminus, e.g., through a short GPI-independent protein anchor. Aiming to advance such N-terminally anchored surface display, we employed *in silico* and machine-learning strategies to study the 3D structure, function, genomic organisation, and evolution of the Pir protein family, whose members evolved to covalently attach themselves near their N-terminus to the β -1,3-glucan of the cell wall. Through the newly-gained insights, we rationally engineered 14 *S. cerevisiae* Hsp150 (Pir2)-based fusion proteins. We quantified their performance, uncovering guidelines for efficient yeast surface display while developing a construct that promoted a 2.5-fold more efficient display of a reporter protein than the full-length Hsp150. Moreover, we developed a Pir-tag, i.e., a peptide spanning only 4.5 kDa but promoting as efficient surface display of a reporter protein as the full-length Hsp150. These constructs fortify the existing surface display toolbox, allowing for a prompt and routine refitting of intracellular proteins into their N-terminally anchored isoforms.

*Correspondence:

Bojan Žunar
bojan.zunar@pbf.unizg.hr

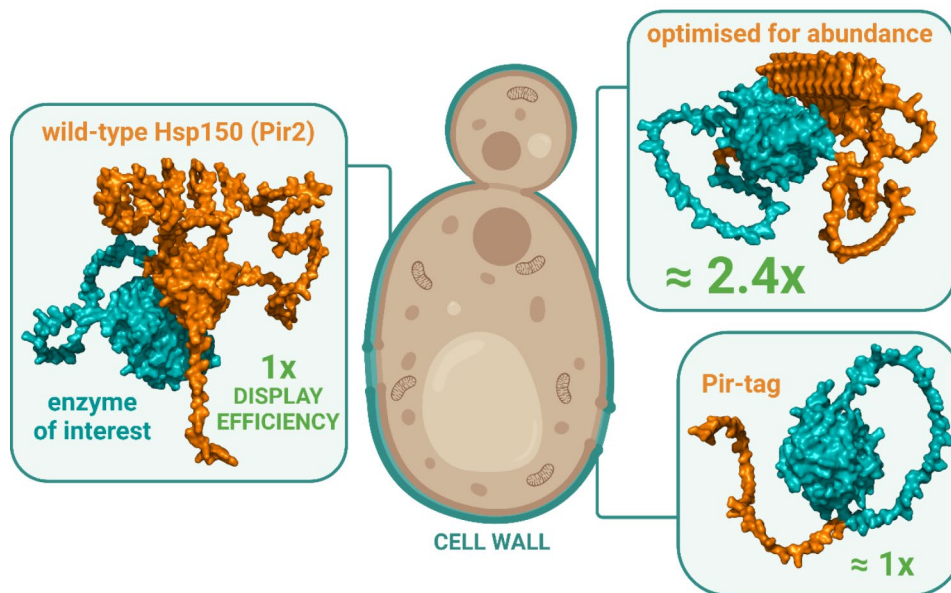
Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Graphical abstract

N-TERMINALLY ANCHORED SURFACE DISPLAY



Keywords Surface display, Pir proteins, Hsp150, Cell wall, *Saccharomyces cerevisiae*

Introduction

Unlike Metazoan cells, microbes encase themselves in a cell wall, a protective layer through which they interact with the environment. While they build this structure primarily from polysaccharides, they also modify it heavily, purposefully meshing it with other macromolecules to adapt its properties and function [1, 2]. Yeast cells, in particular, weave their vegetative cell wall from an inner, polysaccharide-based layer, to which they covalently and noncovalently attach mannoproteins, thus moulding the cell wall's outer layer [3, 4].

Being naturally malleable, the yeast cell wall can, in principle, be modified for many biotechnological applications. By re-engineering mannoprotein synthesis, i.e., displaying enzymes of interest on the cell's surface, we could convert the cell wall's outer layer into a growing and self-renewing catalytic surface, a living material [5–7]. Such a surface is ideal for catalysing reactions reliant on toxic, highly-concentrated, or cell-impermeable substrates, while obtaining readily purifiable products of enzymatic reactions [8, 9].

The most promising approach to cell surface engineering relies on surface display [10–13]. This protein engineering technique reconfigures intracellular proteins for secretion and cell wall binding, commonly by fusing them with native cell wall proteins while respecting the mechanisms that bind native proteins to the cell wall's

inner, polysaccharide layer [14–18]. To bind to the cell wall covalently, the enzyme of interest is fused to either a GPI-anchored cell wall protein or a Pir protein, at its C- or N-terminus, respectively. However, some enzymes tolerate fusions with native cell wall proteins at only one of their termini, otherwise failing to fold properly, to pass quality control in the secretory pathway, or to be flexible enough to catalyse the intended reaction [9, 19, 20]. Thus, only a subset of theoretically possible fusion proteins performs well in vivo, which hinders the full-scale application of surface display.

Pir proteins (proteins with internal repeats) are a family of five *Saccharomyces cerevisiae* proteins (Pir1, Hsp150 (Pir2), Pir3, Cis3 (Pir4), and Pir5) whose sequence features are well documented. All five Pir proteins are directed to the secretory pathway through their N-terminal signal peptides, which are followed by motifs of 18 amino acid residues, after which they were named and which they encode one to eleven times, repeating them in tandem. Through these repeats, Pir proteins bind to the cell surface covalently, with the γ -carboxyl group of the second glutamate within the repeat binding to β -1,3-glucan of the inner polysaccharide layer [21]. Following the repeat region, Pir proteins encode a well-conserved C-terminal region that potentially binds the cell wall through disulphide bonds [22].

Less is known about the structure and function of Pir proteins. Pir1–Pir4 are highly O-glycosylated and important for cell wall stability, as their quadruple deletion mutant grows slowly, has aberrant morphology, and is sensitive to cell wall-stressing compounds [23]. Furthermore, Pir proteins are distinctly expressed throughout the cell cycle and under stress conditions [24, 25], with Pir5 being expressed only during meiotic development [26].

Pir proteins can be used as anchors for surface display [27]. In addition to binding proteins on the cell wall near proteins' N-termini, Pir proteins are the only GPI-independent protein anchors [9]. As such, they avoid GPI-specific bottlenecks, e.g., membrane-to-cell-wall transfer catalysed by the Dfg5 protein family [28] and may be used for efficient concurrent display of GPI-dependently and GPI-independently anchored proteins. However, previously employed Pir-based surface display strategies relied only on known sequence features and native restriction sites. Moreover, as the proteins of interest had to be anchored to the near-full-length Pir protein, they could display only smaller proteins or would otherwise produce a construct too big to be secreted efficiently. Therefore, further innovation of N-terminally anchored surface display was hampered by the unfamiliar 3D structure and ambiguous function of Pir proteins.

Aiming to enhance N-terminally anchored yeast surface display, we focused on essential properties of the Pir protein family, i.e., their 3D structure and putative function, genomic organisation, and the number and conservation of Pir repeats. We explored these topics through modern genomic, structural, and evolutionary *in silico* approaches and machine-learning strategies. Using these newly-gained insights, we rationally engineered *S. cerevisiae* Hsp150 (Pir2), developing constructs and guidelines for highly efficient and easily implementable N-terminally anchored yeast surface display.

Methods

Predicting and exploring protein structures and conformation switching

Protein sequences were aligned with NCBI's blastp [29], using standard parameters. For Pir1–Pir5, tertiary structures and the corresponding structural confidence scores were obtained from EMBL-EBI AlphaFold Protein Structure Database [30, 31]. The structures were visualised in PyMOL 2.5 [32] and structurally superposed with PyMOL's align command. The structural similarity of the β -trefoil domain was investigated with Dali [33], using a heuristic PDB search. The structures of the identified structurally similar β -trefoil domains were superposed with PyMOL's cealign command. The structures of rationally engineered Hsp150- β -lactamase fusion proteins were predicted with Colabfold [34], a version of the deep

neural network AlphaFold2 coupled with Google Collaboratory, using default parameters.

To model the Hsp150 (Pir2) conformation-switching, we emulated Wayment-Steele, et al. [35], i.e., we clustered the Colabfold-generated multiple sequence alignment and specified each alignment branch as a separate Colabfold input.

Investigating pir genomic loci and phylogeny

Homologous genomic loci within the *Saccharomyces* clade were identified by the Smith-Waterman alignment of *S. cerevisiae* Pir regions and whole genome sequences of the neighbouring *Saccharomyces* species. The analysis was performed in Geneie Prime [36] and the results manually curated, with the number of Pir repeats counted by HMMER [37].

Pir phylogeny was inferred in MEGA 11 [38] by applying to the 406 Pir homologues identified in Lozančić, et al. [11] the maximum likelihood method and Dayhoff matrix-based model [39]. Positions with less than 95% site coverage were eliminated. An initial heuristic search was performed with neighbour-join and BioNJ algorithms, and evolutionary rate differences among sites were modelled with a discrete Gamma distribution (5 categories, parameter=2.3669).

The Ykl162c-a signal sequence was analysed with specialised web-hosted machine learning methods DeepSig [40] and SignalP [41]. Gene expression heatmaps were obtained from *Saccharomyces* Genomics Viewer (<http://sgv.genouest.org/>) [42].

Media and growth conditions

E. coli were grown overnight at 37 °C at either 200 rpm in liquid 2xYT media (16.0 g/l tryptone, 10.0 g/l yeast extract, 5.0 g/l NaCl) or on LB plates (10.0 g/l tryptone, 5.0 g/l yeast extract, 5.0 g/l NaCl, 15.0 g/l agar), supplemented with 100 μ g/ml of ampicillin. Wild-type *S. cerevisiae* BY 4741 was grown at 30 °C/180 rpm in standard YPD medium (20.0 g/l peptone, 10.0 g/l yeast extract, 20.0 g/l glucose, and 20.0 g/l agar for solid media). *S. cerevisiae* strains carrying reporter plasmids were grown at 30 °C/180 rpm in chemically defined media without histidine (-his: 6.70 g/l Difco Yeast nitrogen base without amino acids, 20.0 g/l glucose, 1.6 g/l drop-out without histidine, and 20.0 g/l agar for solid media). To induce the *PHO5* promoter, yeast cells were first grown in a repressive and then in a permissive medium, i.e., in -his medium supplemented with 1 g/l of KH_2PO_4 and then in -his medium lacking all phosphate sources, respectively, as in Novačić, et al. [43].

Plasmid and strain construction

Plasmids were constructed in NEB Stable *E. coli* using restriction cloning, Q5 polymerase, and HiFi Assembly

(New England Biolabs, Frankfurt am Main, Germany), according to the manufacturer's instructions. *S. cerevisiae* BY 4741 was transformed as in Gietz, et al. [44]. The details of the constructions are provided in the Supplementary material. Plasmid DNA was isolated with NucleoSpin Plasmid Mini Kit and extracted from an agarose gel with NucleoSpin Gel and PCR Clean up Kit (Macherey-Nagel, Duren, Germany). All constructs were verified by restriction digest and Sanger sequencing. The results of the plasmid analyses agreed with computed restriction maps. Primer synthesis was outsourced to MacroGen Europe (Amsterdam, Netherlands), *de novo* gene synthesis to Genewiz (Leipzig, Germany), and Sanger sequencing to Microsynth (Balgach, Switzerland).

Isolation of cell wall proteins

Cell wall proteins were isolated using the same methodology as in Lozančić, et al. [11]. In short, cells were grown in the -his medium without phosphate until the stationary phase, resuspended in the potassium phosphate (KP) buffer (50 mM, pH 8), mechanically disrupted with glass beads, and washed with KP buffer four times. Non-covalently bound proteins were isolated by boiling the cell wall fraction twice in the reducing Laemmli buffer (0.0625 M Tris, pH 6.8, 2% SDS, 5% v/v β -mercaptoethanol, 0.001% bromophenol blue), which extracted non-covalently bound proteins to the supernatant. To isolate Pir proteins bound to the cell wall covalently, the cell walls stripped of the non-covalently bound proteins, contained within the solid fraction from the previous step, were washed four times with KP buffer, twice with deionised water, and incubated overnight at 4 °C in 30 mM NaOH. Protein samples were standardised by normalising the amount of the extraction buffer and extract used for electrophoresis to the mass of wet cell walls, which was measured before stripping non-covalently bound proteins from the cell walls.

Immunoblotting

Isolated proteins were resolved on 10% SDS-PAGE gels, transferred to a nitrocellulose membrane by semidry transfer (Trans-Blot TurboTM Transfer System, BioRad, Hercules, United States), and analysed by immunoblotting using standard procedures, as in Novačić, et al. [45]. Blots were developed with Clarity Western ECL substrates (BioRad, Hercules, United States) and visualised with a C-DiGit Blot scanner (LI-COR Biosciences, Lincoln, United States).

For Ykl162c-a experiments, strains were grown to the exponential (OD₆₀₀ 2) phase in the YPD medium, and total proteins extracted as in Kushnirov [46]. Myc-tagged Ykl162c-a was probed with anti-c-Myc (1:1,000 dilution; 9E10, Santa Cruz Biotechnology, Dallas, United States) as a primary and mouse IgG κ -binding protein horseradish

peroxidase (1:50,000 dilution; Santa Cruz Biotechnology, Dallas, United States) as a secondary antibody. For Hsp150- β -lactamase experiments, isolated cell wall proteins were probed with anti-HA peroxidase-conjugated antibody (1:1,250 dilution; 11,667,475,001, Roche, Basel, Switzerland).

Measuring β -lactamase activity

Cells were inoculated into 5 ml of -his medium supplemented with KH₂PO₄ and grown overnight at 30 °C/180 rpm. The following day, stationary phase cells were diluted to 0.5 OD₆₀₀/ml in 15 ml -his medium supplemented with KH₂PO₄ and incubated at 30 °C/180 rpm until reaching 2 OD₆₀₀/ml. To induce the *PHO5* promoter, cells were washed with sterile deionised water (sdH₂O), diluted to 0.3 OD₆₀₀/ml in 15 ml of -his medium without phosphate, and grown overnight at 30 °C/180 rpm, thus reaching 2 OD₆₀₀/ml in the morning. Such cells were washed in sdH₂O, then in potassium phosphate (KP) buffer (pH 7, 50 mM), and resuspended in the same buffer to 100 OD₆₀₀/ml. Next, 1 OD₆₀₀ of cells was resuspended in 475 μ l of KP buffer, thermostated at 30 °C/1,200 rpm/2 min, mixed with 25 μ l of nitrocefin (1 mM, dissolved in 50 mM KP buffer with 5% DMSO, pH 7), and incubated at 30 °C/1,200 rpm/5 min. The reaction was stopped by centrifuging the cells (8,000 rpm/30 s) and measuring A₄₈₂ of the supernatant. The activity of each strain was measured in technical triplicates and at least biological duplicates.

Yeast staining, confocal microscopy, and flow cytometry

The yeast culture was grown to the exponential phase (OD₆₀₀ 2), centrifuged (13,000 rpm/30 s), washed with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4), and cells resuspended in either 100 μ g/ml of concanavalin A-FITC (to visualise mannoproteins; C7642, Sigma-Aldrich, St. Louis, United States), 1,000 μ g/ml of calcofluor white (to visualise chitin; 18,909, Sigma-Aldrich, St. Louis, United States), or 500 μ l/ml of aniline blue (to visualise β -glucans; 415,049, Sigma-Aldrich, St. Louis, United States) as in Okada and Ohya (2016). The staining proceeded for 10 min. Each experiment was performed in triplicate.

After staining, the suspension was mixed with 3 ml of PBS and analysed by flow cytometry. For each condition, BD LSRFortessa X-20 (Franklin Lakes, United States) recorded 50,000 events, using a 488 nm laser and 530/30 nm emission filter for concanavalin A-FITC, a 405 nm laser and 450/50 nm emission filter for calcofluor white, or a 405 nm laser and 525/50 nm emission filter for aniline blue. Data were analysed in the R computing environment [47] using openCyto [48] and ggcyto packages [49], with the fluorescence assessed on the flowJo biexponential scale.

For confocal microscopy, following staining, the cells were centrifuged (13,000 g/30 sec), resuspended in 10 μ l of PBS and imaged on a ZEISS LSM 980 confocal microscope (Carl Zeiss, Jena, Germany) under a 40x oil objective, using laser wavelengths of 488 and 561 nm, as in Žunar, et al. [50]. Images were processed using Zeiss Zen Blue software.

To measure supernatant fluorescence, cells were grown in 10 ml of standard -ura medium to stationary phase, at 30 °C/180 rpm/overnight, and centrifuged (3,000 rpm/5 min). The red fluorescence was measured with a Perkin Elmer LS55 fluorescence spectrometer (Perkin Elmer, Waltham, United States), by exciting 3 ml of supernatant in quartz cuvettes at 569 nm and measuring emission at 594 nm.

Results

S. cerevisiae Pir proteins encode lectin-like β -trefoil domains

Aiming to advance the N-terminally anchored Pir-based surface display, we first sought more insight into the 3D structures of native *S. cerevisiae* Pir proteins. *S. cerevisiae* encodes five Pir proteins, Pir1, Hsp150 (Pir2), Pir3, Cis3 (Pir4), and Pir5, each with a unique number of Pir repeats (Fig. 1A). All five Pir proteins are encoded in just two homologous loci, which arose by whole genome duplication, approximately 100 million years ago [51]. As such, Pir1 and Pir5 form one, and Pir3 and Hsp150 another paralogue pair. Interestingly, while Cis3 does not have a genuine paralogue pair, homologous locus encodes Ykl162c-a, a 50-amino-acid peptide annotated as a dubious open reading frame unlikely to encode a functional protein. However, the final 35 residues of Ykl162c-a are homologous to the Cis3 C-terminus.

At the amino acid sequence level, all five *S. cerevisiae* Pir proteins are homologous to each other (Fig. 1B). Each protein encodes a signal peptide, a subunit I, a Kex2-processing site (Lys-Arg) at which Kex2 cleaves the Pir protein into two subunits that remain non-covalently bound to each other, and a subunit II, with the subunit terminology being initially developed to describe the Hsp150 features [52]. Subunit II can be further divided into an N-terminal region, i.e., the internal repeat-encoding region of variable length, and a well-conserved C-terminal region.

To infer the 3D structural features of Pir proteins, we examined their Alphafold2-predicted models. For each Pir homologue, Alphafold2 suggested that only some regions of the protein form well-structured domains, specifically subunit I, one of the internal repeats, and the C-terminal region of subunit II, which together folded into thirteen β -strands per protein (Fig. 1B). Interestingly, one β -strand was always encoded by the subunit I, two by one of the Pir repeats, and ten by the C-terminal

end of the subunit II. Conversely, Alphafold2 suggested that the remaining parts of the proteins, i.e., signal peptide and all-but-one internal repeats, were intrinsically disordered.

We closely analysed the Alphafold2-predicted structure of Hsp150, an archetypal Pir protein. Alphafold2 modelled this protein as topologically knotted, with thirteen β -strands folding into a single well-structured β -trefoil domain (Fig. 1C). As in other *S. cerevisiae* Pir proteins, β -strands originated from non-adjacent parts of the protein. Moreover, Alphafold2 predicted that the remaining ten Pir repeats form a disordered loop, which started and ended at the β -trefoil domain. Other *S. cerevisiae* Pir proteins followed the same basic structure, forming β -trefoil domains with highly conserved β -stranded cores yet divergent β -trefoil turns (Fig. 1D).

To deduce the function of the Pir proteins, we compared the Hsp150-predicted β -strain domain with experimentally solved protein structures available in Uniprot, using Dali [33]. This methodology uncovered significant similarities to several lectin β -trefoil domains (Fig. 1E), e.g., domains contained in nontoxic non-hemagglutinin subcomponent (NTNHA) from *Clostridium botulinum* (PDB: 3VUO), agglutinin from *Amaranthus caudatus* (PDB: 1JLX), and β -trefoil lectin from *Entamoeba histolytica* (PDB: 6IFB). Thus, Pir proteins are likely a family of topologically knotted cell wall-bound lectin-like proteins.

Pir repeats have the potential to fold into a two-faced β -helical domain

Next, we focused on the most notable feature of Pir proteins, their internal (Pir) repeats. We first traced the evolutionary origin of Pir repeats. The Alphafold2 structural model of Cis3, which carries only one Pir repeat, showed that the repeat folds into two β -sheets that formed an integral part of the β -trefoil domain, suggesting it represents the original Pir repeat. Accordingly, the additional Pir repeats likely arose from the expansion of the original two β -sheets of the β -trefoil domain.

Next, we investigated the evolutionary stability of the repeats. At the DNA level, Pir repeats are encoded as near-perfect tandem repeats and thus are predisposed to be intrinsically evolutionary unstable. An evolutionary stably-maintained repeat number would suggest positive selection and imply additional repeat functionality, given that a single Pir repeat suffices to form the β -trefoil domain and covalently bind the protein to the cell wall [21].

To assess the evolutionary stability of Pir repeats, we first mapped their numbers across the *Saccharomyces* clade. We identified homologous Pir loci in non-laboratory probiotic *S. cerevisiae* var. *boulardii* and the remaining seven *Saccharomyces* species, which diverged up to 17 million years ago [53], and we counted the repeats.

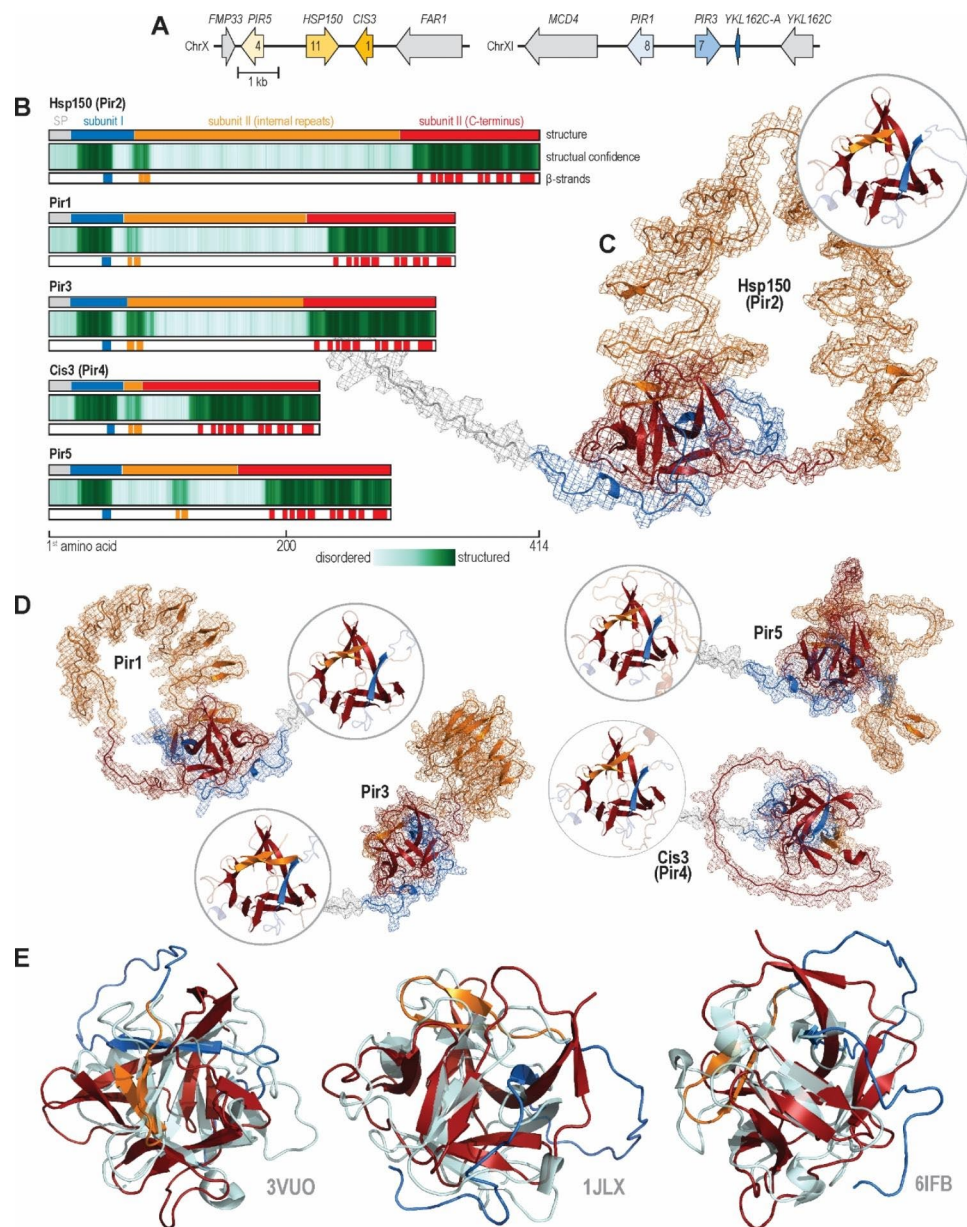


Fig. 1 *S. cerevisiae* Pir proteins encode lectin-like β -trefoil domains. **(A)** *S. cerevisiae* Pir proteins are encoded by the two homologous loci on chromosomes X and XI. Arrows denote gene orientation and the numbers within them the number of Pir repeats. **(B)** The protein sequences of five *S. cerevisiae* Pir proteins (upper track) are homologous, encoding a signal peptide (SP, grey), a small subunit I (blue), and a large subunit II. Subunit II consists of an N-terminal part carrying internal Pir repeats (yellow) and a well-conserved C-terminal part (red). For each Pir protein, AlphaFold2 models only part of the subunit I, one internal repeat, and the C-terminus of subunit II as well structured (middle track, dark green) while suggesting that the remaining parts of the proteins are intrinsically unstructured (faint green). Moreover, AlphaFold2 predicts that well-structured parts of the proteins encode thirteen β -strands (lower track). **(C)** AlphaFold2-structural model of Hsp150 (Pir2), an archetypal Pir protein. Hsp150 is a knotted protein with a single well-structured β -trefoil domain made of thirteen β -strands. The colour scheme follows panel B. The details of the β -trefoil domain are presented in the circular inset, which plots only β -strands in full colour. **(D)** Predicted structures of the remaining four *S. cerevisiae* Pir proteins. Circular insets show structurally-aligned β -trefoil domains. **(E)** PyMOL's structural alignment of Hsp150 β -trefoil domain and several structurally similar crystallised proteins, as determined by Dali. The panels show structural alignments between the Hsp150 domain and nontoxic non-hemagglutinin subcomponent (NTNHA) from *Clostridium botulinum* (PDB: 3VUO), agglutinin from *Amaranthus caudatus* (PDB: 1JLX), and β -trefoil lectin from *Entamoeba histolytica* (PDB: 6IFB). The blue-orange-red colour scheme of the Hsp150 domain follows panel B. The aligned PDB structures are shown in grey

The analysis showed that each of the five *S. cerevisiae* Pir proteins had a characteristic repeat number, which they maintained with only minor variations (Fig. 2A). Interestingly, this analysis also uncovered that Pir6, a lost *S. cerevisiae* Cis3 paralogue, persists in other *Saccharomyces* species.

To test whether Pir repeats remain stably expanded even beyond the *Saccharomyces* clade, we looked at their distribution within predicted proteomes of neighbouring budding yeasts. We analysed 406 Pir proteins from 76 species that diverged from *S. cerevisiae* up to 125 million years ago [11] (Fig. 2B). While the repeat number varied appreciably across the dataset, only Cis3 homologues retained a single Pir repeat. Other Pir homologues consistently maintained more than one Pir repeat, once again suggesting a positive selection for multiple repeats.

Although AlphaFold2 modelled all but one internal repeat as an intrinsically disordered loop (Fig. 1C), we wondered whether such a disordered structure could nevertheless switch conformation, i.e., fold. Emulating a recent AlphaFold2-based study of protein conformation switching [35], we focused on the *S. cerevisiae* Hsp150, which carries 11 internal repeats. We clustered the Colabfold-generated multiple sequence alignment of Hsp150 and performed the AlphaFold2 modelling using each alignment branch as a separate input. Strikingly, the most populated branch (212/288 sequences) produced a 3D structure whose internal repeats stacked one upon other, folding into a well-ordered right-handed, two-faced β -helix (Fig. 2C). In such β -helix, each internal repeat formed a well-conserved β -sheet-turn- β -sheet motif. Notably, in such a structure, glutamine residues, through which Pir proteins covalently link to β -1,3-glucan [21], pointed away from the β -helix, theoretically allowing for cell wall binding (Fig. 2D). Thus, internal repeats have the potential to fold into a two-faced β -helical domain, through which Pir proteins remain covalently bound to the cell wall.

S. cerevisiae encodes a Pir6 remnant

Next, we focused on Ykl162c-a, a seemingly devolved remnant of *S. cerevisiae* Pir6. Ykl162c-a is not specific to the S288c background as it is also present in other commonly used laboratory *S. cerevisiae* strains (*Saccharomyces* Genome Database, data not shown) and probiotic *S. cerevisiae* var. *bouardii* (Fig. 2A). As full-length Pir6 persists in the remaining seven *Saccharomyces* species, including closely related *S. paradoxus*, *S. cerevisiae* must have lost it recently, within the last 5 million years [53], probably through frameshift, as the Pir6 loci of *S. cerevisiae* and *S. paradoxus* differ in several indels.

Out of the 50 amino acid residues of Ykl162c, only the final 35 are homologous to the C-terminus of its paralogue Cis3. Unexpectedly, a bioinformatic analysis

suggested Ykl162c-a evolved a new signal sequence at its N-terminus (Fig. 3A, DeepSig reliability: 1.0, SignalP 6.0 probability: 0.572234, cleavage site between residues 21 and 22). This signal sequence, made of a characteristic hydrophobic core (H-region) flanked by N- and C-regions [54], evolved through two single base-pair frameshifts 40 bp apart, with the resulting region encoding 15 novel amino acid residues. To functionally validate the predicted signalling peptide, we Myc-tagged the native, chromosomally-encoded Ykl162c-a at its C-terminus and preceded it with the constitutive *TEF1* promoter. By immunoblotting, we detected such tagged and overexpressed Ykl162c-a in the whole cell extracts (Fig. 3B), thus confirming it can be stably expressed. To validate the putative signal sequence, we replaced the C-terminal Myc-tag of the constitutively overexpressed Ykl162c-a with the red fluorescent protein ymScarletI and, by measuring the fluorescence of the medium, followed its secretion. Compared to the control strain, in which ymScarletI was not preceded by Ykl162c-a, the strain with ymScarletI-tagged Ykl162c-a produced a significantly more fluorescent medium (17.7328 ± 1.208898 AU vs. 25.0024 ± 5.023018 AU, $p < 0.03$). Moreover, epifluorescent microscopy confirmed that in stationary phase only cells encoding ymScarletI, but not Ykl162c-ymScarletI fusion protein, retained red fluorescence as a strong diffuse signal in the cytosol (data not shown). Thus, Ykl162c-a encodes a novel functional signal sequence.

Moreover, we investigated the expression pattern of native Ykl162c-a. With immunoblotting, we were unable to detect expression of the Myc-tagged Ykl162c-a from its native promoter in the BY 4741 total protein extracts (Fig. 3B), despite analysing exponentially-growing fermenting cells (grown in YPD medium to OD_{600} of 0.2, 2, and 4), respiring cells (grown in standard SP and YPA medium), stationary phase cells (grown in YPD medium to OD_{600} of 10), cell wall-stressed cells (grown for 5 h in YPD supplemented with calcofluor white), or heat-shocked cells (grown at 37 or 42 °C for 5 h). Furthermore, we were unable to detect secreted Ykl162c-a, despite immunoblotting acetone-precipitated proteins secreted to the medium (data not shown).

However, the *Saccharomyces* Genomics Viewer [42] indicated that the *YKL162C-A* locus is transcribed to a potentially significant level in respiring and sporulating but not in fermenting diploid cells (Fig. 3C) and only poorly transcribed in cell-cycle synchronised haploid cells (data not shown). Moreover, the viewer indicated Ykl162c-a transcription footprint spans only *YKL162C-A* and not the entire ancestral *PIR6* locus. Finally, ribosome profiling [55] suggested such transcripts might also be translated (data not shown). Together, the newly-evolved signal sequence and the finely-tuned transcription response suggest that Ykl162c-a might

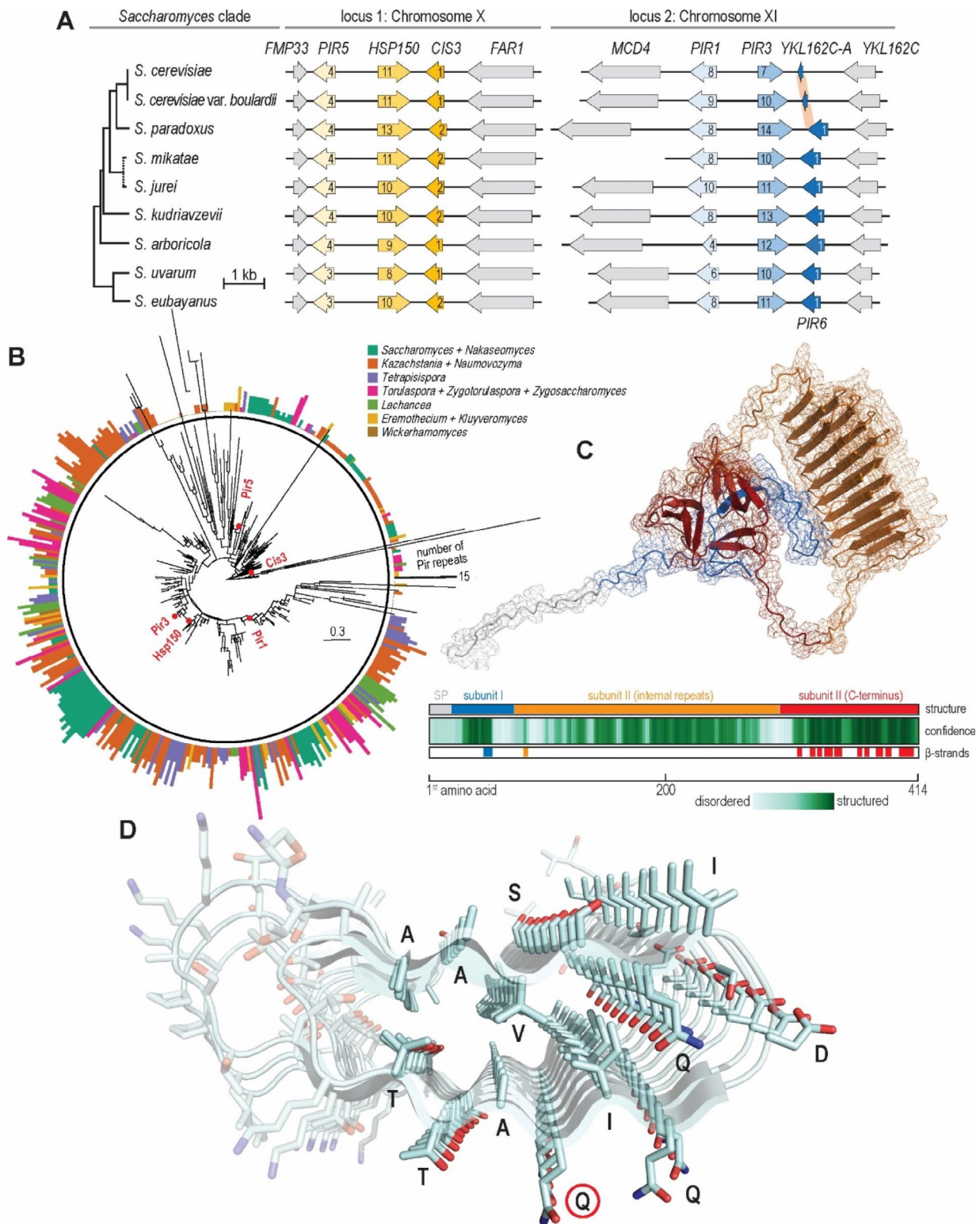


Fig. 2 Distribution and putative folding of Pir repeats. **(A)** Homologous Pir loci across the *Saccharomyces* clade. Only *S. cerevisiae* does not encode Pir6, a Cis3 homologue. Arrows denote gene orientation and numbers within them the number of Pir repeats. **(B)** Unrooted phylogenetic tree of 406 Pir homologues across 76 species of budding yeasts. Red points denote the positions of *S. cerevisiae* homologues. The bar height and colour in the outer rim show the number of Pir repeats in each homologue and its clade of origin, respectively. **(C)** AlphaFold2 model of Hsp150 (Pir2) with folded Pir repeats. The colour scheme of the structural model and the summary tracks follow Fig. 1. **(D)** Sideview of the Hsp150 Pir repeats folded into two-faced β -helix, accentuating ordered folding of amino acid residues. Glutamine residues through which Pir proteins covalently bind to the cell wall are encircled in red

4741 *his3Δ1* locus, expecting it to label the cell surface, as well as the organelles of the secretory pathway that the fusion protein traverses. We stained such cells with FITC-conjugated concanavalin A, which binds to the mannoprotein layer and thus outlines the cell wall, and imaged the cells by confocal microscopy. In contrast to the exclusively green-fluorescing wild-type cells (Fig. 4B), the Hsp150-ymScarletI cells fluoresced in both green and red (Fig. 4C, Supplementary Fig. 1), indicating ymScarletI folded successfully. In the exponentially growing cultures ($OD_{600}=2$), the Hsp150-ymScarletI was detected throughout the cells nonuniformly, mostly in brightly-lit globules, suggesting the fusion protein was contained in the endoplasmic reticulum and transported towards the cell periphery, with a small proportion of Hsp150-ymScarletI localising adjacent to the cell wall, as seen by the faint red ring on the inner border of the mannoprotein layer, indicating the fusion protein was successfully targeted to the cell wall. The faint cell wall-adjacent signal implied that either extracellular conditions dimmed the fluorescence of the fusion protein or that its overexpression had saturated the exporting capacity of the secretory pathway.

We wondered whether such saturation of secretory pathway hampered surface display by triggering cell wall remodelling, which could lower the amount of β -glucan available for mannoprotein binding. To test this hypothesis, we quantified three main components of the yeast cell wall via flow cytometry. We stained wild-type and Hsp150-ymScarletI-overexpressing cells with FITC-conjugated concanavalin A, calcofluor white, and aniline blue, which specifically bound to mannoproteins, chitin, and β -glucan, respectively. However, in all three cases, the distributions of the cell wall components between wild-type and Hsp150-ymScarletI cells remained similar (Fig. 4D), as did the cell size, suggesting overexpressing Hsp150-ymScarletI did not trigger cell wall stress or remodelling. Thus, strong expression of Hsp150-based fusion proteins does not hamper surface display.

Positioning of the Hsp150's fusion partner alters the efficiency of the surface display

Next, we measured the efficiency of the Hsp150-based surface display. We constructed an Hsp150- β -lactamase fusion protein and quantified its presence on the cell surface via a colourimetric assay based on nitrocefin,

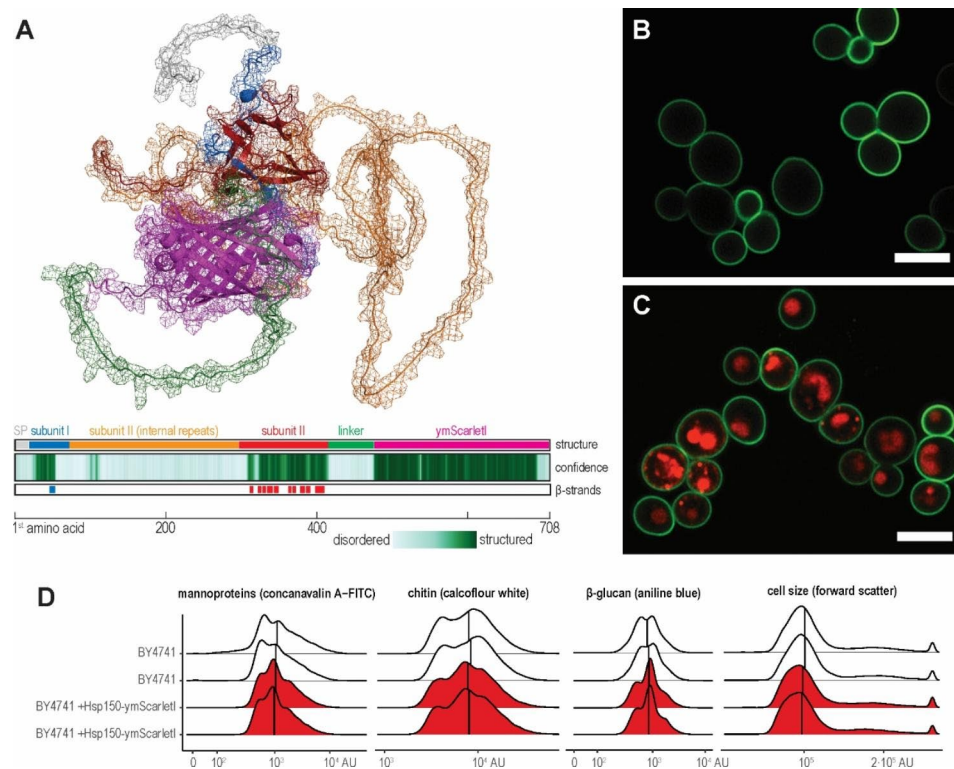


Fig. 4 Saturating secretory pathway with Hsp150-ymScarletI fusion protein does not affect the cell wall. **(A)** Alphafold2-structural model of the Hsp150-ymScarletI fusion protein. The colour scheme of the structural model and the summary tracks follows Fig. 1. **(B)** Confocal micrograph of the wild-type cells stained with FITC-conjugated concanavalin A. Scale bar denotes 5.0 μ m. **(C)** Confocal micrograph of Hsp150-ymScarletI-expressing cells stained with FITC-conjugated concanavalin A. **(D)** Cell size and the amounts of mannoprotein, chitin, and β -glucan of wild-type (white) and Hsp150-ymScarletI-expressing cells (red), quantified via flow cytometry and FITC-conjugated concanavalin A-, calcofluor white-, and aniline blue-staining, respectively. Graphs show two biological replicates

a cephalosporin derivative that does not cross the cell membrane [57, 58]. In this assay, β -lactamase cleaves nitrocefin's β -lactam ring, shifting its peak absorbance from 390 to 482 nm. Using this assay, we tested five Hsp150- β -lactamase fusion proteins, in which β -lactamase was inserted at positions inferred as relevant from the AlphaFold2's structural prediction, i.e., at the beginning of the Hsp150's subunit I (v1), after its first internal Pir repeat (v2), at the beginning of the well-structured part of subunit II (v3), immediately after the Hsp150's C-terminus (v4), or after the disordered tag-rich linker (v5) (Fig. 5A, Supplementary Fig. 2).

While all five fusion constructs were functional, shuffling the β -lactamase had a striking effect, with constructs differing up to 9-fold in their activity (Fig. 5B). Construct v2, carrying β -lactamase after the first internal repeat, performed the best, while construct v4, carrying β -lactamase immediately after Hsp150 C-terminus, performed the worst, reaching only 11% of the activity of construct v2. Interestingly, constructs v1, v3, and

v5, carrying β -lactamase at the beginning of subunit I, i.e., immediately after the signal sequence, at the beginning of the well-structured part of subunit II, and after the disordered tag-rich linker, respectively, performed equivalently, reaching 55% of the activity of construct v2. Thus, the choice of β -lactamase insertion site affected the Hsp150-based surface display profoundly.

We then tested whether Hsp150- β -lactamase constructs still bind to the cell surface, covalently via their internal repeats and noncovalently through steric interactions with other cell wall components (Fig. 5C). The anti-HA immunoblotting of different fractions of cell wall proteins, targeting the tag-carrying disordered linker, showed that all five constructs bonded to the cell wall covalently and noncovalently, with the most active construct v2 also producing the strongest signal in the covalently-bound fraction. Thus, shuffling β -lactamase within the Hsp150 construct does not preclude covalent and non-covalent binding of the Hsp150-fusion proteins to the cell wall.

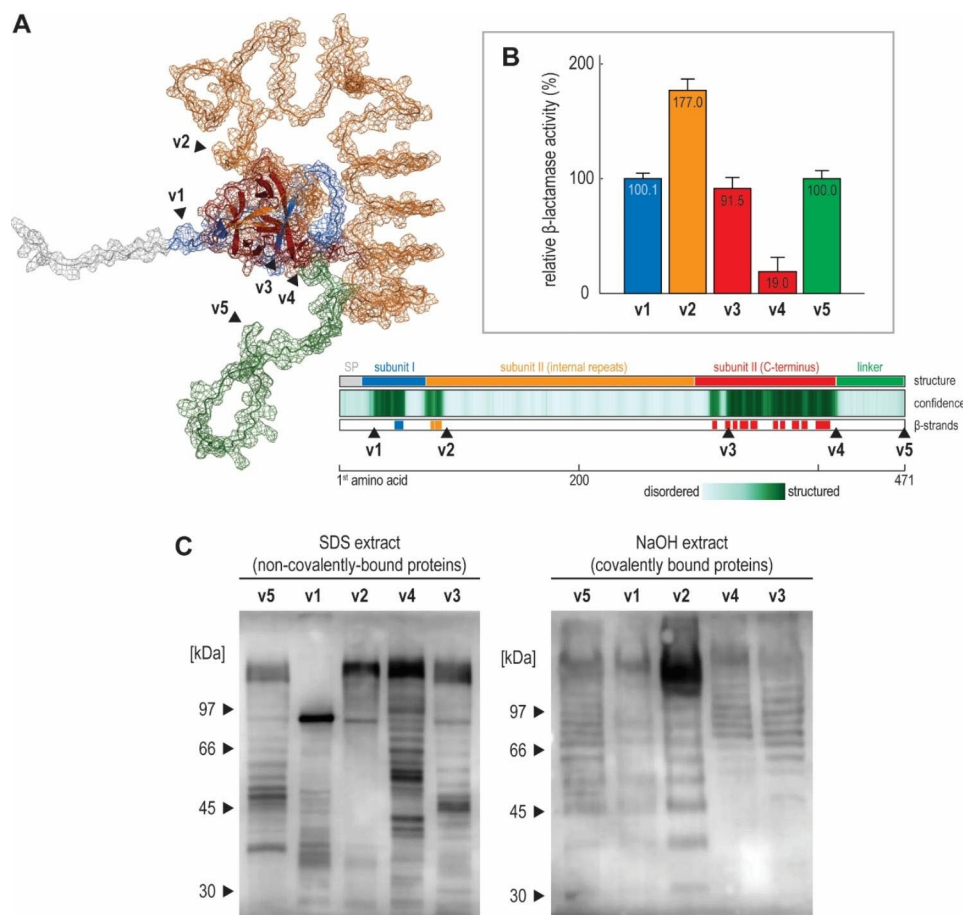


Fig. 5 Shuffling the β -lactamase within the Hsp150-fusion protein affects surface display. **(A)** AlphaFold2-structural model of Hsp150 protein fused with the tag-rich disordered linker. Black triangles point to the β -lactamase insertion sites. The colour scheme of the structural model and the summary tracks follow Fig. 1. **(B)** Relative β -lactamase activity of fusion constructs v1, v2, v3, v4, and v5, normalised to that of construct v5, with error bars denoting standard deviations and bar colours indicating the position of the inserted β -lactamase within the structural model. **(C)** Anti-HA immunoblotting of the noncovalently- and covalently-bound cell wall proteins

The immunoblotting also uncovered an anomalously low molecular mass of noncovalently-bound construct v1, suggesting this construct was cleaved. Indeed, as is the original Hsp150, all five Hsp150- β -lactamase constructs are substrates for Kex2 [52], which cleaves them into two polypeptide chains, by hydrolysing the peptide bond between subunit I and II. However, as only the construct v1 carries 29 kDa β -lactamase within subunit I, before the Kex2 processing site, its cleavage is particularly obvious. Interestingly, the non-Kex2-processed form appears only in the covalently-bound fraction, suggesting it more often binds covalently to the cell wall.

The immunoblots also pointed to pronounced yet reproducible and regular degradation of Hsp150-fusion proteins. Such degradation was also observed when detecting chromosomally-tagged Hsp150 and was unresponsive to protease inhibitors (data not shown). With the number of minor bands mostly matching the number of the internal repeats, the signal points either to uncontrolled proteolytic cleavage during protein isolation or a controlled processing event, either of which likely occurs at the unstructured and thus exposed turns of the β -helix.

Identifying minimal Hsp150 region needed for N-terminally anchored surface display

Attempting to design the Pir-tag, i.e., a short peptide sequence that would allow for N-terminally anchored surface display of any protein, based on the AlphaFold2 prediction of notable Hsp150 structural features, we initially constructed six truncated versions of construct v5, lacking various combinations of subunits I and II (Fig. 6A, Supplementary Fig. 3). The constructs varied in their display efficiency, which ranged from 2% for construct $\Delta 5$, lacking the entire subunit I and the N-terminal part of subunit II, including all internal repeats, to 243% for construct $\Delta 1$, lacking only the C-terminal part of subunit II (Fig. 6B). The comparison of constructs indicated that the C-terminal part of subunit II strongly impaired surface display, lowering it 1.9-fold ($\Delta 2$ vs. $\Delta 4$), 2.4-fold (v5 vs. $\Delta 1$) and 29.5-fold ($\Delta 3$ vs. $\Delta 5$). Conversely, the presence of internal repeats promoted surface display. Accordingly, removing the entire subunit II except one Pir repeat from construct v2, which produced construct v6, lowered its surface display efficiency from 177 to 137%. Finally, to test whether the Hsp150 signal sequence acted as a bottleneck, as Hsp150 is secreted non-conventionally [59–61], we exchanged the Hsp150 signal sequence with that of the Ccw12 cell wall protein (v7). However, both the original construct v5 and the derived construct v7 were displayed equally efficiently. Thus, the Hsp150 C-terminal part of subunit II hinders, while its internal repeats promote surface display, likely through two separate mechanisms, as their effects are additive.

We also tested whether truncated constructs still bind to the cell surface (Fig. 6C). The anti-HA immunoblotting showed all constructs bind to the cell wall covalently and noncovalently, save for the $\Delta 3$ and $\Delta 5$, which, lacking all internal repeats, bind only noncovalently. In line with the previous immunoblotting observations, only constructs with an array of internal repeats produced ladder-like degradation products. Finally, construct $\Delta 3$, despite being bound noncovalently, retained significant β -lactamase activity, demonstrating that covalent binding is not indispensable for surface display.

Finally, to produce a minimal Pir-tag, we designed construct $\Delta 7$, which lacked the entire Hsp150 subunit I and most of its subunit II, thus preceding the HA-linker and β -lactamase only with a signal peptide and one Pir repeat (Supplementary Fig. 3). Intriguingly, this 42-amino-acid peptide displayed β -lactamase as efficiently as the full, 413-residues long Hsp150, while binding the displayed protein both noncovalently and covalently to the cell wall, in approximately the same amount as the full Hsp150 (Supplementary Fig. 4). As such, the Pir-tag demonstrates the bulk of the Hsp150 is dispensable for efficient surface display, which can be recapitulated using merely the Hsp150 signal sequence and one of its internal repeats. Thus, construct $\Delta 7$ defines a minimal, Hsp150-based peptide for efficient N-terminally anchored surface display.

Discussion

In this work, we advance yeast surface display by investigating Pir proteins, i.e., by combining genomic, structural and evolutionary insights to optimally position a protein of interest within *S. cerevisiae* Hsp150. Moreover, we distil the Hsp150 salient features into a minimal Pir-tag, thus allowing for an efficient and easily implementable N-terminally anchored surface display.

Our structural insights, obtained by comparing AlphaFold2 predictions with experimentally determined protein structures, suggested that each Pir protein encodes one lectin-like domain. Such an approach could be expanded to other proteins localised to the cell wall, as many still lack clearly-defined functions, being highly and heterogeneously glycosylated and thus challenging to crystallise [4]. Such studies would also hint at the potential physiological roles of cell wall proteins, as most of their single-gene deletion mutants are in standard phenotypic assays indistinguishable from the wild type, owing to their high functional redundancy [62]. Moreover, as demonstrated here, combining genetic with structural analysis can point to hidden functionality and, as such, should be extended to other incompletely characterised cell wall proteins.

As they carry lectin-like domains, Pir proteins might bind yeast cells to the sugars displayed on the surfaces

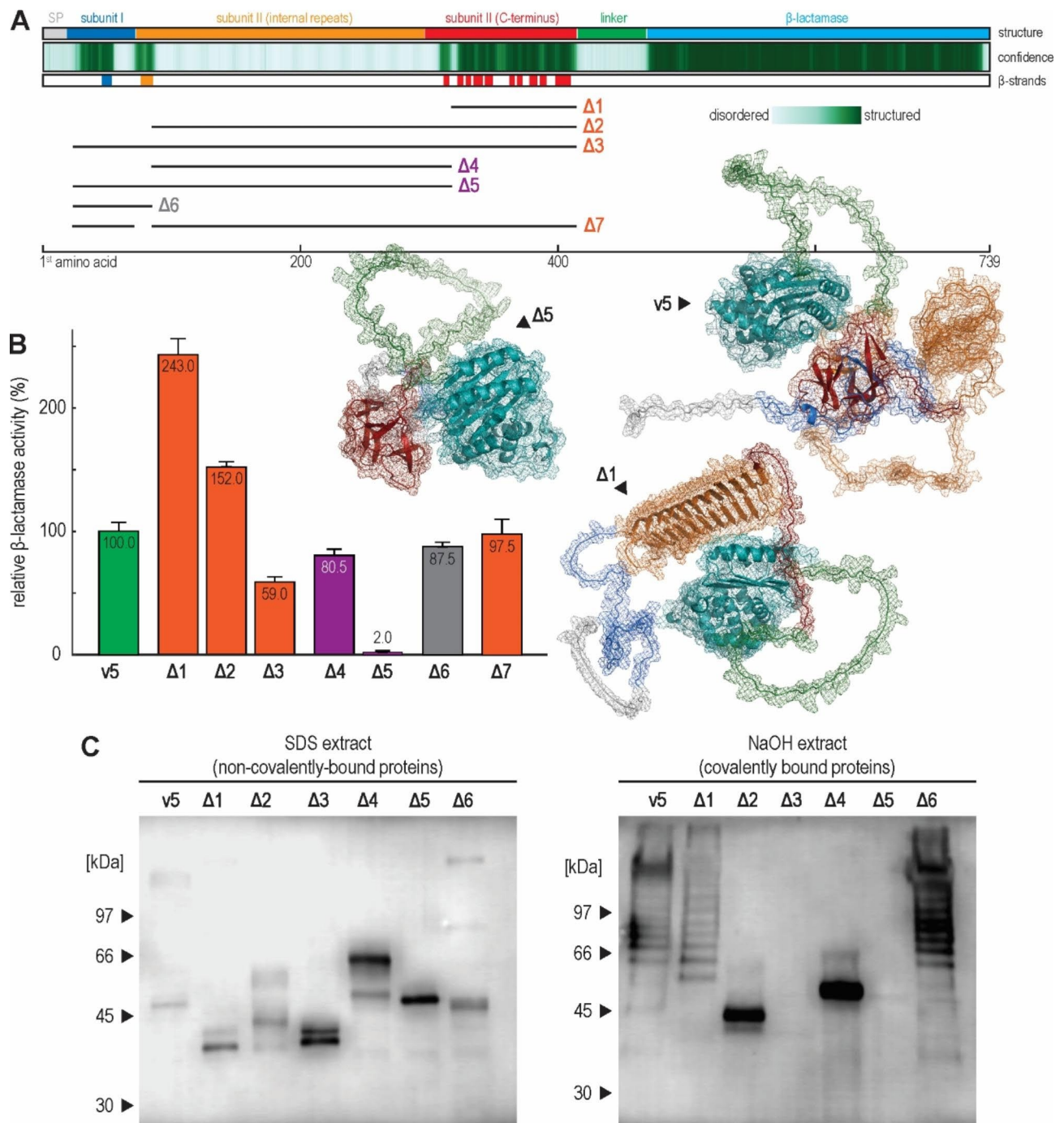


Fig. 6 Pinpointing minimal Hsp150 region needed for efficient N-terminally anchored surface display. **(A)** Summary tracks describing construct v5 and deletions introduced into constructs $\Delta 1$, $\Delta 2$, $\Delta 3$, $\Delta 4$, $\Delta 5$, $\Delta 6$ and $\Delta 7$, with deleted regions marked as black lines, and the visualisations of AlphaFold2-structural models comparing constructs v5, $\Delta 1$, and $\Delta 5$. **(B)** Relative β -lactamase activity of fusion constructs v5, $\Delta 1$, $\Delta 2$, $\Delta 3$, $\Delta 4$, $\Delta 5$, $\Delta 6$ and $\Delta 7$, normalised to that of construct v5, with error bars denoting standard deviations and bar colours corresponding to those in panel A. **(C)** Anti-HA immunoblotting of the noncovalently- and covalently-bound cell wall proteins

of other microorganisms, plants, or animals. By binding to other microorganisms, *S. cerevisiae* could boost the lethality of its environment-sterilising make-accumulate-consume lifestyle [63], i.e., it could attach itself to its unicellular competitors and selectively augment

local ethanol concentration beyond tolerable levels. Conversely, by binding to the plant cell wall, yeast cells could fix themselves to an abundant food source while increasing their wildlife-assisted dispersal. Finally, lectin-like proteins could help yeast colonise the animal digestive

tract, e.g., by aiding overwintering in the wasp gut, which acts as an *S. cerevisiae* reservoir [64], or by populating the mammalian gut. The latter hypothesis is enticing, as Hsp150 (heat shock protein 150) is induced 7-fold at 37 °C but not required for *S. cerevisiae* thermotolerance [52] and was twice as abundant in clinical isolates of *S. cerevisiae*, when compared to the standard laboratory strains [65]. Moreover, being modelled as knotted, Pir proteins should fare better in harsher conditions than their unknotted counterparts and thus perform well in the gut of warm-blooded animals [66, 67]. As standard laboratory tests do not assess interspecies dynamics, such an ecological role would also explain the perceived lack of phenotype in single Pir deletion mutants.

With the cell wall being the primary interface between the microbial cell and the environment [62], the sugar-binding character of Pir proteins likely drives their maintenance, diversification, and subfunctionalisation. Indeed, Lozančić, et al. [11] detected Pir proteins in each of the 78 ascomycete yeasts most closely related to *S. cerevisiae*, with several of these species encoding over a dozen of Pir proteins. Moreover, while lectin-like *S. cerevisiae* Pir domains consist of a well-conserved β -stranded core, their interstrand turns, which determine sugar-binding specificity, differ. The tendency toward subfunctionalisation also explains why *S. cerevisiae* would retain both Pir loci after the whole genome duplication and why Pir proteins are differentially expressed throughout the cell cycle and growth phases [42]. Finally, it is interesting to speculate whether the loss of Pir6 and the evolution of Ykl162c-a was positively selected for by the domestication of *S. cerevisiae*, which affected its growth and life cycle remarkably [68].

The Ykl162c-a is an interesting peptide, encoding only 50 amino acid residues, 21 of which form a newly evolved functional signal sequence. The peptide also avoids all four C-terminal cysteine residues conserved across the Pir family [69]. While the here-presented results do not confirm that the native transcript is translated and secreted, the putative secreted peptide could act as an (interspecies) pheromone or even an antimicrobial molecule [70]. Its presence also suggests that other such peptides with high secretion potential could be encoded yet overlooked within the yeast genome.

The hallmarks of the Pir proteins are their eponymous tandem repeats, whose purpose currently remains unknown. The here-presented models of Pirs as knotted proteins partly discredit the simplest hypothesis, that the repeats simply tether the C-terminal part of the protein far from the cell's surface. An alternative hypothesis is that Pir proteins serve as stretchable proteinaceous cell wall netting covalently linking β -glucans and thus helping to stabilise the cell wall during growth and remodelling [4]. That hypothesis seeks to explain why the deletion of

multiple Pir proteins weakens the cell wall by presuming many Pir proteins bind to β -1,3-glucan through two or more Pir repeats. However, our modelling suggests a third possibility, that by folding into a two-faced β -helix, Pir repeats build a structure with which the cell actively modulates its environment, as do other such structures found in biofilm matrices, adhesins, and antifreeze proteins [71].

Our efforts to advance N-terminally anchored surface display focused on *S. cerevisiae* Hsp150, an archetypal and well-investigated member of the Pir protein family [58, 72]. However, although this protein was previously employed for surface display, a meticulous and called-for benchmarking of its potential insertion sites remained unrealised [27], partly because a proper comparison would need to rely on the previously unknown Hsp150 structure.

To rigorously quantify the Hsp150-based surface display, we implemented several strategies. We fused Hsp150 with a β -lactamase reporter protein and employed a nitrocefin assay, which relies on nitrocefin's cell-membrane impenetrability. Moreover, we placed the Hsp150-fusion protein under the strong inducible *PHO5* promoter [73], ensuring that during preculture, the Hsp150-constructs did not interfere with cell growth. Finally, with confocal microscopy and flow cytometry, we verified at the single-cell and population levels that the strongly expressed fusion protein did not perturb cell wall homeostasis, i.e., did not change the amounts of mannoproteins, chitin, and β -glucans in the cell wall.

By inserting β -lactamase at varying positions throughout Hsp150, which were selected based on the genetic and structural studies of the Pir protein family, we demonstrated that the C-terminal Pir domain inhibits surface display. As immunoblotted constructs produced equally strong signals, despite having a 5-fold difference in activity, impaired surface display probably stemmed from improper protein folding and not degradation of the fusion protein in the secretory pathway. Conversely, inserting β -lactamase in the long unstructured region, immediately after the first Pir repeat, helped it to fold and possibly helped Hsp150 to bind to the cell wall covalently, as evidenced by the strong immunoblot signal.

Construct truncations also indicated that the C-terminal Hsp150 domain undermined while the unstructured repeat-rich region promoted the efficiency of the surface display. The constructs containing the repeat region possibly performed better because cell wall proteins could require unstructured regions to pass through the inner polysaccharide layer efficiently, as was recently shown for bacterial cell wall proteins [74]. Notably, the effects of the C-terminal domain and the repeat-region were additive, as the best-performing construct lacked the C-terminal domain but carried the entire subunit I and all repeats.

Accordingly, as the regions AlphaFold2 deemed unstructured promoted and those it regards as structured hampered surface display, this tool could routinely be used to assess and improve future constructs.

While readily generalisable, the presented results may not extend to every surface-displayed protein. We based our system on β -lactamase, a reporter protein whose activity can be conveniently colourimetrically measured with a cell membrane-impenetrable substrate, a β -lactam nitrocefin. As such, we could reliably quantify only secreted, and not intracellular, β -lactamase. However, while β -lactamase is a superior reporter protein for surface display, it may not be an all-encompassing model for surface display targets. For example, enzymes whose functionality hinges on their N-terminus's flexibility might benefit more from C-terminal GPI-based anchoring. Moreover, diverse enzymatic shapes and catalytic properties could challenge our understanding of the beneficial and detrimental impacts of Hsp150's unstructured and structured regions, respectively.

Finally, from an applicable point of view, we described a Hsp150-derived construct which promoted a 2.5-fold more efficient surface display than the full-length Hsp150. Moreover, through extensive cropping, we developed a Pir-tag, as efficient as full-length Hsp150 but spanning only 4.5 kDa, i.e., 45 amino acids, 18 of which encode a signal sequence. Of course, as the here presented results rely on the β -lactamase/nitrocefin test, they might vary on a protein-to-protein basis. Nevertheless, these two constructs are a much-needed addition to the yeast surface display toolbox that will allow for the efficient and routine refitting of biotechnologically-interesting proteins into their N-terminally anchored surface-displayed isoforms.

Conclusion

In this work, we leveraged the structural and evolutionary insights into Pir proteins to augment the N-terminally anchored yeast surface display. With this strategy, we designed two constructs, one 2.5-fold more efficient than the wild-type Hsp150 and another as efficient as the full-length Hsp150 but spanning only 4.5 kDa. As such, we outline a blueprint for streamlined refitting of proteins for their N-terminally-anchored display on the outer yeast surface.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12934-023-02183-2>.

Supplementary Material 1: Details of the plasmid and strain construction.

Supplementary Material 2: Supplementary Fig. 1 Single-channel micrographs and their overlays of (A) wild-type cells stained with FITC-conjugated concanavalin A and (B) Hsp150-ymScarlet1-expressing cells stained with FITC-conjugated concanavalin A, as shown in Fig. 4. Scale

bar denotes 5.0 μ m. Fig. 2 AlphaFold2-structural models of constructs v1-v5 of the Hsp150- β -lactamase fusion protein. Percentages denote the relative β -lactamase activity of the fusion construct, normalised to that of construct v5. The colour schemes of the structural models follow Fig. 5A. Fig. 3 AlphaFold2-structural models of constructs Δ 1- Δ 7 of the Hsp150- β -lactamase fusion protein. Percentages denote the relative β -lactamase activity of the fusion construct, normalised to that of construct v5. The colour schemes of the structural models follow Fig. 6A. Supplementary Fig. 4 Comparison of the noncovalently- and covalently-bound cell wall proteins v5, Δ 2, and Δ 7, with anti-HA immunoblotting.

Acknowledgements

We thank David Gosset for his help with fluorescent microscopy and flow cytometry on the MO2VING platform (Centre de Biophysique Moléculaire, CNRS, Orleans). Graphical abstract was created with BioRender.com.

Author contributions

CRedit author statement: T.M.C.: Methodology, Investigation, Formal analysis. M.L.: Methodology, Investigation, Formal analysis. A.N.: Methodology, Investigation, Formal analysis, Writing - Review & Editing, A.M.: Investigation. D.M.: Investigation. B.V.: Resources, Supervision. V.M.: Resources, Supervision, Project administration, Funding acquisition. R.T.: Methodology, Resources, Supervision, Writing - Review & Editing, Project administration, Funding acquisition. B.Ž.: Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Visualization, Writing - Original Draft, Writing - Review & Editing.

Funding

This research was funded by The Croatian Science Foundation, grants No. IP-2019-04-2891 and DOK-2021-02-9672.

Declarations

Competing interests

The authors declare no competing interests.

Author details

¹Laboratory for Biochemistry, Department of Chemistry and Biochemistry, Faculty of Food Technology and Biotechnology, University of Zagreb, Pierottijeva 6, Zagreb 10000, Croatia

²Centre de Biophysique Moléculaire (CBM), CNRS, University of Orléans and INSERM, Orléans Cedex 2, UPR 4301, 45071, France

Received: 21 July 2023 / Accepted: 18 August 2023

Published online: 07 September 2023

References

1. Klis FM, Mol P, Hellingwerf K, Brul S. Dynamics of cell wall structure in *Saccharomyces cerevisiae*. *FEMS Microbiol Rev.* 2002;26(3):239–56. <https://doi.org/10.1111/j.1574-6976.2002.tb00613.x>.
2. Orlean P. Cell Wall Biogenesis. The Molecular and Cellular Biology of the yeast *Saccharomyces cerevisiae*. New York: Cold Spring Harbour Laboratory Press; 1988: 229–362.
3. Cabib E, Roberts R, Bowers B. Synthesis of the yeast cell wall and its regulation. *Annu Rev Biochem.* 1982;51(1):763–93. <https://doi.org/10.1146/annurev.bi.51.070182.003555>.
4. Orlean P. Architecture and Biosynthesis of the *Saccharomyces cerevisiae* Cell Wall. *Genetics.* 2012;192(3):775–818. <https://doi.org/10.1534/genetics.112.144485>.
5. Molinari S, Tesoriero RF, Ajo-Franklin CM. Bottom-up approaches to engineered living materials: Challenges and future directions. *Matter.* 2021;4(10):3095–120. <https://doi.org/10.1016/j.matt.2021.08.001>.
6. Nguyen PQ, Courchesne N-MD, Duraj-Thatte A, Praveschotinunt P, Joshi NS. Engineered living materials: prospects and Challenges for using Biological Systems to Direct the Assembly of Smart materials. *Adv Mater.* 2018;30(19):1704847. <https://doi.org/10.1002/adma.201704847>.

7. Žunar B, Ito T, Mosrin C, Sugahara Y, Bénédetti H, Guégan R, et al. Confocal imaging of biomarkers at a single-cell resolution: quantifying 'living' in 3D-printable engineered living material based on pluronic F-127 and yeast *Saccharomyces cerevisiae*. *Biomater Res.* 2022;26(1):85. <https://doi.org/10.1186/s40824-022-00337-8>.
8. Han L, Zhao Y, Cui S, Liang B. Redesigning of Microbial Cell Surface and its application to whole-cell Biocatalysis and Biosensors. *Appl Biochem Biotechnol.* 2018;185(2):396–418. <https://doi.org/10.1007/s12010-017-2662-6>.
9. Teymennet-Ramirez KV, Martínez-Morales F, Trejo-Hernández MR. Yeast surface Display System: strategies for improvement and Biotechnological Applications. *Front Bioeng Biotechnol.* 2022;9. <https://doi.org/10.3389/fbioe.2021.794742>.
10. Gai SA, Wittrup KD. Yeast surface display for protein engineering and characterization. *Curr Opin Struct Biol.* 2007;17(4):467–73. <https://doi.org/10.1016/j.sbi.2007.08.012>.
11. Ložančić M, Žunar B, Hrestak D, Lopandić K, Teparić R, Mrša V. Systematic comparison of Cell Wall-Related proteins of different yeasts. *J Fungi.* 2021;7(2):1–19. <https://doi.org/10.3390/jof7020128>.
12. Tanaka T, Yamada R, Ogino C, Kondo A. Recent developments in yeast cell surface display toward extended applications in biotechnology. *Appl Microbiol Biotechnol.* 2012;95(3):577–91. <https://doi.org/10.1007/s00253-012-4175-0>.
13. Ueda M. Establishment of cell surface engineering and its development. *Biosci Biotechnol Biochem.* 2016;80(7):1243–53. <https://doi.org/10.1080/09168451.2016.1153953>.
14. Abe H, Ohba M, Shimma Y-i, Jigami Y. Yeast cells harboring human α -1,3-fucosyltransferase at the cell surface engineered using pir, a cell wall-anchored protein. *FEMS Yeast Res.* 2004;4(4–5):417–25. [https://doi.org/10.1016/S1567-1356\(03\)00193-4](https://doi.org/10.1016/S1567-1356(03)00193-4).
15. Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol.* 1997;15(6):553–7. <https://doi.org/10.1038/nbt0697-553>.
16. Matsumoto T, Fukuda H, Ueda M, Tanaka A, Kondo A. Construction of yeast strains with high cell surface lipase activity by using Novel Display Systems based on the Flo1p Flocculation Functional Domain. *Appl Environ Microbiol.* 2002;68(9):4517–22. <https://doi.org/10.1128/AEM.68.9.4517-4522.2002>.
17. Nakamura Y, Shibasaki S, Ueda M, Tanaka A, Fukuda H, Kondo A. Development of novel whole-cell immunoadsorbents by yeast surface display of the IgG-binding domain. *Appl Microbiol Biotechnol.* 2001;57(4):500–5. <https://doi.org/10.1007/s002530100802>.
18. Van der Vaart JM, te Biesebeke R, Chapman JW, Toschka HY, Klis FM, Verrips CT. Comparison of cell wall proteins of *Saccharomyces cerevisiae* as anchors for cell surface expression of heterologous proteins. *Appl Environ Microbiol.* 1997;63(2):615–20. <https://doi.org/10.1128/aem.63.2.615-620.1997>.
19. Teparić R, Mrša V. Overview of systems and techniques for surface display of recombinant proteins in yeast *S. cerevisiae*. *Appl Food Biotechnol.* 2015;3(1):3–14. <https://doi.org/10.22037/afbv3i1.9457>.
20. Wang Z, Mathias A, Stavrou S, Neville DM. Jr. A new yeast display vector permitting free scFv amino termini can augment ligand binding affinities. *Protein Eng Des Sel.* 2005;18(7):337–43. <https://doi.org/10.1093/protein/gzi036>.
21. Ecker M, Deutzmann R, Lehle L, Mrša V, Tanner W. Pir proteins of *Saccharomyces cerevisiae* are attached to β -1,3-Glucan by a new protein-carbohydrate linkage. *J Biol Chem.* 2006;281(17):11523–9. <https://doi.org/10.1074/jbc.M600314200>.
22. Castillo L, Martínez AI, Garcera A, Victoria Elorza M, Valentín E, Sentandreu R. Functional analysis of the cysteine residues and the repetitive sequence of *Saccharomyces cerevisiae* Pir4/Cis3: the repetitive sequence is needed for binding to the cell wall β -1,3-glucan. *Yeast.* 2003;20(11):973–83. <https://doi.org/10.1002/yea.1016>.
23. Mrša V, Tanner W. Role of NaOH-extractable cell wall proteins Ccw5p, Ccw6p, Ccw7p and Ccw8p (members of the pir protein family) in stability of the *Saccharomyces cerevisiae* cell wall. *Yeast.* 1999;15(10A):813–20. [https://doi.org/10.1002/\(SICI\)1097-0061\(199907\)15:10A<813::AID-YEA421<3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-0061(199907)15:10A<813::AID-YEA421<3.0.CO;2-Y).
24. Jung US, Levin DE. Genome-wide analysis of gene expression regulated by the yeast cell wall integrity signalling pathway. *Mol Microbiol.* 1999;34(5):1049–57. <https://doi.org/10.1046/j.1365-2958.1999.01667.x>.
25. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, et al. Comprehensive Identification of Cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell.* 1998;9(12):3273–97. <https://doi.org/10.1091/mbc.9.12.3273>.
26. Enyenihi AH, Saunders WS. Large-scale functional genomic analysis of Sporulation and Meiosis in *Saccharomyces cerevisiae*. *Genetics.* 2003;163(1):47–54. <https://doi.org/10.1093/genetics/163.1.47>.
27. Yang N, Yu Z, Jia D, Xie Z, Zhang K, Xia Z, et al. The contribution of pir protein family to yeast cell surface display. *Appl Microbiol Biotechnol.* 2014;98(7):2897–905. <https://doi.org/10.1007/s00253-014-5538-5>.
28. Vogt MS, Schmitz GF, Varón Silva D, Mösch H-U, Essen L-O. Structural base for the transfer of GPI-anchored glycoproteins into fungal cell walls. *Proc Natl Acad Sci USA.* 2020;117(36):22061–7. <https://doi.org/10.1073/pnas.2010661117>.
29. Johnson M, Zaretskaya I, Raytselis Y, Merezukh Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008;36(suppl2):W5–W9. <https://doi.org/10.1093/nar/gkn201>.
30. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2>.
31. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2022;50(D1):D439–44. <https://doi.org/10.1093/nar/gkab1061>.
32. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 2.5. In.; 2022.
33. Holm L. DALI and the persistence of protein shape. *Protein Sci.* 2020;29(1):128–40. <https://doi.org/10.1002/pro.3749>.
34. Mirdita M, Schütze K, Moriawaki Y, Heo L, Ovchinnikov S, Steinegger M. Colab-Fold: making protein folding accessible to all. *Nat Methods.* 2022;19(6):679–82. <https://doi.org/10.1038/s41592-022-01488-1>.
35. Wayment-Steele HK, Ovchinnikov S, Colwell L, Kern D. Prediction of multiple conformational states by combining sequence clustering with AlphaFold2. *bioRxiv* 2022:2022.2010.2017.512570. <https://doi.org/10.1101/2022.10.17.512570>.
36. Dotmatics. Geneious Prime 2022. 2022.
37. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39(suppl2):W29–W37. <https://doi.org/10.1093/nar/gkr367>.
38. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol.* 2021;38(7):3022–7. <https://doi.org/10.1093/molbev/msab120>.
39. Schwarz R, Dayhoff M. Matrices for detecting distant relationships. In: *Atlas of protein sequences* Edited by Dayhoff M: National Biomedical Research Foundation; 1979: 353–358.
40. Savojardo C, Martelli PL, Fariselli P, Casadio R. DeepSig: deep learning improves signal peptide detection in proteins. *Bioinformatics.* 2018;34(10):1690–6. <https://doi.org/10.1093/bioinformatics/btx818>.
41. Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol.* 2022;40(7):1023–5. <https://doi.org/10.1038/s41587-021-01156-3>.
42. Lardenois A, Liu Y, Walther T, Chalmel F, Evrard B, Granovskaia M, et al. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. *Proc Natl Acad Sci USA.* 2011;108(3):1058–63. <https://doi.org/10.1073/pnas.1016459108>.
43. Novačić A, Menéndez D, Ljubas J, Barbarić S, Stutz F, Soudet J, et al. Antisense non-coding transcription represses the PHO5 model gene at the level of promoter chromatin structure. *PLoS Genet.* 2022;18(10):e1010432. <https://doi.org/10.1371/journal.pgen.1010432>.
44. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc.* 2007;2(1):31–4. <https://doi.org/10.1038/nprot.2007.13>.
45. Novačić A, Beauvais V, Oskomić M, Štrbac L, Dantec AL, Rahmouni AR, et al. Yeast RNA exosome activity is necessary for maintaining cell wall stability through proper protein glycosylation. *Mol Biol Cell.* 2021;32(5):363–75. <https://doi.org/10.1091/mbc.E20-08-0544-T>.
46. Kushnirov VV. Rapid and reliable protein extraction from yeast. *Yeast.* 2000;16(9):857–60. [https://doi.org/10.1002/1097-0061\(20000630\)16:9<857::AID-YEA561<3.0.CO;2-B](https://doi.org/10.1002/1097-0061(20000630)16:9<857::AID-YEA561<3.0.CO;2-B).
47. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. In. Vienna; 2022.
48. Finak G, Frelinger J, Jiang W, Newell EW, Ramey J, Davis MM, et al. Open-Cyto: an Open source infrastructure for scalable, robust, reproducible, and

- automated, end-to-end Flow Cytometry Data Analysis. *PLoS Comput Biol*. 2014;10(8):e1003806. <https://doi.org/10.1371/journal.pcbi.1003806>.
49. Van P, Jiang W, Gottardo R, Finak G. ggCyt: next generation open-source visualization software for cytometry. *Bioinformatics*. 2018;34(22):3951–3. <https://doi.org/10.1093/bioinformatics/bty441>.
 50. Žunar B, Mosrin C, Bénédicti H, Vallée B. Re-engineering of *CUP1* promoter and *Cup2/Ace1* transactivator to convert *Saccharomyces cerevisiae* into a whole-cell eukaryotic biosensor capable of detecting 10 nM of bioavailable copper. *Biosens Bioelectron*. 2022;214:114502. <https://doi.org/10.1016/j.bios.2022.114502>.
 51. Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*. 2004;428(6983):617–24. <https://doi.org/10.1038/nature02424>.
 52. Russo P, Kalkkinen N, Sareneva H, Paakkola J, Makarow M. A heat shock gene from *Saccharomyces cerevisiae* encoding a secretory glycoprotein. *Proc Natl Acad Sci USA*. 1992;89(9):3671–5. <https://doi.org/10.1073/pnas.89.9.3671>.
 53. Shen X-X, Steenyk JL, LaBella AL, Oplulent DA, Zhou X, Kominek J, et al. Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Sci Adv*. 2020;6(45):eabd0079. <https://doi.org/10.1126/sciadv.abd0079>.
 54. Liaci AM, Förster F. Take me home, protein roads: structural insights into Signal peptide interactions during ER translocation. *Int J Mol Sci*. 2021;22(21):1–19. <https://doi.org/10.3390/ijms222111871>.
 55. Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*. 2012;335(6068):552–7. <https://doi.org/10.1126/science.1215110>.
 56. Ho B, Baryshnikova A, Brown GW. Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* Proteome. *Cell Syst*. 2018;6(2):192–205e193. <https://doi.org/10.1016/j.cels.2017.12.004>.
 57. O'Callaghan Cynthia H, Morris A, Kirby Susan M, Shingler AH. Novel method for detection of β -Lactamases by using a chromogenic cephalosporin substrate. *Antimicrob Agents Chemother*. 1972;1(4):283–8. <https://doi.org/10.1128/AAC.1.4.283>.
 58. Simonen M, Jämsä E, Makarow M. The role of the carrier protein and disulfide formation in the folding of beta-lactamase fusion proteins in the endoplasmic reticulum of yeast. *J Biol Chem*. 1994;269(19):13887–92. [https://doi.org/10.1016/S0021-9258\(17\)36730-3](https://doi.org/10.1016/S0021-9258(17)36730-3).
 59. Fatal N, Karhinen L, Jokitalo E, Makarow M. Active and specific recruitment of a soluble cargo protein for endoplasmic reticulum exit in the absence of functional COPII component Sec24p. *J Cell Sci*. 2004;117(9):1665–73. <https://doi.org/10.1242/jcs.01019>.
 60. Fatal N, Suntio T, Makarow M. Selective protein exit from yeast endoplasmic reticulum in absence of functional COPII coat component Sec13p. *Mol Biol Cell*. 2002;13(12):4130–40. <https://doi.org/10.1091/mbc.02-05-0082>.
 61. Grieve AG, Rabouille C. Golgi bypass: skirting around the heart of classical secretion. *Cold Spring Harb Perspect Biol*. 2011;3(4):1–15. <https://doi.org/10.1101/cshperspect.a005298>.
 62. Teparić R, Lozančić M, Mrša V. Evolutionary overview of molecular interactions and enzymatic activities in the yeast cell walls. *Int J Mol Sci*. 2020;21(23). <https://doi.org/10.3390/ijms21238996>.
 63. Hagman A, Säll T, Compagno C, Piskur J. Yeast make-Accumulate-Consume Life Strategy evolved as a multi-step process that predates the whole genome duplication. *PLoS ONE*. 2013;8(7):e68734. <https://doi.org/10.1371/journal.pone.0068734>.
 64. Stefanini I, Dapporto L, Legras J-L, Calabretta A, Di Paola M, De Filippo C, et al. Role of social wasps in *Saccharomyces cerevisiae* ecology and evolution. *Proc Natl Acad Sci USA*. 2012;109(33):13398–403. <https://doi.org/10.1073/pnas.1208362109>.
 65. Hsu P-H, Chiang P-C, Liu C-H, Chang Y-W. Characterization of Cell Wall Proteins in *Saccharomyces cerevisiae* Clinical isolates elucidates Hsp150p in virulence. *PLoS ONE*. 2015;10(8):e0135174. <https://doi.org/10.1371/journal.pone.0135174>.
 66. Sulkowska JI, Sulkowski P, Szymczak P, Cieplak M. Stabilizing effect of knots on proteins. *Proc Natl Acad Sci USA*. 2008;105(50):19714–9. <https://doi.org/10.1073/pnas.0805468105>.
 67. Taylor WR. Protein knots and fold complexity: some new twists. *Comput Biol Chem*. 2007;31(3):151–62. <https://doi.org/10.1016/j.compbiolchem.2007.03.002>.
 68. De Chiara M, Barré BP, Persson K, Irizar A, Vischioni C, Khaiwal S, et al. Domestication reprogrammed the budding yeast life cycle. *Nat Ecol Evol*. 2022;6(4):448–60. <https://doi.org/10.1038/s41559-022-01671-9>.
 69. Toh -EA, Oguchi T, Matsui Y, Yasunaga S, Nisogi H, Tanaka K. Three yeast genes, PIR1, PIR2 and PIR3, containing internal tandem repeats, are related to each other, and PIR1 and PIR2 are required for tolerance to heat shock. *Yeast*. 1993;9(5):481–94. <https://doi.org/10.1002/yea.320090504>.
 70. Rebuffat S. Ribosomally synthesized peptides, foreground players in microbial interactions: recent developments and unanswered questions. *Nat Prod Rep*. 2022;39(2):273–310. <https://doi.org/10.1039/D1NP00052G>.
 71. DeBenedictis EP, Keten S. Mechanical unfolding of alpha- and beta-helical protein motifs. *Soft Matter*. 2019;15(6):1243–52. <https://doi.org/10.1039/C8SM02046A>.
 72. Makarow M, Hänninen A-L, Suntio T, Bastos RN. Production of Heterologous Proteins in Yeast With the Aid of the Hsp150 Δ Carrier. In: *Yeast Protocol* Edited by Xiao W. Totowa, NJ: Humana Press; 2006: 333–343.
 73. Korber P, Barbaric S. The yeast PHO5 promoter: from single locus to systems biology of a paradigm for gene regulation through chromatin. *Nucleic Acids Res*. 2014;42(17):10888–902. <https://doi.org/10.1093/nar/gku784>.
 74. Halladin DK, Ortega FE, Ng KM, Footer MJ, Mitić NS, Malkov SN, et al. Entropy-driven translocation of disordered proteins through the Gram-positive bacterial cell wall. *Nat Microbiol*. 2021;6(8):1055–65. <https://doi.org/10.1038/s41564-021-00942-8>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.