

RESEARCH

Open Access



A3DyDB: exploring structural aggregation propensities in the yeast proteome

Javier García-Pardo^{1†}, Aleksandra E. Badaczewska-Dawid^{2†}, Carlos Pintado-Grima¹, Valentín Iglesias¹, Aleksander Kuriata³, Sebastian Kmiecik^{3*} and Salvador Ventura^{1*}

Abstract

Background The budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) is a well-established model system for studying protein aggregation due to the conservation of essential cellular structures and pathways found across eukaryotes. However, limited structural knowledge of its proteome has prevented a deeper understanding of yeast functionalities, interactions, and aggregation.

Results In this study, we introduce the A3D yeast database (A3DyDB), which offers an extensive catalog of aggregation propensity predictions for the *S. cerevisiae* proteome. We used Aggrescan 3D (A3D) and the newly released protein models from AlphaFold2 (AF2) to compute the structure-based aggregation predictions for 6039 yeast proteins. The A3D algorithm exploits the information from 3D protein structures to calculate their intrinsic aggregation propensities. To facilitate simple and intuitive data analysis, A3DyDB provides a user-friendly interface for querying, browsing, and visualizing information on aggregation predictions from yeast protein structures. The A3DyDB also allows for the evaluation of the influence of natural or engineered mutations on protein stability and solubility. The A3DyDB is freely available at <http://biocomp.chem.uw.edu.pl/A3D2/yeast>.

Conclusion The A3DyDB addresses a gap in yeast resources by facilitating the exploration of correlations between structural aggregation propensity and diverse protein properties at the proteome level. We anticipate that this comprehensive database will become a standard tool in the modeling of protein aggregation and its implications in budding yeast.

Keywords Protein aggregation, Aggrescan 3D, AlphaFold, Yeast, *Saccharomyces cerevisiae*

Background

The notable conservation of essential cellular structures and pathways such as cell cycle regulation, DNA repair, RNA processing, signal transduction pathways, metabolism, protein quality control mechanisms, or stress response, has positioned the budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) as an ideal eukaryotic model organism. *S. cerevisiae* is also a reference expression system for the heterologous production of recombinant proteins. One of the major advantages of *S. cerevisiae* resides on its well-characterized and fully annotated genome. Indeed, several databases dedicated to yeast have been

[†]Javier García-Pardo and Aleksandra E. Badaczewska-Dawid are co-first authors.

*Correspondence:
Sebastian Kmiecik
sekmi@chem.uw.edu.pl
Salvador Ventura
Salvador.Ventura@uab.cat

¹Institut de Biotecnologia i de Biomedicina (IBB) and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Barcelona 08193, Spain

²Genome Informatics Facility, Office of Biotechnology, Iowa State University, Ames, IA 50011, USA

³Biological and Chemical Research Center, Faculty of Chemistry, University of Warsaw, Pasteura 1, Warsaw 02-093, Poland



© 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.

published, including the *Saccharomyces* Genome Database (SGD) [1], the Yeast Metabolome Database (YMDB) [2] or YEASTRACT+ [3]. Among them, the SGD database stands as the gold-standard repository with a wealth of integrated biological information for this microorganism. It also provides access to analytic tools to explore these data, enabling the discovery of functional relationships between sequence and gene products in fungi and higher organisms. In addition, the availability of comprehensive genomic resources, such as deletion libraries and collections of temperature-sensitive mutants, facilitates the identification and characterization of factors involved in different biological processes [4]. Indeed, *S. cerevisiae* stands out as one of the most widely employed model systems for studying protein aggregation in vivo. Notably, the yeast proteome is relatively small compared to higher eukaryotes, which makes it a privileged model for developing systematic studies and high-throughput screenings targeting protein aggregation [5]. Despite the relevance of this model organism in the field, aggregation propensity predictions are not currently reported in dedicated databases.

The identification of intermolecular interactions mediated by solvent-exposed aggregation-prone regions (APRs) embedded in the protein sequence has proven successful in predicting protein aggregation [6–10]. This approach is particularly suitable in the context of intrinsically disordered proteins (IDPs) or for newly synthesized polypeptide chains, but it often overpredicts when applied to folded proteins. Indeed, in folded proteins the identified APRs are often located within their hydrophobic cores or at inaccessible regions characterized by the presence of highly stable secondary structures [11, 12]. It is now widely accepted that globular proteins aggregate by the spatial clustering of often non-contiguous sequence regions of hydrophobic amino acids, forming structural APRs in the protein surface (STAPs) [13]. The aggregation may happen by local or global structural destabilization [14] or by stochastic fluctuations that lead to the exposure of previously buried APRs [15]. Consequently, considering a protein's spatial environment becomes crucial in understanding the underlying forces driving its aggregation. In this context, we developed the Aggrescan 3D (A3D) algorithm [16–18], which makes use of the experimentally determined Aggrescan's aggregation propensity scale [7, 19] and projects it into a three-dimensional protein structure. The versatility and accuracy of this algorithm have converted it into one of the default methods to study the aggregation of proteins in their natively folded states and how dynamic fluctuations and mutations impact this reaction.

Over the years, accurate predictions of structure-based aggregation propensities in yeast at the proteome level were hampered by the limited availability of structural

information. However, the newly developed AlphaFold2 (AF2) database has released the prediction of thousands of structures from different organisms, including *S. cerevisiae* [20]. The overall quality of these computed models was shown to be comparable to experimentally determined structures [21]. Therefore, the newly reported structural information from AF2 allows the generation of proteome-wide repositories reporting yeast globular proteins' aggregation properties.

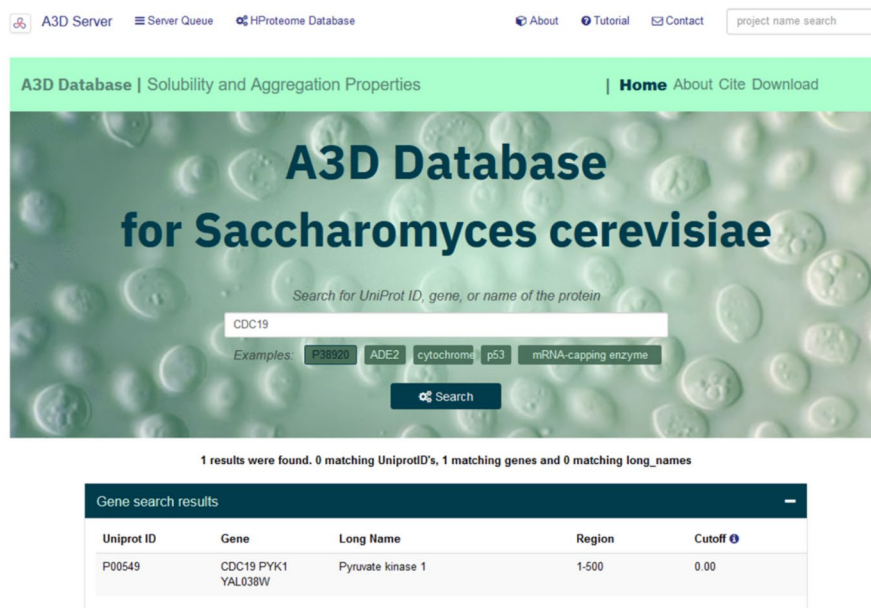
Herein, we present the A3D yeast database (A3DyDB), which compiles the structure-based aggregation propensity predictions for 6039 *S. saccharomyces* protein models from the AF database [20]. Since many yeast proteins are not fully structured, the A3DyDB allows for the customization of jobs to adapt structural predictions according to AF2 confident cutoffs. The database also includes a tool for evaluating the influence of user-defined mutations on protein solubility and stability. We believe that the A3DyDB will serve as a useful resource for the study of protein aggregation in yeast. It will also allow the investigation of correlations between structural aggregation propensity and protein function, stability, architecture, location, and protein abundance, among other factors associated with protein aggregation. Ultimately, we illustrate the performance and utility of the database with selected case reports.

Results

A3DyDB summary and interface description

The A3DyDB incorporates A3D predictions for 6039 proteins from the *S. cerevisiae* proteome. To perform aggregation predictions, we used a large dataset of structural models generated with AF2, which were downloaded from the AF database [20]. These structures were analyzed using the latest A3D implementation developed by our group [17, 22]. The resultant A3D data have been stored in the first comprehensive database describing the structure-based aggregation predictions for yeast (<http://biocomp.chem.uw.edu.pl/A3D2/yeast>). The A3DyDB is endowed with a search tab on the front page, which allows users to query for the content by using the gene, protein name, or Uniprot Accession (See Fig. 1a). Selecting entries from the query list leads to a page containing the A3D analysis. The analysis is distributed in different tabs containing the following information: (I) protein information and project details, (II) an interactive A3D score profile and annotation of transmembrane regions (only applicable to membrane proteins), (III) a detailed table containing the precalculated A3D scores and AF structure prediction confidence scores (pLDDTs), (IV) the protein structure colored by A3D and pLDDTs scores, (V) custom jobs with pre-calculated pLDDT cutoffs and (VI) collection of images. The A3DyDB *Project details* tab contains relevant information regarding the

a



A3D Server Server Queue HPProteome Database About Tutorial Contact project name search

A3D Database | Solubility and Aggregation Properties | Home About Cite Download

A3D Database for *Saccharomyces cerevisiae*

Search for UniProt ID, gene, or name of the protein

CDC19

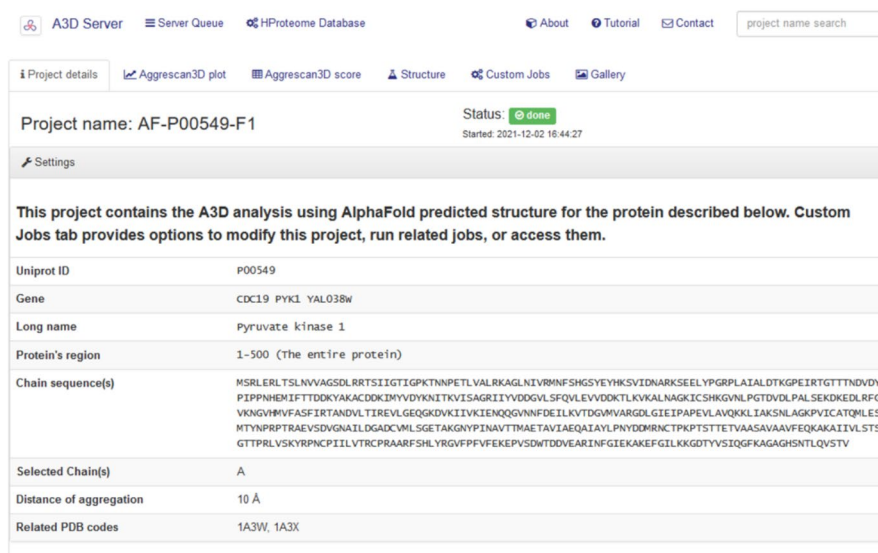
Examples: P38920 ADE2 cytochrome p53 mRNA-capping enzyme

Search

1 results were found. 0 matching UniProtID's, 1 matching genes and 0 matching long_names

| UniProt ID | Gene | Long Name | Region | Cutoff |
|------------|--------------------|-------------------|--------|--------|
| P00549 | CDC19 PYK1 YAL038W | Pyruvate kinase 1 | 1-500 | 0.00 |

b



A3D Server Server Queue HPProteome Database About Tutorial Contact project name search

Project details Aggrescan3D plot Aggrescan3D score Structure Custom Jobs Gallery

Project name: AF-P00549-F1 Status: done Started: 2021-12-02 16:44:27

Settings

This project contains the A3D analysis using AlphaFold predicted structure for the protein described below. Custom Jobs tab provides options to modify this project, run related jobs, or access them.

| | |
|-------------------------|--|
| UniProt ID | P00549 |
| Gene | CDC19 PYK1 YAL038W |
| Long name | Pyruvate kinase 1 |
| Protein's region | 1-500 (The entire protein) |
| Chain sequence(s) | MSRLRLTSLNVAAGSDLRRTSIIIGTIGPKTNNPETLVALRKAGLNIVRMNFSHGSEYHKSVIDNARKSEELYRGRPLAIALDTKGEIRITGTTNDVDYPIPPNHEMIFTDDKYAKACDOKIMYDYKNTKVISAGRIIYVDDGVLDFQLEVVDDKTLKVKALNAGKICSHKGVNLPQTDVLPALSEKDKEDLRFQVKNQGHMVFASFIRANDVLTIREVLGEQKDKVKKIIVKIEHQQGVNFDLKVYTDGVMVARGDLGIEIPAPEVLAQKLLIAKSNLAKGKPVICATQMLSEMTYNNRPRTRAEVSDVGNAILDGADCVMLSGETAKGNYPINAVTTMAETAVIAEQAIYALPNYDDMRNCTPKPTSTTETVAASAAVAVEQKAKAIIIVLSTSGTTPRLVSKYRPNCPILVTRCPRAARFSHLYRGVFPVFEKEPVSDDTDDVEARINFGIEKAKEFGILKKGDTYYSIQGFKAGAGHSNTLQVSTV |
| Selected Chain(s) | A |
| Distance of aggregation | 10 Å |
| Related PDB codes | 1A3W, 1A3X |

Fig. 1 Overview of the A3DyDB home page and example of a search result for CDC19. (a) A3DyDB home page and query result for CDC19. (b) A3DyDB results page showing the project details tab for the CDC19 entry (UniProt accession P00549)

selected entry (Fig. 1b). *A3D plot/score* tabs display a detailed analysis of the per-residue aggregation propensity scores (A3D scores) and the profile and additional annotations for transmembrane and intermembrane proteins can be found in the subsequent sections. For each entry, two different structural models are provided in the *Structure* tab (top panel in Fig. 2a). The top structure shows a per-residue prediction of the A3D aggregation propensity (A3D score), while at the lower structural representation AF2 confidence score (pLDDT) is depicted (lower panel in Fig. 2a).

Membrane proteins were also included in the database as A3D accurately predicts hydrophobic transmembrane segments as highly aggregation-prone regions (Fig. 2b). Although they do not contribute to typical aggregation mechanisms, these highly hydrophobic stretches are often inserted in membrane lipidic bilayers or involved in protein oligomerization. Based on UniProt annotations, a total of 1219 transmembrane and intramembrane proteins from *S. cerevisiae* were identified. For these cases, we have included a complementary tab with relevant information, including predictions of consensus

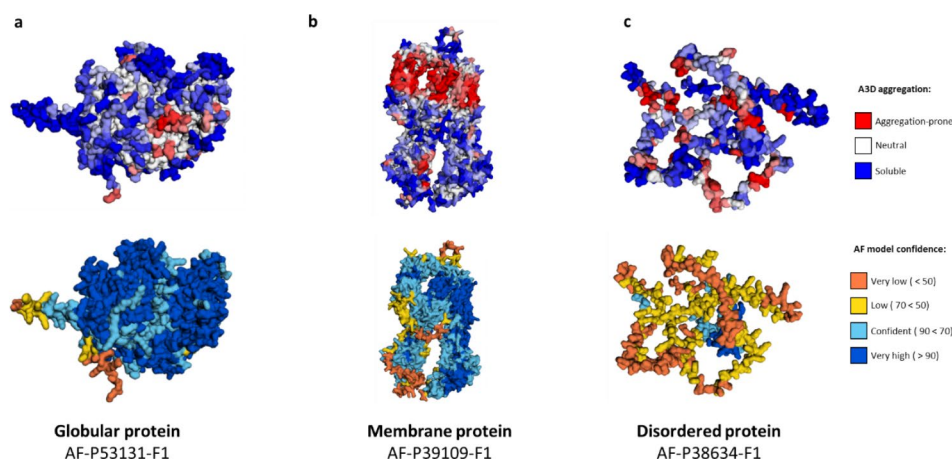


Fig. 2 Examples of protein structural models as reported in the A3D Yeast database. Visualization of A3D predicted aggregation propensity ranges from blue (more solubilizing) to red (more aggregation-prone). A3D Yeast database provides the per-residue local confidence score (pDDLT), a metric that has been shown to inversely correlate with protein disorder [29]. Recent analysis suggests pDDLT scores < 70 usually correspond to unstructured in isolation. Examples of the A3D Yeast database output for (a) a globular protein, (b) a membrane protein and (c) a disordered protein are shown

membrane segments from protein sequences by means of the TOPCONS server [23]. TOPCONS reports five different predictive algorithms, including OCTOPUS [24], Philius v [25], Polyphobius [26], SCAMPI [27], and SPOCTOPUS algorithms [28]. The inclusion of this module will facilitate the visual inspection and analysis of transmembrane regions, with the aim of performing comparative analyses.

A significant number of proteins from the A3DyDB contain structurally disordered regions. This is consistent with the observation that eukaryotes have a higher proportion of IDPs relative to bacteria or archaea [30]. AlphaFold2 outputs the pDDLT score, a per-residue estimate of the model's confidence on a scale from 0 to 100 [20] which reports on the quality of the AF prediction [21]. Interestingly, regions of low confidence often correspond to intrinsically disordered regions (IDRs) [29]. Not surprisingly, manual data curation revealed that low pLDDT scores might result in misleading A3D predictions, either because these solvent-exposed regions are more exposed or compact in the model than they should be. A significant number of yeast proteins showed the presence of regions with low predicted accuracy (low pLDDT score in Fig. 2c). After testing different pLDDT thresholds, we decided to precompute A3D on top of three different AF models for each protein entry: the full-length protein model and two additional models in which residues with pLDDT < 70 or residues with pLDDT < 50 were removed (see Figure S1). These precalculated models can be directly accessed from the *Custom Jobs* tab. The A3DyDB implementation also allows users to model the effects of custom mutations on the stability and aggregation propensity of a given particular protein entry using the FoldX force field [31]. Using the mutation editor at the *Custom Jobs* tab it is possible to evaluate the effects

of single or multiple mutations in a custom A3D analysis. These new jobs will be immediately listed and accessible in the A3DyDB queue.

Discussion

Protein aggregation is increasingly recognized as a contributing factor to various pathologies in eukaryotes [32] and constitutes a major limitation to produce functional recombinant proteins in yeast [33]. Numerous past studies, mostly performed using simple prokaryotic and eukaryotic model organisms such as bacteria and yeast, have led to a detailed understanding of how highly aggregation-prone proteins form insoluble species and how these proteins are toxic for the cells [34]. These seminal investigations have allowed the identification of important principles of protein aggregation, which has led during the last decade to the development of a series of predictive algorithms to identify aggregation-prone sites [35]. Our current understanding of the structural landscape of the yeast proteome has radically changed with the development of deep-learning-based approaches, such as RoseTTA fold [36] or AF2 [21]. This new wealth of structural data can be exploited to predict the aggregation properties of the whole yeast proteome and undertake the redesign of yeast proteins to improve their solubility and stability for diverse purposes. Herein, we have launched the A3DyDB, which contains the pre-computed aggregation predictions for the *S. cerevisiae* proteome, and we have tested the performance of the database in a variety of case reports. Below, we provide selected case examples demonstrating the suitability of our comprehensive repository in diverse scenarios.

Case examples

Exploiting the A3DyDB to study cellular organization and metabolism in yeast

Yeast organisms live in a wide range of different environments which require local adaptations to transient conditions [37]. For this reason, proteins have evolved to self-organize in the cellular milieu in response to specific stimuli such as nutrient starvation. Under these circumstances, metabolic enzymes from yeast proteins undergo a widespread reorganization into reversible punctate cytoplasmic foci that are disassembled when the stress is released [38]. Interestingly, the evidence strongly suggests that the formation of reversible protein assemblies, specific to metabolites, is potentially widespread in the realm of cell biology [38]. In this context, the ability of

proteins to form assemblies is closely linked to their propensity for aggregation. Linear aggregation predictors such as TANGO [39] have been used to study differences in protein aggregation in these dynamic assemblies [38], but most of the proteins involved in foci formation contained globular regions which require dedicated structure-based aggregation predictive tools.

Herein, we used the data from the A3DyDB to investigate possible differences in structural aggregation between the 180 foci and 27 non-foci forming proteins described by Narayanaswamy et al. [38] (Supplementary Table 1). Our results showed that foci-forming proteins had a significantly higher A3D average score than non-foci-forming proteins (Fig. 3a). Indeed, the visual inspection of proteins from the two independent datasets

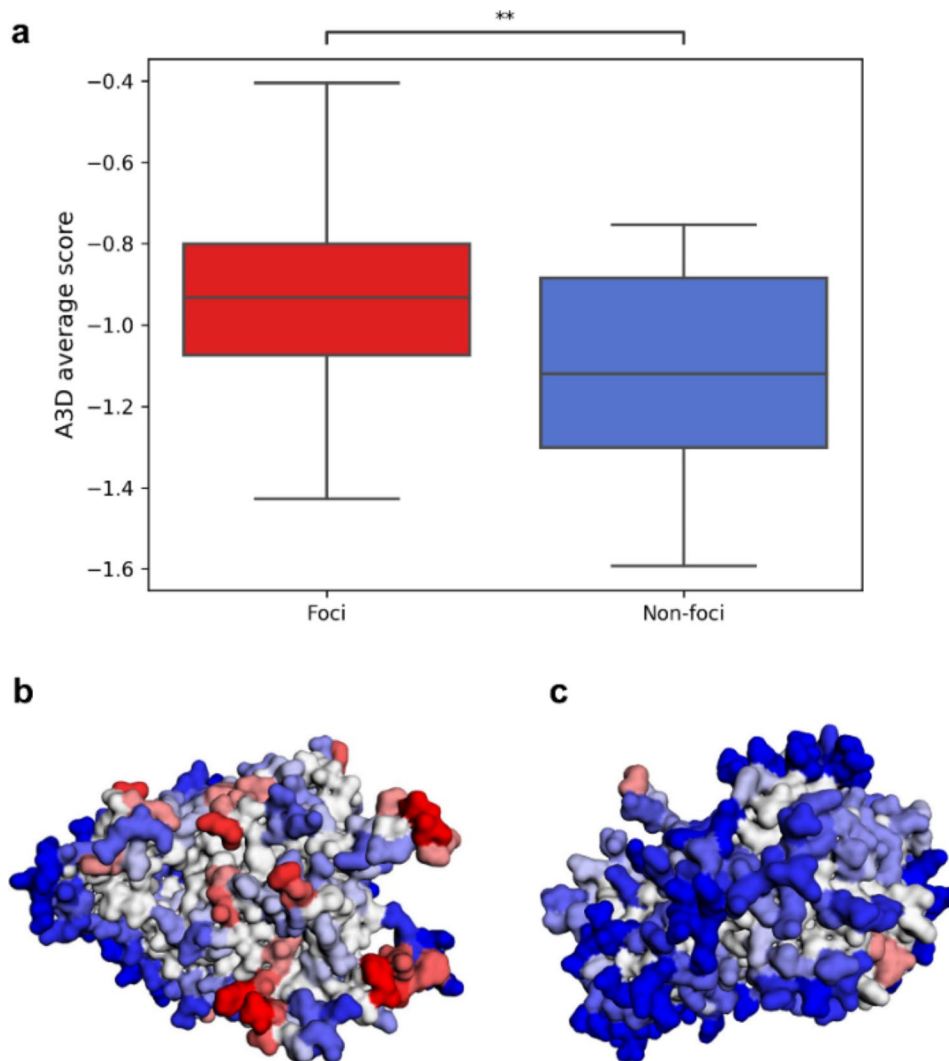


Fig. 3 Structural aggregation propensity differences between foci and non-foci forming proteins from yeast. **(a)** The A3D average score is higher for proteins observed to form punctate foci (in red) than those that did not form (in blue). A statistically significant difference between the two groups was observed (Mann-Whitney-Wilcoxon two-sided test- t ; $p=0.0047$). The A3D predictions from AF ARO2 (AF-P28777-F1) **(b)** and TIF2 (AF-P10081-F1) **(c)** structures, two representative cases of foci and non-foci proteins respectively. The foci-forming protein ARO2 presents a larger number of STAPs than the non-foci protein TIF2.

revealed that foci proteins, such as ARO2, contain a larger number of STAPs than non-foci proteins like TIF2 (Fig. 3b and c, respectively). Overall, it seems that protein organization in yeast is a very dynamic process that could be, at least, partially understood by protein aggregation. In this framework, the release of thousands of structural predictions in the yeast A3D database can help in identifying STAPs in several proteins involved in cellular reorganizations.

Predicting STAPs to study functional protein assemblies

We have investigated a case example in which STAPs are important for mediating functional protein-protein interactions. The actin fold is found in cytoskeletal polymers, chaperones, and various yeast enzymes involved in metabolic pathways. Most actin-fold proteins, such as the carbohydrate kinases are monomeric proteins and do not polymerize. However, it has been recently reported that the *S. cerevisiae* glucokinase GLK1 can form polymers in response to its substrates and products (Fig. 4a) [40].

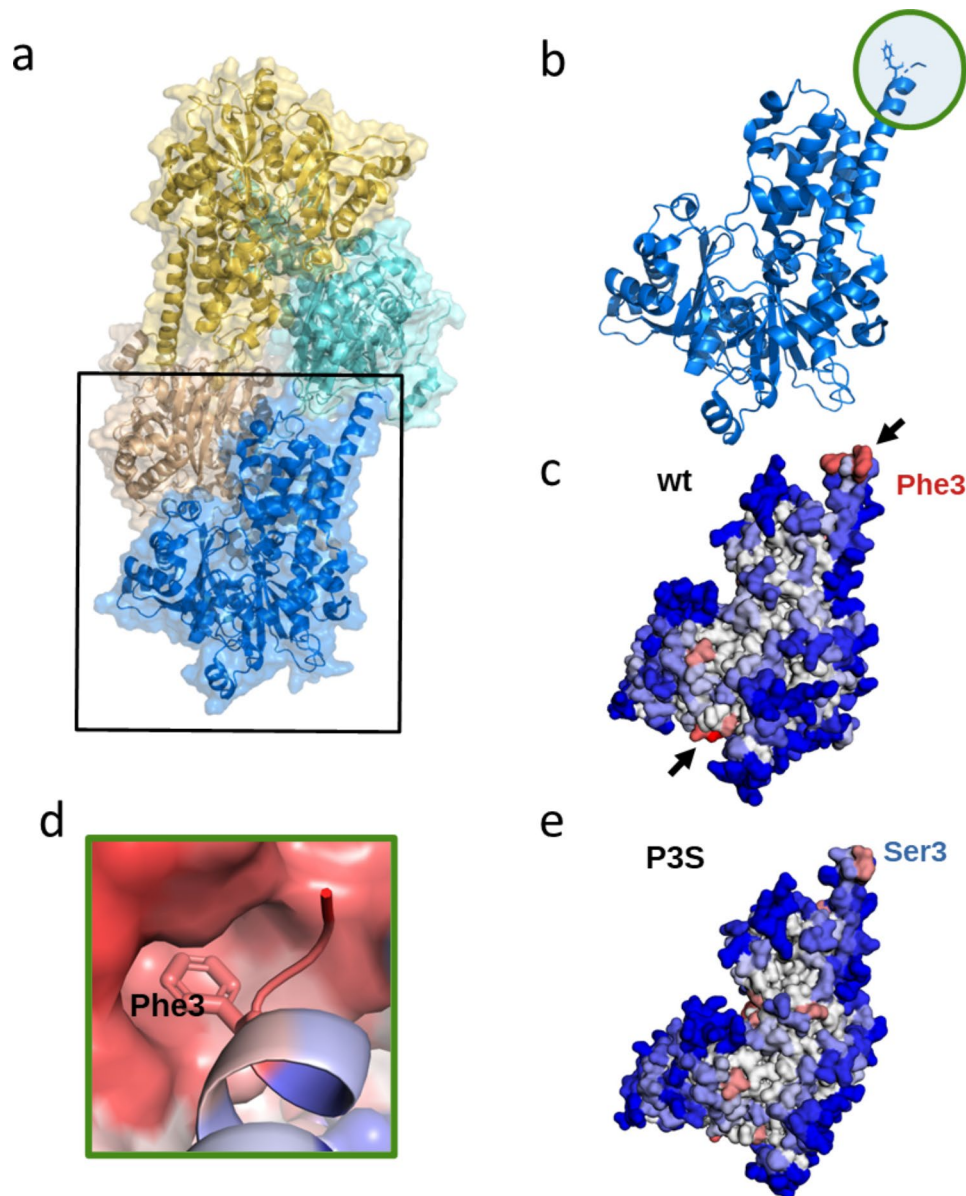


Fig. 4 A3D predictions for yeast hexokinase GLK1 self-assembly. **(a)** Cryo-EM structure of the yeast hexokinase GLK1 filament (PDB: 6PDT). **(b)** Structure of a GLK1 monomer showing the location of the N-terminal Phenylalanine residue involved in filament formation (Phe3, indicated with a circle). **(c)** A3D prediction of the wild-type active monomeric GLK1 obtained from the A3DyDB. Predicted aggregation propensities are colored in the structure ranging from blue (more solubilizing) to red (more aggregation-prone). The position of the two predicted hydrophobic STAPs is indicated with arrows. **(d)** Detailed representation of the insertion site of Phe3 into the C-terminus hydrophobic cavity. **(e)** A3D prediction of the Phe3Ser (P3S) solubilizing mutant. This solubilizing mutation eliminates filament polymerization

The polymerization of this actin-fold protein inhibits its kinase catalytic activity, a mechanism directly coupled to cell viability and the adaptation of the yeast to stochastic changes in the environment. Recently, cryo-EM studies have revealed that glucokinase GLK1 from *S. cerevisiae* is able to form two-stranded filaments with a molecular architecture different from that of cytoskeletal polymers [40]. These filaments are built up by GLK1 monomer stabilized by hydrophobic interactions between GLK1 subunits along a strand. In this structure, a solvent exposed Phe3 of one GLK1 subunit is inserted into the hydrophobic pocket at the C-terminus of the next GLK1 moiety, effectively mediating the stabilization of the filament (Fig. 4b).

A detailed analysis of the structural aggregation propensity of the monomeric GLK1 was obtained from the AF2-derived model collected in A3DyDB (Fig. 4c). Our analysis showed the presence of two strong STAPs in this protein that overlap with the observed GLK1 oligomerization interfaces, where the N-terminus Phe3 is inserted in the hydrophobic C-terminus cavity (Fig. 4d). This is consistent with the view that functional and aberrant polymerization surfaces share very similar physicochemical properties and do frequently overlap [13, 41]. Garner

and coworkers mutated the N-terminal Phe3 involved in GLK1 assembly contacts to Ser, to change the non-polar character of this [21] protein position [40]. Based on our A3D predictions, this mutation significantly decreases the potency of the N-terminal STAP (Fig. 4e), which coincides with the experimental observation that it eliminated polymerization both in vitro and in vivo.

Using the A3DyDB to study membrane proteins

The endoplasmic reticulum (ER) network is built up by tubules with high membrane curvature in cross-section, which are generated and stabilized by reticulons and receptor expression-enhancing proteins (REEPs). Reticulons and REEPs are integral membrane proteins resident at the ER that are evolutionary conserved across all eukaryotes [42, 43]. These proteins share a common architecture that has been also identified in other human proteins that function as ER-phagy intramembrane receptors (i.e. ATG40 in *S. cerevisiae* and FAM134B in mammals) [44, 45]. Here, we used the A3DyDB to investigate the structure and aggregation propensities of YOP1, as a model membrane protein (Fig. 5).

YOP1 is a yeast reticulon that is highly enriched in the tubular portions of the ER and virtually excluded from

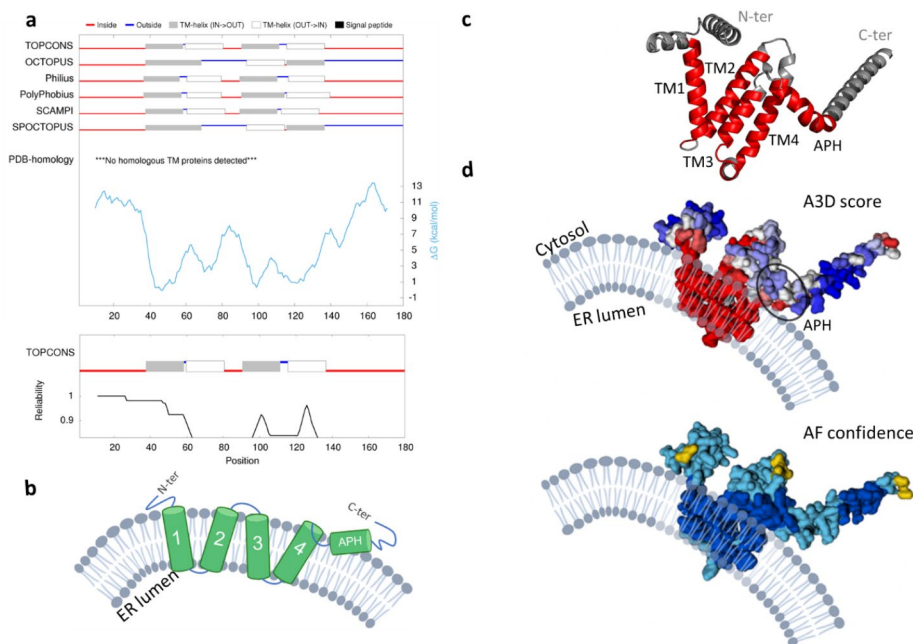


Fig. 5 A3DyDB allows the study of membrane proteins and their functional implications. **(a)** A3DyDB results page or Transmembrane regions tab for YOP1 entry (Uniprot Accession Q12402). **(b)** Scheme of YOP1 predicted topology deduced from A2DyDB TOPCONS predictions and previous experimental data [44, 46]. **(c)** YOP1 AF2 model (AF-Q12402-F1), showing the location of predicted TMs and APH regions. **(d)** A3D analysis (upper panel) and AlphaFold2 (AF2) model confidence score (lower panel). Note that YOP1 exposes STAPs to the hydrophobic interior of the membrane bilayer and at the APH region

other regions, which induces high membrane curvature of tubules by an unknown mechanism. Indeed, deletion of the reticulons and YOP1 in yeast has been linked to the loss of tubular ER. Previous studies have suggested that YOP1 generates high membrane curvature by hydrophobic insertion and scaffolding mechanisms, which is mediated by the insertion of the TM regions on the highly hydrophobic lipid bilayers [47–49]. The suggested mechanism relies on functional data and the protein's structure, but its experimental verification remains incomplete at this stage.

We have analyzed the presence of membrane segments at the transmembrane regions tab from the A3DyDB output. We found that YOP1 has two pairs of trans-membrane (TM) segments, which is one of the characteristic features of reticulons (Fig. 5a). These TM regions are well predicted by the different algorithms implemented in the database and by the consensus TOPCONS prediction, which is accompanied by an overall high prediction reliability (Fig. 5a). Next, we have compared the structure-based aggregation predictions from A3D with the proposed membrane topology. Most of the TM regions were predicted as large highly aggregation-prone segments, as expected for such hydrophobic regions. The main four STAPs in YOP1 were followed by an additional strong short aggregation-prone stretch (residues Ile143-Ile152), a region that matches with an amphipathic helix (APH) that is required for maintaining the characteristic reticulon's ER-tubule localization (Fig. 5c and d). Both the TMs and APH have been shown as essential elements to generate high membrane curvature and for maintaining relevant protein-protein interactions [46]. It has been previously demonstrated that YOP1 undergoes homotypic and heterotypic oligomerization [46, 47, 50, 51]. This behavior was mostly due to homotypic interactions mostly between the TMs regions, as suggested by Cystein-based crosslinking experiments [50, 51]. Taken together, our results indicate that STAPs are present in YOP1, and it is likely that these regions are involved in membrane binding, as well as in maintaining YOP1 oligomerization interaction interfaces.

Conclusions

The A3DyDB provides a unique repository of structural aggregation predictions for thousands of yeast structures collected in the AF database, a resource that for a long time remained elusive due to the limited amount of available structural data. Given the importance of *S. cerevisiae* as a model organism in aggregation, we see A3DyDB as a valuable resource to inspect STAPs in yeast proteins and find associations between aggregation propensity and other functional aspects of yeast biology. Besides, A3DyDB can be used to analyze the effect of mutations on the 3D surface of proteins and engineer variants that

could become more stable and soluble. The presented here *in silico* approach can serve to make faster and more cost-efficient yeast mutants for different applications such as reconstructing metabolic networks, improving the solubility of endogenous proteins, recombinant protein production, and anticipating improved protein variants in synthetic biology approaches.

Methods

Data collection and A3D analysis

S. cerevisiae (UP000002311) protein structures (n=6039) were downloaded from the AF database (October 20, 2022; structural model version v4, available at https://ftp.ebi.ac.uk/pub/databases/alphafold/v4/UP000002311_559292_YEAST_v4.tar) and run with A3D in static mode with a distance of aggregation analysis of 10Å and FoldX-based energy minimization for stability calculations. Custom jobs were created for all predicted structures with two defined AF cutoffs 50 and 70, with residues of pLDDT ≤ 50 or ≤ 70 removed for the A3D aggregation prediction.

Database construction

The user interface of the A3DyDB online database was developed utilizing HTML and integrated with custom JavaScript functions to enhance interactivity. The visual design of the website is a combination of standard Bootstrap components along with a touch of custom CSS styles. An Apache2 web server is employed to host the website, leveraging MySQL integration to handle data storage, retrieval, and querying of the pre-calculated A3DyDB entries. Interactive plots are dynamically generated using the D3.js library, while molecular visualization tasks are handled by the Open Source PyMOL tool [52]. Additionally, the database is seamlessly integrated with the A3D Server, enabling direct submission of custom mutation analysis. The transmembrane analysis for membrane proteins was performed with the consensus algorithm TOPCONS [23], which includes predictions from OCTOPUS [24], Philius [25], Polyphobius [26], SCAMPI [27], and SPOCTOPUS [28].

Foci and mutation structural analyses

Foci (n=180) and non-foci (n=27) forming proteins were obtained from [38]. Structural aggregation propensities for both protein datasets were obtained from the yeast A3DyDB. Statistical significance between variables and/or datasets was assessed with Mann-Whitney-Wilcoxon two-sided test with Bonferroni correction. p-value was marked with asterisks to better convey statistical significance (p > 0.05 (ns), p ≤ 0.05 (*), p ≤ 0.01 (**), p ≤ 0.001 (***), p ≤ 0.0001 (****)).

GLK1 mutation analysis was performed with A3DyDB mutation mode under the Custom jobs tab. Structural

representations of A3D mutants were obtained from the Structure tab visualization tool.

Acknowledgements

Not applicable.

Authors' contributions

JG-P, CP-G, and VI performed the analyses and wrote the initial draft of the manuscript. AEB-D, AK and SK contributed with the implementation of the database. JG-P, SK and SV acquired funding and reviewed the final version of the manuscript with inputs from all authors. JG-P, SK and SV conceptualized the project. All authors have read and agreed to the final version of the manuscript.

Funding

This work was funded by the Spanish Ministry of Science and Innovation (MICINN) PID2019-105017RB-I00 to S.V. by ICREA, ICREA-Academia 2020 and by EU (PhasAge/H2020-WIDESPREAD-2020-5) to S.V. J.G.-P. was supported by the Spanish Ministry of Science and Innovation with a Juan de la Cierva Incorporación IJC2019-041039-I.VI. was supported by the Spanish Ministry of Science and Innovation and the European Union-NextGenerationEU (Universitat Autònoma de Barcelona 02/07/2021). C.P.-G. was supported by the Secretariat of Universities and Research of the Catalan Government and the European Social Fund (2023 FI_3 00018). This article is partially based upon work from COST Action ML4NGP, CA21160, supported by COST (European Cooperation in Science and Technology). SK acknowledges financial support by National Science Centre, Sheng grant number 2021/40/Q/NZ2/00078.

Data Availability

All data generated or analyzed during this study are included in this published article and the A3DyDB is freely available at <http://biocomp.chem.uw.edu.pl/A3D2/yeast>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 30 June 2023 / Accepted: 18 August 2023

Published online: 16 September 2023

References

1. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* 2012;40(Database issue):D700–5.
2. Ramirez-Gaona M, Marcu A, Pon A, Guo AC, Sajed T, Wishart NA, et al. YMDB 2.0: a significantly expanded version of the yeast metabolome database. *Nucleic Acids Res.* 2017;45(D1):D440–D5.
3. Monteiro PT, Oliveira J, Pais P, Antunes M, Palma M, Cavalheiro M, et al. YEAST-RACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. *Nucleic Acids Res.* 2020;48(D1):D642–D9.
4. Jin K, Li J, Vizeacoumar FS, Li Z, Min R, Zamparo L, et al. PhenoM: a database of morphological phenotypes caused by mutation of essential genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 2012;40(Database issue):D687–94.
5. Di Gregorio SE, Duennwald ML. Yeast as a model to study protein misfolding in aged cells. *FEMS Yeast Res.* 2018;18(6).
6. Belli M, Ramazzotti M, Chiti F. Prediction of amyloid aggregation in vivo. *EMBO Rep.* 2011;12(7):657–63.
7. Conchillo-Sole O, de Groot NS, Aviles FX, Vendrell J, Daura X, Ventura S. AGGR-ESCAN: a server for the prediction and evaluation of hot spots of aggregation in polypeptides. *BMC Bioinformatics.* 2007;8:65.
8. Maurer-Stroh S, Debulpaep M, Kuemmerer N, Lopez de la Paz M, Martins IC, Reumers J, et al. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods.* 2010;7(3):237–42.
9. Walsh I, Seno F, Tosatto SC, Trovato A. PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.* 2014;42(Web Server issue):W301–7.
10. Santos J, Pujols J, Pallares I, Iglesias V, Ventura S. Computational prediction of protein aggregation: advances in proteomics, conformation-specific algorithms and biotechnological applications. *Comput Struct Biotechnol J.* 2020;18:1403–13.
11. Castillo V, Chiti F, Ventura S. The N-terminal helix controls the transition between the soluble and amyloid states of an FF domain. *PLoS ONE.* 2013;8(3):e58297.
12. Santos J, Iglesias V, Ventura S. Computational prediction and redesign of aberrant protein oligomerization. *Prog Mol Biol Transl Sci.* 2020;169:43–83.
13. Castillo V, Ventura S. Amyloidogenic regions and interaction surfaces overlap in globular proteins related to conformational diseases. *PLoS Comput Biol.* 2009;5(8):e1000476.
14. Castillo V, Espargaro A, Gordo V, Vendrell J, Ventura S. Deciphering the role of the thermodynamic and kinetic stabilities of SH3 domains on their aggregation inside bacteria. *Proteomics.* 2010;10(23):4172–85.
15. Grana-Montes R, de Groot NS, Castillo V, Sancho J, Velazquez-Campoy A, Ventura S. Contribution of disulfide bonds to stability, folding, and amyloid fibril formation: the PI3-SH3 domain case. *Antioxid Redox Signal.* 2012;16(1):1–15.
16. Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S. AGGR-ESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Res.* 2015;43(W1):W306–13.
17. Kuriata A, Iglesias V, Pujols J, Kurcinski M, Kmiecik S, Ventura S. Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility. *Nucleic Acids Res.* 2019;47(W1):W300–W7.
18. Pujols J, Iglesias V, Santos J, Kuriata A, Kmiecik S, Ventura S. A3D 2.0 update for the prediction and optimization of protein solubility. *Methods Mol Biol.* 2022;2406:65–84.
19. Badaczewska-Dawid AE, Garcia-Pardo J, Kuriata A, Pujols J, Ventura S, Kmiecik S. A3D database: structure-based predictions of protein aggregation for the human proteome. *Bioinformatics.* 2022;38(11):3121–3.
20. 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–D44.
21. 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.
22. Kuriata A, Iglesias V, Kurcinski M, Ventura S, Kmiecik S. Aggrescan3D stand-alone package for structure-based prediction of protein aggregation properties. *Bioinformatics.* 2019;35(19):3834–5.
23. Tsirigos KD, Peters C, Shu N, Kall L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* 2015;43(W1):W401–7.
24. Viklund H, Elofsson A. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics.* 2008;24(15):1662–8.
25. Reynolds SM, Kall L, Riffle ME, Bilmes JA, Noble WS. Transmembrane topology and signal peptide prediction using dynamic bayesian networks. *PLoS Comput Biol.* 2008;4(11):e1000213.
26. Kall L, Krogh A, Sonnhammer EL. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics.* 2005;21(Suppl 1):i251–7.
27. Bernsel A, Viklund H, Falk J, Lindahl E, von Heijne G, Elofsson A. Prediction of membrane-protein topology from first principles. *Proc Natl Acad Sci U S A.* 2008;105(20):7177–81.
28. Viklund H, Bernsel A, Skwark M, Elofsson A. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics.* 2008;24(24):2928–9.
29. Piovesan D, Monzon AM, Tosatto SCE. Intrinsic protein disorder and conditional folding in AlphaFoldDB. *Protein Sci.* 2022;31(11):e4466.
30. Pancsa R, Tompa P. Structural disorder in eukaryotes. *PLoS ONE.* 2012;7(4):e34687.
31. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res.* 2005;33(Web Server issue):W382–8.

32. Chiti F, Dobson CM. Protein misfolding, amyloid formation, and Human Disease: a Summary of Progress over the last decade. *Annu Rev Biochem.* 2017;86:27–68.
33. Romero-Suarez D, Wulff T, Rong Y, Jakociu Nas T, Yuzawa S, Keasling JD, et al. A reporter system for cytosolic protein aggregates in yeast. *ACS Synth Biol.* 2021;10(3):466–77.
34. Ibstedt S, Sideri TC, Grant CM, Tamas MJ. Global analysis of protein aggregation in yeast during physiological conditions and arsenite stress. *Biol Open.* 2014;3(10):913–23.
35. Pintado-Grima C, Barcenas O, Bartolomé-Nafria A, Fornt-Suñe M, Iglesias V, García-Pardo J, et al. A review of Fifteen Years developing computational tools to study protein aggregation. *Biophysica.* 2023;3(1):1–20.
36. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* 2021;373(6557):871–6.
37. Dhar R, Sagesser R, Weikert C, Wagner A. Yeast adapts to a changing stressful environment by evolving cross-protection and anticipatory gene regulation. *Mol Biol Evol.* 2013;30(3):573–88.
38. Narayanaswamy R, Levy M, Tsechansky M, Stovall GM, O'Connell JD, Mirrielees J, et al. Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proc Natl Acad Sci U S A.* 2009;106(25):10147–52.
39. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol.* 2004;22(10):1302–6.
40. Stoddard PR, Lynch EM, Farrell DP, Dosey AM, DiMaio F, Williams TA, et al. Polymerization in the actin ATPase clan regulates hexokinase activity in yeast. *Science.* 2020;367(6481):1039–42.
41. Pechmann S, Levy ED, Tartaglia GG, Vendruscolo M. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc Natl Acad Sci U S A.* 2009;106(25):10159–64.
42. Chen S, Novick P, Ferro-Novick S. ER structure and function. *Curr Opin Cell Biol.* 2013;25(4):428–33.
43. Yang YS, Strittmatter SM. The reticulons: a family of proteins with diverse functions. *Genome Biol.* 2007;8(12):234.
44. Khaminets A, Heinrich T, Mari M, Grumati P, Huebner AK, Akutsu M, et al. Regulation of endoplasmic reticulum turnover by selective autophagy. *Nature.* 2015;522(7556):354–8.
45. Bhaskara RM, Grumati P, García-Pardo J, Kalayil S, Covarrubias-Pinto A, Chen W, et al. Curvature induction and membrane remodeling by FAM134B reticulon homology domain assist selective ER-phagy. *Nat Commun.* 2019;10(1):2370.
46. Brady JP, Claridge JK, Smith PG, Schnell JR. A conserved amphipathic helix is required for membrane tubule formation by Yop1p. *Proc Natl Acad Sci U S A.* 2015;112(7):E639–48.
47. Voeltz GK, Prinz WA, Shibata Y, Rist JM, Rapoport TA. A class of membrane proteins shaping the tubular endoplasmic reticulum. *Cell.* 2006;124(3):573–86.
48. Campelo F, McMahon HT, Kozlov MM. The hydrophobic insertion mechanism of membrane curvature generation by proteins. *Biophys J.* 2008;95(5):2325–39.
49. Wang N, Clark LD, Gao Y, Kozlov MM, Shemesh T, Rapoport TA. Mechanism of membrane-curvature generation by ER-tubule shaping proteins. *Nat Commun.* 2021;12(1):568.
50. Shibata Y, Voss C, Rist JM, Hu J, Rapoport TA, Prinz WA, et al. The reticulon and DP1/Yop1p proteins form immobile oligomers in the tubular endoplasmic reticulum. *J Biol Chem.* 2008;283(27):18892–904.
51. Xiang Y, Lyu R, Hu J. Oligomeric scaffolding for curvature generation by ER tubule-forming proteins. *Nat Commun.* 2023;14(1):2617.
52. The PyMOL Molecular. Graphics System, Version 2.0 Schrödinger, LLC.

Publisher's Note

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