Exploring the Theoretical Propensity for Liquid-Liquid Phase Separation in the Organization of Golgi Matrix Proteins

Luis Felipe S. Mendes[1], Antonio J. Costa-Filho[2]

[1] Group of Biophysics and Structural Biology "Sergio Mascarenhas". São Carlos Institute of Physics, University of São Paulo, São Carlos-SP, Brazil

[2] Molecular Biophysics Laboratory, Department of Physics, Faculty of Philosophy, Sciences, and Letters at Ribeirão Preto, University of São Paulo, Ribeirão Preto-SP, Brazil

To whom correspondence should be addressed: Luis F. S. Mendes (luis.felipe.mendes@usp.br)

Abstract

The Golgi apparatus, an essential cellular organelle, is fundamental in protein sorting and lipid metabolism. Comprising stacked, flattened sack-like structures, or cisternae, the Golgi displays distinct polarity with cis- and trans-faces that orchestrate different protein maturation and transport stages. Crucial to the Golgi integrity and organisation are the Golgi Matrix Proteins (GMPs), predominantly composed of Golgins and GRASPs. These proteins contribute to the Golgi's distinctive stacked, polarised structure and ensure the appropriate localisation of resident Golgi enzymes, which are key for accurate protein processing. Even though the Golgi apparatus was unveiled over a century ago, the detailed mechanisms that control its structure remain elusive. Here, we highlight the significant prevalence of intrinsically disordered regions within GMPs across Eukarya. Intriguingly, these regions are linked to a high propensity for those proteins to form liquid-like condensates via Liquid-Liquid Phase Separation (LLPS). Biomolecular condensates are crucial in organising intracellular components into membrane-less organelles, such as Lewy bodies, the Corneal lens, and the nucleolus. Our findings pave the way for future investigations into the roles of disorder and LLPS in GMPs. Moreover, they provide a framework for understanding how these proteins might shape the architecture of the Golgi apparatus and the regulatory implications of phase separation in these processes.

Keywords:

Liquid-Liquid Phase Separation, Golgi Complex, Golgins, GRASPs, Intrinsically Disordered Regions

Introduction:

Eukaryotic cells evolved to possess efficient secretory machinery capable of transporting a significant fraction of their proteome. The number of proteins predicted to enter the secretory pathway can reach up to 36% of the total proteome in humans (*1*). Proteins entering the secretory pathway are efficiently sorted to a specific destination: the extracellular space, the plasma membrane, or the interior of the endomembrane system. The machinery also has efficient quality control, avoiding the transport of misfolded proteins or complexes (*2*). Proteins destined for this classical secretory pathway are synthesised in the endoplasmic reticulum (ER) interior, where they are first folded and possibly modified by the post-translation ER machinery (*2*). Subsequently, these secretory cargoes are transferred to the ER exit sites (ERES), where they are assembled into secretory vesicles coated with coatomers (COPII), routered for the ER-Golgi intermediate compartment (ERGIC), and later to the *cis*-Golgi face (*3, 4*).

The Golgi apparatus, a central organelle in eukaryotic cells, plays a pivotal role in trafficking, processing, and sorting proteins and lipids. This membrane-bound organelle exhibits a unique architecture consisting of a stack of flattened cisternae interconnected by tubular-vesicular networks. These cisternae can be further classified into three distinct regions: *cis*, *medial*, and *trans*, each responsible for specific functions in protein processing and glycosylation. The maintenance of this intricate structure is crucial for the Golgi's function and relies on a diverse group of proteins, collectively known as the Golgi matrix proteins (GMPs). In the organisation of the Golgi stacks and cisternae formation, it was observed that in cells treated with brefeldin A (an inhibitor of protein transport from the ER to the Golgi), the Golgi membrane and its integral proteins relocated to the ER by membrane fusion (*5*). On the other hand, the GMPs accumulated in the cytoplasm formed a ribbon-like reticulum in the correct Golgi location (*5*). Based on this observation, it seems that GMPs could generate the higher-order architecture of the Golgi by a mechanism independent of other Golgi proteins. Although an orchestrated action involving all the GMPS provides the total adhesive energy dictating Golgi cisternal stacking (*6*), how these proteins act in the general organisation/dynamics remains unknown.

The Golgins and the Golgi Reassembly and Stacking Proteins (GRASP) are the main protein families composing the GMPs. The Golgins are predominantly extended coiled-coil proteins predicted to reach an extension between 100-600 nm (*7, 8*). They attach to the Golgi cisternae through either their carboxy terminus, directly interacting with the membrane surface (*7*) or a membrane-associated partner (*9*). Golgins participate in vesicular traffic at the Golgi apparatus, Golgi maintenance and dynamics, and the positioning of the Golgi apparatus within cells (*8, 10*). On the other hand, GRASPs constitute a family of peripheral membrane-associated

proteins essential for general Golgi organisation into the ribbon-like structure (*9*) and Golgi integrity for the correct protein sorting and glycosylation along the secretory pathway (*11*). GRASPs are observed in all branches of the eukaryotic tree of life, although not equally spread in all supergroups (*12*). GRASPs are IDPs with high promiscuity for mediating protein-protein interactions with several different protein partners (*13-15*). Their unusual structural plasticity (*16*) might be the converging feature of GRASPs in multiple cell functional tasks, including their direct participation in Golgi dynamics during mitotic and apoptotic times (*17-19*) and their self-association during stress conditions triggering unconventional protein secretion pathways (*20, 21*). GRASPs also anchor some Golgin representatives, including GM130, Golgin45, and p155 in mammalian cells and the Golgin BUG1 in Yeast (*9, 22*). This anchoring is essential for the correct Golgi organisation (*9, 23*).

Focusing on the Golgi complex, it is striking to observe the rapid, dramatic, and reversible structural transformations this organelle undergoes during mitosis. During cell division, the Golgi cisternae disassemble into considerably smaller vesicles, which, subsequently, reassemble in a highly regulated manner within each daughter cell after mitosis (*24*). This process bears a striking resemblance to the behaviour of the nucleolus. Furthermore, an even more comprehensive disassembly of the Golgi complex is governed by specific drugs acting through distinct pathways, with the Golgi reappearing within minutes after removing those drugs (*25, 26*). These observations suggest that the Golgi cisternae are more malleable and dynamic than typical membrane-bound organelles.

Another intriguing aspect of the Golgi complex is its ability to maintain a preferential polarisation towards the endoplasmic reticulum (ER) (Figure 1) despite its components being in a constant state of dynamic movement across the secretory pathway (*27-29*). Therefore, it is intriguing how a sophisticated structure like the Golgi stack can possess a steady-state polarity while most of its components are fluid and interchanging. Moreover, what enables the Golgi to exhibit such remarkable elasticity, allowing for rapid fragmentation and reorganisation into flattened structures? Exploring these issues will undoubtedly enhance our understanding of the unique characteristics and functions of the Golgi complex in cell biology.

In 2009, James Rothman suggested that the Golgi, analogous to the Nucleolus, would fundamentally be liquid with a phase-separated internal organisation (*30*). This suggestion came after discovering that many human Golgins can populate a liquid-like droplet state quite efficiently once they are overexpressed *in vivo* and dispersed in the cytosol (*31*). In recent years, liquid-liquid phase separation (LLPS) has emerged as a critical mechanism for forming and organising biomolecular condensates within cells. LLPS is a process by which a homogenous solution of macromolecules separates into distinct liquid phases, often leading to the assembly of

membrane-less compartments with unique biophysical properties (*32*). This phenomenon has been implicated in cellular functions, such as signal transduction, RNA processing, and stress response (*33-35*).

The remodelling of cell membranes requires their bending. Membrane bending is an essential cellular process with implications in numerous aspects of cell biology, such as membrane traffic, cell motility, organelle creation, and cell division (*36-38*). Disturbances in this process can lead to various diseases, including neurodegenerative disorders, muscular dystrophies, and viral infections (*39-41*). The bending is typically attributed to proteins with specific structural features, including wedge-like amphipathic helices and crescent-shaped scaffolds (*42*). These proteins bind to membranes and induce stress on the membrane surface, promoting curvature (*37*). Besides the traditional bending induced by well-shaped proteins, it has been recently shown that protein phase separation occurring on the surfaces of both synthetic and cell-derived membrane vesicles induces significant compressive stress within the membrane plane, which prompts the membrane to bend inward, forming protein-lined membrane tubules (*37, 43*). In this case, the proteins undergoing phase separation contain substantial regions of intrinsic disorder, which do not possess a stable three-dimensional structure (*43, 44*). Fascinatingly, disordered regions have recently been observed to form networks sustained by weak, multivalent contacts, facilitating the assembly of protein liquid phases on membrane surfaces (*45, 46*).

In the context of LLPS-induced membrane remodelling, it was demonstrated that autophagosomes could envelop droplets containing the p62 protein, and this process could be modelled physically, with droplet surface tension supporting the formation of membrane sheets (*47*). The importance of wetting on membrane morphology was also described recently by Mangiarotti et al. It was found that changes in solution salinity or membrane composition could drive these systems through two wetting transitions, leading to intricate, curved membrane structures due to the “fingering” or “ruffling” of the condensate-membrane interface (*48*). The observed morphologies resulted from a balance of adhesion, membrane elasticity, and interfacial tension (*48*).

Since LLPS can affect membrane bending dramatically (*37, 45, 48*) and some Golgins have been shown to undergo LLPS (*31*), it would be reasonable to argue that there might be a direct correlation between the LLPS-propensity of GMPs and the unusual Golgi cisternae shape. However, there is still a lack of knowledge on the molecular determinants driving LLPS of GMPs, their evolutionary conservation, and the possible interplay between the unique architecture of the Golgi apparatus, its matrix proteins, and the functional relevance of LLPS in maintaining and regulating this essential organelle. Understanding the relationship between LLPS and the GMPs is particularly relevant given the dynamic nature of the Golgi apparatus. Furthermore,

investigating the evolutionary conservation of LLPS propensity in the GMP family across a diverse set of organisms, including those theoretically predicted in the Last Eukaryotic Common Ancestor (LECA), may provide insights into the functional significance of this phenomenon in shaping the Golgi apparatus' structure and its functional evolution.

Results and discussions

Identifying all GMPs in vertebrates is an active research area, and the comprehensive list is not definitively established. An extra difficulty is that different organisms can express different sets of proteins. In this work, we selected some well-known GMPs and associated proteins that maintain Golgi structure and function. It is important to emphasise that we might not have exhausted the full list of GMPs. Some GMPs identified in metazoans are: GM130 (GOLGA2), Golgin-45 (BLZF1), Golgin-160 (GOLGA3), AKAP9 (CG-NAP), Giantin (GOLGB1), GCP60 (ACBD3), GCC88 (GCC1), Golgin-97, Golgin-245 (GOLGA4), SCYL1BP1 (GORAB), GCC185 (GCC2), Bicaudal D (BCD1 and BCD2), NEEC1 (JAKMIP2), GRASP65 (GORASP1) and GRASP55 (GORASP2). These families form the list to be discussed now in the text. Other important GMPs like CASP (GOLGA8), Golgin-84 (GOLGA5), p115 (also known as USO1), Zinc Finger Protein-Like 1 (ZFPL1), and GMAP-210 (also known as TRIP11) are also found in metazoans but are not exclusive, and so will be discussed later in the text in a broader context. It is important to note that different proteins can have multiple names or synonyms, adding to the complexity of cataloguing them all.

Not all the proteins listed in the previous paragraph are Golgins. They typically have coiled-coil regions; many have a C-terminal GRIP domain or a similar structure. In the previous list, proteins like GRASP65 and GRASP55, while crucial to the function and structure of the Golgi, are not classified as Golgins. The same is true for the ZFPL1. The distinction between Golgins and other Golgi-associated proteins is not always clear, and there may be exceptions or debates over how specific proteins are classified. The pervasive presence of Golgins with their long rod-like structures establishes an extensive ribosome-free zone around the Golgi apparatus, a feature readily observable in microscopy images (*49*). Despite the per-residue confidence score (pLDDT) of the Alphafold models of human Golgins primarily falling within the "Confident" to "Very High" range, the observed "Expected Position Error" (EPE) indicates a relatively low level of accuracy at the inter-domain level (AlphaFold Uniprot IDs Q08379, Q9H2G9, Q08378, Q9H3P7, Q96CN9, Q92805, Q13439, Q5T7V8, Q8IWJ2, Q96G01, and Q96AA8). This discrepancy suggests that while the model is confident in the predicted structure of individual

residues and local structures, there is less certainty about the precise positioning and orientation of different structural domains relative to each other. Nevertheless, a consistent theme observed across all vertebrate Golgins is the concomitant presence of extended helical structures, likely the coiled coils mentioned above, and regions of intrinsic disorder. This combination of structured and unstructured domains suggests a level of structural complexity and flexibility that may be essential to the function of these proteins in the Golgi apparatus.

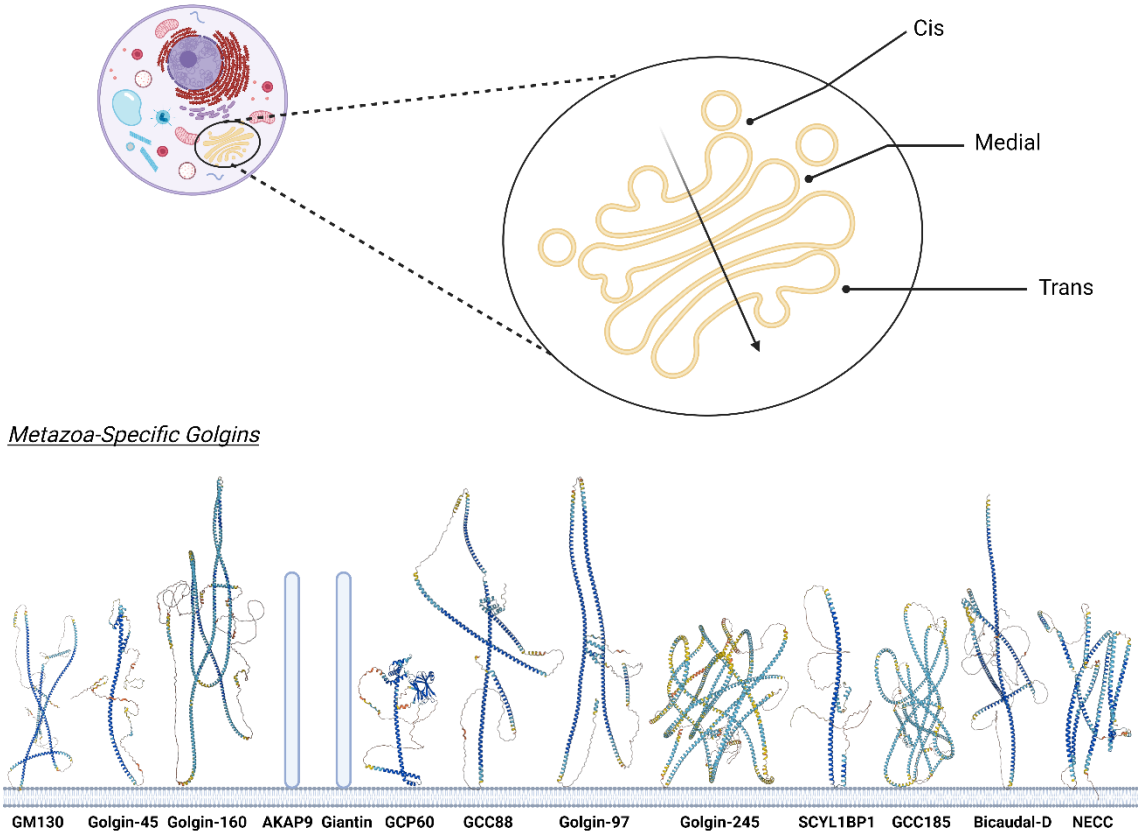

Figure 1: Depiction of a Polarised Golgi Apparatus Stacked Adjacent to the Endoplasmic Reticulum (ER). The black arrow denotes the route of anterograde cargo transport within the conventional secretory pathway. The structural configurations of metazoan-specific Golgins are informed by the AlphaFold model, based on homologues derived from the UniProt database. The Golgins are colour-coded according to the AlphaFold per-residue confidence score (pLDDT). We have chosen the directed orientation towards the membrane surface to highlight the significant cytosolic extension of these proteins, contributing to the formation of a ribosome-exclusion zone. AKAP9 and Giantin are considered too big to be modelled by Alphafold (3907 and 3259 amino acids, respectively).

Despite the widely acknowledged presence of coiled-coil regions in Golgins, the extent of structural disorder within these proteins has received limited attention. We employed the Hydrophobicity versus Net Charge analysis to level that asymmetry. Initially proposed by Uversky, this method provides a binary classification of proteins, serving as an important tool in

our initial investigation (*50*). The method builds a clear boundary (red line in Figure 2) separating "natively disordered proteins" from "ordered proteins". Predominantly used to differentiate extended IDPs from globular proteins, this analytical approach unveiled a critical trait: most GMPs exhibited behaviours akin to extended IDPs (Figure 2). This finding underscores the conservation of sequence features within GMPs, including low net-hydrophobicity and high net charge.

Notable exceptions to this trend are GRASPs and Golgin-45, which appear to reside more frequently on the "well-folded" side of the Uversky plot (Figure 2). GRASPs consist of two primary regions: the structured GRASP domain and a fully-disordered C-terminus, the latter typically being the larger portion of those proteins (*15*). Previous molecular biophysics analyses have indicated GRASPs as molten-globule proteins, suggesting an intermediate level of structural order (*16, 21*). Golgin-45 is the principal binding partner of mammalian GRASP55 (*13, 51*). Unlike GRASP, which is predicted to be present in the Last Eukaryotic Common Ancestor (LECA), Golgin-45 appears to be holozoan-specific (*52*). Prior research has demonstrated that the progenitor of the paralogous proteins GRASP55 and GRASP65 could bind Golgin-45 with high affinity (*13*). In this context, GRASPs and Golgin-45 neither exhibit the mean net charge/hydrophobicity ratio characteristic of extended IDPs, like the other GMPs, nor fit in the classic well-folded realm of protein structure.

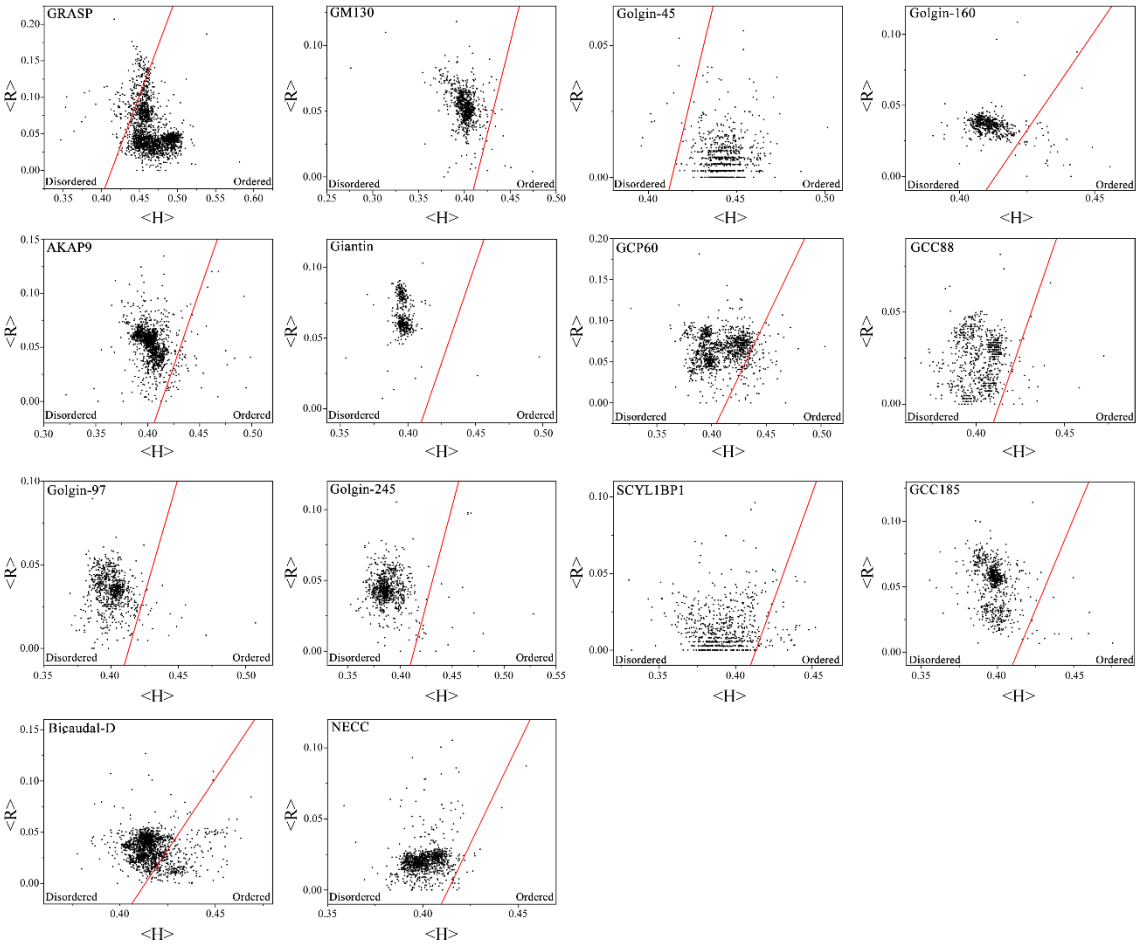

Figure 2: Charge-hydrophobicity phase space of the individual GMPs database where each black dot represents a specific GMP. The solid red line represents the border between "natively disordered" and "native ordered" proteins calculated according to the equation $< R > = 2.785 < H > - 1.151$. In this work, $< H >$ and $< R >$ are the mean hydrophobicity and the mean net charge of the protein, respectively. Both parameters and the borderline between ordered and disordered were calculated according to the Uversky methodology (*50*). The number N of protein sequences ranges from ~500 to >2000 depending on the protein family, as presented in the material and methods section.

We next explored the relative amount of IDRs inside the protein sequences to get more quantitative details of the disordered behaviour. Drawing from the rich tapestry of protein families, we singled out two distinct families for comparison with our database. On one end of the spectrum, we considered the mammalian Albumins— representative of structurally well-ordered proteins—and on the opposite end, we included the RNA binding protein FUS, a classic representative of the extended IDPs found in vertebrates. The median of disorder fractions varied substantially from around 30% of the total protein sequence for Golgin45 and NEEC to nearly 70% for SCYL1BP1 (Figure 3A). Our data showed that most families were composed of a median ranging from 40-60%, explaining why these proteins behave more similarly to extended IDPs in the Uversky plot analyses.

An inherent challenge in predicting disordered regions in proteins is the often-observed overlap in predicting intrinsic disorder within coiled-coil regions, even though the converse is unlikely (*53, 54*). Regions within coiled coils are typically highly disordered in a monomeric state and only gain helical structure upon coiled-coil formation (*53*). To deepen our understanding of the structural propensities of the GMPs, we analysed the overall composition of order/disorder-promoting residues. In the realm of protein structure, the concepts of "disorder-promoting residues" and "order-promoting residues" refer to the propensities of specific amino acids to encourage either a structured (ordered) or unstructured (disordered) conformation within the protein chain. Disorder-promoting residues are those amino acids that, upon their inclusion within a protein, tend to enhance the propensity for the surrounding regions to adopt an unstructured or disordered conformation. These residues often exhibit distinct characteristics, such as small size (for instance, Glycine), net charge (like Lysine or Glutamic acid), or uncharged polarity (such as Serine or Threonine).

On the contrary, order-promoting residues are predominantly hydrophobic (examples include Leucine, Isoleucine, and Valine) and may participate in the constitution of the protein core. Alternatively, they can participate in forming hydrogen bonds (as Asparagine and Threonine), contributing to establishing regular secondary structure elements such as α-helices and β-sheets. Compared with the order-promoting residues observed for FUS (~15%) and Albumin (~33%), the composition observed for the GMPs tested was between these extreme situations (Figure 3B), again suggesting an intrinsically high degree of structural flexibility. GMPs harboured fewer regions of order and exhibited high flexibility (Figure 3B), suggesting that their function is likely linked to these structural characteristics. For example, GMPs tend to adopt extended structures, which give rise to the extensive ribosome-exclusion area surrounding the Golgi in microscopy images (*49, 55*).

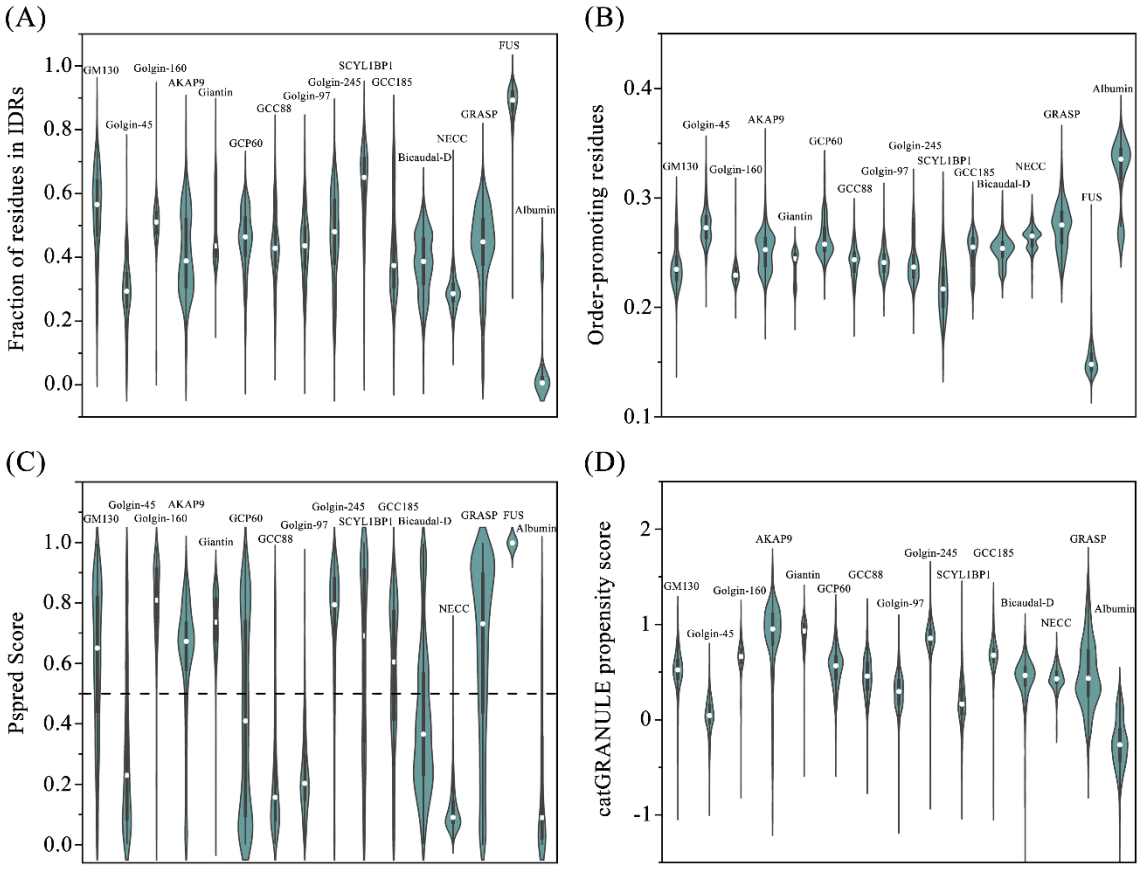

Figure 3: Analysis of Amino Acid Composition and LLPS-Propensity Across Metazoan GMPs. A) Representation of residue fractions exhibiting a score exceeding the 0.5 thresholds, as per IUPred2 long disorder predictions. B) Enumeration of order-promoting residue fractions ('I', 'C', 'L', 'V', 'W', 'Y', 'F'), as defined by reference (*56*). C) Application of the machine-learning-based PSPredictor (*57*). for LLPS prediction. A protein sequence is categorised as phase separating if the PsPred Score exceeds 0.5. D) Assessment of LLPS propensity utilising the catGRANULE algorithm, where propensity scores are evaluated in the context of proteome distribution, as conveyed through a cumulative distribution function. A higher score corresponds to a higher propensity for a protein to phase separate. As dosage-sensitive proteins are known to score approximately 0.5 (as per reference (*58*)), we designate this score as our threshold in the analyses.

The flexibility and diverse binding capability of IDPs/IDRs have been linked to liquid-liquid phase separation (LLPS) (*46, 57*). LLPS is like oil droplets forming in water: specific proteins or protein complexes separate from the surrounding solution to form concentrated, liquid-like droplets within the cell, and this phenomenon is now recognised to be widely spread *in vivo* (*59*). Predicting the propensity of a protein or a protein region to undergo LLPS is a vibrant area of investigation in current bioinformatics and biophysical research. Several computational tools are available to predict LLPS propensity, with many focusing on the presence of IDRs and sequence-based features like amino acid composition, charge, and hydrophobicity (*57, 58*). These include tools like PScore and catGRANULE (*57, 58*).

We utilised these two analytical tools to assess our GMP database, focusing on evaluating the propensity of LLPS and its conservation within specific protein families. Previously seen metazoan-specific Golgins, such as GM130, Golgin-97, Golgin-245, GCC88, and GCC185, were observed to undergo phase separation when overexpressed *in vivo* (*31, 60*). Interestingly, the Golgin-97 and GCC88 families did not demonstrate a strong propensity towards LLPS (Figures 3C and D). Similarly, Bicaudal D, NEEC, and Golgin45 lacked this strong propensity (Figure 3C and D). In order to delve deeper into these protein families with lower propensities, we conducted tests on the human protein representatives using the PhasePred Algorithm. This tool is a meta-predictor that aggregates the prediction scores of several LLPS-associated predictors (*61*). Notably, the human Golgin-97 protein (Uniprot Q92805) returned a Feature PS-Part score (a metric related to proteins whose phase separation behaviours are regulated by protein or nucleic acid partner components) exceeding 75%. This score points to a strong, albeit partner-dependent, propensity for LLPS. Similar patterns were detected for GCC88 (Uniprot Q96CN9), Bicaudal-D (Uniprot Q96G01), and Golgin-45 (Uniprot Q9H2G9), which produced PS-Part scores of approximately 64%, 77%, and 52% respectively. In contrast, NEEC (Uniprot Q96AA8) generated a markedly lower score, at around 32%.

Interestingly, the PhasePred provides an immunofluorescence image for each tested human protein sourced from the Protein Atlas (*62*) (http://v18.proteinatlas.org/ accessed in 2023). While these images do not conclusively demonstrate that the proteins were in a condensed state under the specific conditions of the experiment, each image revealed a punctate distribution within cells, implying that these proteins may tend to form discrete clusters. Therefore, all GM protein families tested (except for NEEC) showed a pronounced tendency for LLPS, either spontaneously or partner/context-dependent. This prevalence was particularly notable, considering the unique evolutionary paths of these proteins across Eukaria and their low primary sequence conservation. The significant conservation of this physicochemical property stood out, underscoring its potentially crucial role in those proteins' functions and interactions.

Since LLPS appears inherent to metazoan GMPs, it is tempting to speculate that this property is important to the proteins' functionality. As previously discussed, one of the principal roles of GMPs lies in organising and preserving the Golgi apparatus. However, characteristics such as flattened cisternae and stacked Golgi are not unique to metazoans. Furthermore, considering the evolutionary pathways and conservation of certain GMPs, there is a compelling argument that Golgi stacking may have originated in the common eukaryotic ancestor. In a study grounded on phylogenetics, Barlow and colleagues identified several proteins implicated in the Golgi structure, which were inferred to be present in the Last Eukaryotic Common Ancestor (LECA). These proteins, which share roles in delineating separate Golgi structures, include GMAP, GRASP, TMF, Golgin-84, CASP, p115, GRIP-containing protein, and zinc finger protein-

like 1 (ZFPL1) (*52*) (figure 4A). Therefore, we expanded our earlier analyses to encompass a database of these "ancient" GMPs. This extension aimed to investigate whether the high LLPS tendency observed in metazoans is also prevalent across Eukarya. Our analyses confirm a high prevalence of protein regions with a high propensity to be intrinsically disordered, with medians ranging from ~20% (p115) to ~60% (TMF) of the entire protein sequence (Figure 4B). Interestingly, even other properties like pI are highly conserved across the eukarya GMPs predicted to be present in LECA (Figure 4C). In alignment with our previous findings, most GMPs also exhibited a high propensity for spontaneous LLPS, including the families of GRASP, GMAP-21, TMF, Golgin-84, and GRIP-CP (Figures 4D and E). The exceptions were ZFPL1, p115, and CASP, with scores as low as those observed in our control dataset of Albumins.

As before, we employed the PhasePred Algorithm to test the human proteins of these low-score families. Notably, ZFPL1 exhibited an exceptionally low score of approximately 6%, substantiating our observations in the database (Figures 4D and E). ZFPL1 is necessary for maintaining the integrity of the cis-Golgi, possibly through its interaction with GM130 (*63*). The latter has been previously seen to form liquid-like condensates *in vivo* (*60*), suggesting ZFPL1 could play a part in this complex condensed state formed by GM130 in metazoans or interact with other Golgins throughout evolution. The human protein p115 registered a score of around 45%, indicating that proteins in the Eukarya family could coexist in a condensed state but in a partner-dependent context. Human CASP, with a score of approximately 20%, showed no signs that this protein family could spontaneously phase separate or that a protein/RNA partner could mediate this process. Our study demonstrates that the high prevalence of disordered regions and the inclination for Liquid-Liquid Phase Separation, whether occurring spontaneously or being partner-dependent, are intrinsic characteristics of a vast majority of GMPs across all Eukaryotic species.

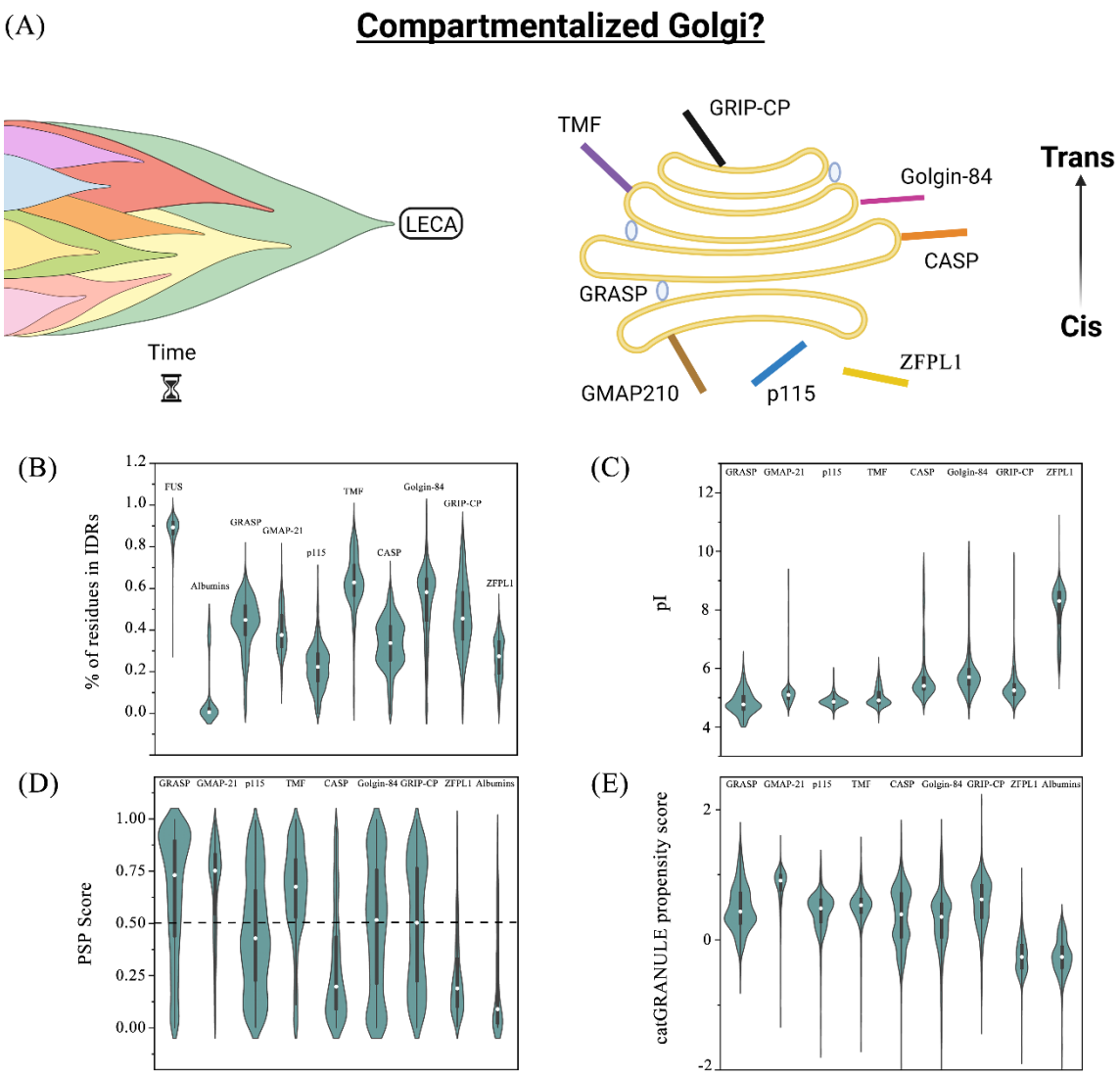

Figure 4: Analyses of the LLPS propensity and IDR content in Golgi structure proteins inferred to be present in the LECA according to (*52*). A) Schematic view of ancient protein families predicted to be present in LECA and their relative position based on modern Golgi compartments. B) Representation of residue fractions exhibiting scores exceeding the 0.5 thresholds, using IUPred2 long disorder predictions. C) pI distribution across the GMP database. The pI values were calculated using an in-house Python script described in the Material and Methods Section. D) Application of the machine-learning-based PSPredictor (*57*). for LLPS prediction. A protein sequence is categorised as phase separating if the PsPred Score exceeds 0.5. E) Assessment of LLPS propensity utilising the catGRANULE algorithm, where propensity scores are evaluated in the context of proteome distribution, as conveyed through a cumulative distribution function. A higher score corresponds to a higher propensity for a protein to phase separate. As dosage-sensitive proteins are known to score approximately 0.5 (as per reference (*58*)), we designate this score as our threshold in the analyses.

Conclusion:

In conclusion, our study provides a novel perspective on the nature of GMPs and critical functional characteristics across Eukarya, focusing on metazoans. Our data distinctly highlight the high prevalence of intrinsic disorder within GMPs, a feature that appears to be conserved

across diverse eukaryotic lineages, albeit with different levels of prevalence and extent. Interestingly, the disorder in GMPs seems to be central to the potential of these proteins for LLPS, a phenomenon widely observed in the intracellular environment and critical for various biological processes. The ability to phase separate appears to be conserved amongst many GMP families, with some proteins showing a strong partner-dependent propensity for this characteristic. The combined findings indicate that disorder and LLPS propensity, despite the low primary sequence conservation across GMP families, have been evolutionarily conserved, suggesting their critical roles in the proteins' functionality. These observations highlight that the functional diversity and flexibility of GMPs may stem from their structural disorder, enabling them to adapt to a wide array of interactions and respond to various cellular cues, underpinning their role in Golgi apparatus organisation and preservation. Our findings open the way for future studies on the role of disorder and LLPS in GMPs and their implication in various cellular processes. It will be particularly interesting to delve further into the precise mechanisms by which these proteins influence the architecture of the Golgi apparatus and the regulatory role of phase separation in these processes. The evolutionary conservation of these traits among GMPs suggests their fundamental importance in cell biology and invites further exploration.

Material and Methods

Database collection:

Our protein database was created using the hierarchical catalogue of orthologs OrthoDB (*64*) and/or search for protein families in the InterPro database (*65*). The protein sequences collected were as follows: TMF (OrthoDB 2054925at2759, 3717503at2759 and 1221595at2759, N = 1629), CASP (InterPro IPR012955, N = 2395), Golgin84 (InterPro IPR019177, N = 1500), Mammalian Serum Albumin-Like (from InterPro IPR020858, N = 1500), GRASP (OrthoDB 36330at2759, N = 2395), GM130 (OrthoDB 2880220at2759, N = 887), Golgin45 (OrthoDB 2879260at2759, N = 830), Golgin160 (OrthoDB 5357451at2759, N = 560), AKAP9 (OrthoDB 158920at33208, N = 1550), Giantin (OrthoDB 4227816at2759, N = 505), GCP60 (OrthoDB 38315at33208, N = 1384), GCC88 (OrthoDB 237371at33208, N = 831), Golgin97 (OrthoDB 42510at33208, N = 832), FUS RNA binding protein (OrthoDB 484351at7742, N = 982), Golgin245 (OrthoDB 2529099at33208, N = 871), protein bicaudal D homolog 2-like (OrthoDB 5350043at2759, N = 1886)), SCYL1BP1 (OrthoDB 2486431at33208, N = 708), p115 (OrthoDB 1774at2759, N = 1927), GCC185 (OrthoDB 814692at33208, N = 807), GRIP-Containing Proteins (N = 1500 randomly selected from InterPro IPR000237), GMAP (OrthoDB Group 5138860at2759, N = 886), ZFPL1 (OrthoDB Group 25738at2759, N = 903) and Janus kinase and microtubule interacting protein 2 - NECC (OrthoDB 5403492at2759, N = 1461). We employed

the online resource found at https://fasta.bioch.virginia.edu/fasta_www2/clean_fasta.html (accessed in 2023) to process and clean our dataset's FASTA formatted protein names. For most larger databases exceeding 1500 sequences, a custom Python script was utilised to randomly truncate the sequence data to the desired size of 1500 (script "fasta_random" in the SI). We have also generated a Python script to filter out unnatural amino acid residues within the protein sequences (script "fasta_filter" in the SI). This script reads the protein sequences from an input file, removes any character that is not a valid amino acid from each sequence, and writes the cleaned sequences to an output file, preserving the headers of the original sequences. This procedure was crucial given that most of the analysis tools we employed lacked the functionality to handle these unconventional letters appropriately. The full protein sequence database can be made available under request.

Prediction of LLPS-Propensity:

The propensity for LLPS was evaluated using the PSPred and catGRANULE web-based platforms. Each algorithm accepts different sequence lengths: PSPred can simultaneously process segments of 100 amino acids, while catGRANULE can handle up to 500 amino acids per run. To streamline the upload process and accommodate these constraints, we developed Python scripts that partition our protein sequence database into appropriately sized segments for submission to each server. This script (script "fasta_split" in the SI) reads a FASTA file, splits the sequences into groups of a specified size, and writes each group to a separate FASTA file. The statistical analyses were performed using the software Origin 2022.

Prediction of IDRs and calculated fractions within the sequence

We quantified each amino acid's intrinsic disorder within a sequence using the IUPred2A tool, specifically designed for assessing long sequences (*66*). We engineered a Python script to interface with our comprehensive protein sequence databases to streamline the process (script "disorder_fractions" in the SI). This script seamlessly integrated with the IUPred2A software package we sourced directly from the official website. It then automated the computation of disorder scores, effectively isolating the fraction of residues with scores exceeding the threshold of 0.5. The statistical analyses were performed using the software Origin 2022.

Charge vs Hydrophobicity Plot:

We crafted a Python script implementing the Uversky methodology (script "fasta_CHA" in the SI) to systematically evaluate each protein sequence's mean hydrophobicity and net charge in our database. Our user-friendly script allows customisation of the hydrophobicity scale per user preference, exploring the potential described in (*67*). For this manuscript, we have utilised the Kyte-Doolittle scale, originally employed by Uversky in (*50*). Our data presentation shows a discrimination borderline, defined by a specific equation, categorising proteins as either "natively disordered" or "natively ordered". The equation is $< R > = 2.785 < H > - 1.151$, according to (*50*). In this work, $< H >$ and $< R >$ are the mean hydrophobicity and the mean net charge of the protein, respectively.

Amino acid sequence general properties determination:

We engineered a Python script to efficiently perform multiple tasks to expedite the protein sequence analysis process. This script reads protein sequences from a FASTA file, conducts a comprehensive evaluation of various properties using the Biopython package, and outputs these properties in an organised tabular format. The analysed properties include fractions of residues promoting order/disorder, isoelectric point (pI), Grand Average of Hydropathy (GRAVY), and the proportion of polar and charged residues. The statistical analyses were performed using the software Origin 2022.

Acknowledgements

The authors thank the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (Grants No. 2022/06006-0, 2021/10465-7 and 2020/15542-7) for the financial support. AJCF thanks Conselho Nacional de Desenvolvimento Científico e Tecnlógico (CNPq) for the partial financial support (Grant no. 306682/2018-4)

Conflict of interest

The authors declare that they have no conflict of interest.

References:

---

1. M. Uhlen *et al.*, Tissue-based map of the human proteome. *Science* **347**, (2015).
2. R. Benyair, E. Ron, G. Z. Lederkremer, Protein quality control, retention and degradation at the endoplasmic reticulum. *International Review of Cell and Molecular Biology, Vol 292* **292**, 197-280 (2011).
3. J. Dancourt, C. Barlowe, Protein Sorting Receptors in the Early Secretory Pathway. *Annual Review of Biochemistry, Vol 79* **79**, 777-802 (2010).
4. G. Zanetti, K. B. Pahuja, S. Studer, S. Shim, R. Schekman, COPII and the regulation of protein sorting in mammals. *Nature Cell Biology* **14**, 20-28 (2012).
5. J. Seemann, E. Jokitalo, M. Pypaert, G. Warren, Matrix proteins can generate the higher order architecture of the Golgi apparatus. *Nature* **407**, 1022-1026 (2000).
6. I. Lee *et al.*, Membrane adhesion dictates Golgi stacking and cisternal morphology. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1849-1854 (2014).
7. T. M. Witkos, M. Lowe, The Golgin Family of Coiled-Coil Tethering Proteins. *Frontiers in Cell and Developmental Biology* **3**, (2016).
8. A. K. Gillingham, S. Munro, Finding the Golgi: Golgin Coiled-Coil Proteins Show the Way. *Trends in Cell Biology* **26**, 399-408 (2016).
9. Y. J. Zhang, J. Seemann, Rapid degradation of GRASP55 and GRASP65 reveals their immediate impact on the Golgi structure. *Journal of Cell Biology* **220**, (2020).
10. M. Wong, S. Munro, The specificity of vesicle traffic to the Golgi is encoded in the golgin coiled-coil proteins. *Science* **346**, 601-+ (2014).
11. Y. Xiang *et al.*, Regulation of protein glycosylation and sorting by the Golgi matrix proteins GRASP55/65. *Nature Communications* **4**, (2013).
12. L. F. S. Mendes *et al.*, The GRASP domain in golgi reassembly and stacking proteins: differences and similarities between lower and higher Eukaryotes. *Febs Journal* **286**, 3340-3358 (2019).
13. L. F. S. Mendes *et al.*, Resurrecting Golgi proteins to grasp Golgi ribbon formation and self-association under stress. *International Journal of Biological Macromolecules*, **194**, 264-275 (2022).
14. L. F. S. Mendes *et al.*, Conformational flexibility of GRASPs and their constituent PDZ subdomains reveals structural basis of their promiscuous interactome. *Febs Journal* **287**, 3255-3272 (2020).
15. L. F. S. Mendes, N. A. Fontana, S. T. Reddy, V. N. Uversky, A. J. Costa, The exquisite structural biophysics of the Golgi Reassembly and Stacking Proteins. *International Journal of Biological Macromolecules* **164**, 3632-3644 (2020).
16. L. F. S. Mendes *et al.*, New structural insights into Golgi Reassembly and Stacking Protein (GRASP) in solution. *Scientific Reports* **6**, (2016).
17. I. Ayala, F. Mascanzoni, A. Colanzi, GRASP65-mediated control of the Golgi structure is crucial for G2/M transition. *Molecular Biology of the Cell* **29**, (2018).
18. D. Tang, H. Yuan, Y. Wang, The Role of GRASP65 in Golgi Cisternal Stacking and Cell Cycle Progression. *Traffic* **11**, 827-842 (2010).
19. J. D. Lane *et al.*, Caspase-mediated cleavage of the stacking protein GRASP65 is required for Golgi fragmentation during apoptosis. *Journal of Cell Biology* **156**, 495-509 (2002).
20. N. A. Fontana, A. D. Rosse, A. Watts, P. S. R. Coelho, A. J. Costa-Filho, *In vivo* amyloid-like fibrils produced under stress. *bioRxiv*, 2021.2002.2002.429251 (2021).

21. N. A. Fontana, R. Fonseca-Maldonado, L. F. S. Mendes, L. P. Meleiro, A. J. Costa, The yeast GRASP Grh1 displays a high polypeptide backbone mobility along with an amyloidogenic behavior. *Scientific Reports* **8**, (2018).
22. R. Behnia, F. A. Barr, J. J. Flanagan, C. Barlowe, S. Munro, The yeast orthologue of GRASP65 forms a complex with a coiled-coil protein that contributes to ER to Golgi traffic. *Journal of Cell Biology* **176**, 255-261 (2007).
23. R. Grond *et al.*, The function of GORASPs in Golgi apparatus organisation in vivo. *J Cell Biol* **219**, (2020).
24. J. Shorter, G. Warren, Golgi architecture and inheritance. *Annual Review of Cell and Developmental Biology* **18**, 379-420 (2002).
25. P. A. Takizawa *et al.*, COMPLETE VESICULATION OF GOLGI MEMBRANES AND INHIBITION OF PROTEIN-TRANSPORT BY A NOVEL SEA SPONGE METABOLITE, ILIMAQUINONE. *Cell* **73**, 1079-1090 (1993).
26. J. Lippincottschwartz, L. C. Yuan, J. S. Bonifacino, R. D. Klausner, RAPID REDISTRIBUTION OF GOLGI PROTEINS INTO THE ER IN CELLS TREATED WITH BREFELDIN-A - EVIDENCE FOR MEMBRANE CYCLING FROM GOLGI TO ER. *Cell* **56**, 801-813 (1989).
27. B. S. Glick, A. Nakano, Membrane Traffic Within the Golgi Apparatus. *Annual Review of Cell and Developmental Biology* **25**, 113-132 (2009).
28. N. Gomez-Navarro, E. Miller, Protein sorting at the ER-Golgi interface. *Journal of Cell Biology* **215**, 769-778 (2016).
29. K. Park, S. Ju, N. Kim, S. Y. Park, The Golgi complex: a hub of the secretory pathway. *Bmb Reports* **54**, 246-252 (2021).
30. J. E. Rothman, Jim's View: Is the Golgi stack a phase-separated liquid crystal? *Febs Letters* **593**, 2701-2705 (2019).
31. P. Ziltener, A. A. Rebane, M. Graham, A. M. Ernst, J. E. Rothman, The golgin family exhibits a propensity to form condensates in living cells. *Febs Letters*.
32. M. Fuxreiter, M. Vendruscolo, Generic nature of the condensed states of proteins. *Nature Cell Biology* **23**, 587-594 (2021).
33. S. Boeynaems *et al.*, Protein Phase Separation: A New Phase in Cell Biology. *Trends in Cell Biology* **28**, 420-435 (2018).
34. S. Alberti, D. Dormann, Liquid-Liquid Phase Separation in Disease. *Annual Review of Genetics, Vol 53* **53**, 171-+ (2019).
35. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).
36. H. T. McMahon, J. L. Gallop, Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* **438**, 590-596 (2005).
37. S. Mondal, T. Baumgart, Membrane reshaping by protein condensates. *Biochimica Et Biophysica Acta-Biomembranes* **1865**, (2023).
38. H. T. McMahon, E. Boucrot, Membrane curvature at a glance. *Journal of Cell Science* **128**, 1065-1070 (2015).
39. G. Fusco *et al.*, Structural basis of membrane disruption and cellular toxicity by alpha-synuclein oligomers. *Science* **358**, 1440-+ (2017).
40. L. G. M. Basso, E. F. Vicente, E. Crusca, E. M. Cilli, A. J. Costa, SARS-CoV fusion peptides induce membrane surface ordering and curvature. *Scientific Reports* **6**, (2016).
41. E. M. Houang, Y. Y. Sham, F. S. Bates, J. M. Metzger, Muscle membrane integrity in Duchenne muscular dystrophy: recent advances in copolymer-based muscle membrane stabilisers. *Skeletal Muscle* **8**, (2018).
42. P. Bassereau *et al.*, The 2018 biomembrane curvature and remodeling roadmap. *Journal of Physics D-Applied Physics* **51**, (2018).
43. F. Yuan *et al.*, Membrane bending by protein phase separation. *Proceedings of the National Academy of Sciences of the United States of America* **118**, (2021).

44. V. N. Uversky, Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current Opinion in Structural Biology* **44**, 18-30 (2017).
45. C. Has, P. Sivadas, S. L. Das, Insights into Membrane Curvature Sensing and Membrane Remodeling by Intrinsically Disordered Proteins and Protein Regions. *Journal of Membrane Biology* **255**, 237-259 (2022).
46. K. K. Turoverov *et al.*, Stochasticity of Biological Soft Matter: Emerging Concepts in Intrinsically Disordered Proteins and Biological Phase Separation. *Trends in Biochemical Sciences* **44**, 716-728 (2019).
47. J. Agudo-Canalejo *et al.*, Wetting regulates autophagy of phase-separated compartments and the cytosol. *Nature* **591**, 142-+ (2021).
48. A. Mangiarotti, N. Chen, Z. Zhao, R. Lipowsky, R. Dimova, Wetting and complex remodeling of membranes by biomolecular condensates. *Nature Communications* **14**, 2809 (2023).
49. F. S. Sjostrand, V. Hanzon, ULTRASTRUCTURE OF GOLGI APPARATUS OF EXOCRINE CELLS OF MOUSE PANCREAS. *Experimental Cell Research* **7**, 415-429 (1954).
50. V. N. Uversky, J. R. Gillespie, A. L. Fink, Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins-Structure Function and Genetics* **41**, 415-427 (2000).
51. J. Zhao, B. Li, X. Huang, X. Morelli, N. Shi, Structural Basis for the Interaction between Golgi Reassembly-stacking Protein GRASP55 and Golgin45. *Journal of Biological Chemistry* **292**, 2956-2965 (2017).
52. L. D. Barlow, E. Nyvltova, M. Aguilar, J. Tachezy, J. B. Dacks, A sophisticated, differentiated Golgi in the ancestor of eukaryotes. *Bmc Biology* **16**, (2018).
53. R. van der Lee *et al.*, Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews* **114**, 6589-6631 (2014).
54. D. Simm, K. Hatje, S. Waack, M. Kollmar, Critical assessment of coiled-coil predictions based on protein structure data. *Scientific Reports* **11**, (2021).
55. S. Mogelsvang, N. Gomez-Ospina, J. Soderholm, B. S. Glick, L. A. Staehelin, Tomographic evidence for continuous turnover of Golgi cisternae in Pichia pastoris. *Molecular Biology of the Cell* **14**, 2277-2291 (2003).
56. L. F. S. Mendes *et al.*, New structural insights into Golgi Reassembly and Stacking Protein (GRASP) in solution. *Scientific Reports* **6**, (2016).
57. X. Q. Chu *et al.*, Prediction of liquid-liquid phase separating proteins using machine learning. *Bmc Bioinformatics* **23**, (2022).
58. B. Bolognesi *et al.*, A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell Reports* **16**, 222-231 (2016).
59. F. C. Keber, T. Nguyen, C. P. Brangwynne, M. Wühr, Evidence for widespread cytoplasmic structuring into mesoscopic condensates. *bioRxiv*, 2021.2012.2017.473234 (2021).
60. A. A. Rebane *et al.*, Liquid-liquid phase separation of the Golgi matrix protein GM130. *Febs Letters* **594**, 1132-1144 (2020).
61. Z. M. Chen *et al.*, Screening membraneless organelle participants with machine-learning models that integrate multimodal features. *Proceedings of the National Academy of Sciences of the United States of America* **119**, (2022).
62. C. Y. Yu *et al.*, Proteome-scale analysis of phase-separated proteins in immunofluorescence images. *Briefings in Bioinformatics* **22**, (2021).
63. C. F. Chiu *et al.*, ZFPL1, a novel ring finger protein required for cis-Golgi integrity and efficient ER-to-Golgi transport. *Embo Journal* **27**, 934-947 (2008).
64. D. Kuznetsov *et al.*, OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research*, (2022).
65. T. Paysan-Lafosse *et al.*, InterPro in 2022. *Nucleic Acids Research*, (2022).

66. B. Meszaros, G. Erdos, Z. Dosztanyi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research* **46**, W329-W337 (2018).
67. F. Huang *et al.*, Improving protein order-disorder classification using charge-hydropathy plots. *Bmc Bioinformatics* **15**, (2014).