scientific reports

OPEN

A statistical mechanics investigation of unfolded protein response across organisms

Nicole Luchetti[1,2✉], Keith M. Smith[3], Margherita A. G. Matarrese[1], Alessandro Loppini[4], Simonetta Filippi[1,5,6✉] & Letizia Chiodo[1]

Living systems rely on coordinated molecular interactions, especially those related to gene expression and protein activity. The Unfolded Protein Response is a crucial mechanism in eukaryotic cells, activated when unfolded proteins exceed a critical threshold. It maintains cell homeostasis by enhancing protein folding, initiating quality control, and activating degradation pathways when damage is irreversible. This response functions as a dynamic signaling network, with proteins as nodes and their interactions as edges. We analyze these protein-protein networks across different organisms to understand their intricate intra-cellular interactions and behaviors. In this work, analyzing twelve organisms, we assess how fundamental measures in network theory can individuate seed proteins and specific pathways across organisms. We employ network robustness to evaluate and compare the strength of the investigated protein-protein interaction networks, and the structural controllability of complex networks to find and compare the sets of driver nodes necessary to control the overall networks. We find that network measures are related to phylogenetics, and advanced network methods can identify main pathways of significance in the complete Unfolded Protein Response mechanism.

Keywords Endoplasmic reticulum stress, Complex networks, Protein-protein interactions, Network analysis

The Unfolded Protein Response (UPR)[1] is a mechanism adopted by cells to maintain homeostasis within the endoplasmic reticulum (ER) compartment in response to an accumulation of unfolded or improperly folded proteins (Fig. 1)[2–4]. When protein concentration exceeds physiological levels, pro-survival mechanisms are activated to restore the balance between folded and unfolded proteins[5–7]. The heat shock protein family A member 5 (HSPA5), also known as binding immunoglobulin protein (BiP)[8], is a key promotor of the UPR, activating three stress sensors in the ER: the activating transcription factor 6 (ATF6), the endoplasmic reticulum to nucleus signaling 1 (ERN1), and the eukaryotic translation initiation factor 2 alpha kinase 3 (EIF2AK3), respectively[9–11]. If the adaptive UPR response fails, other pathways are activated, leading to apoptosis and autophagy[6,12]. This mechanism is essential for cell survival in mammals[1,13,14] and is strongly preserved across various organisms, from vertebrates to yeasts and worms[15–17], as well as in fungi[18] and plants[19,20].

The advancement of network theory has significantly contributed to studying biological networks, particularly protein-protein interaction (PPI) networks[21–24]. Indeed, by conducting network modeling and topological analysis, researchers can gain insights into genes and proteins involved in various biological functions and disease mechanisms[25–27] The UPR pathway can be described as a PPI network[28–30] and analyzed using complex network tools[31–36]. Protein interaction information is stored in public databases[37–40], and obtained via direct and indirect information (i.e., from experimental Y2H test and homology). Classic measures in network theory, whose definition is briefly reported in Table 1, provide valuable insights into network structure and function. Still, they do not adequately address the dynamic aspects of network behavior and vulnerability to disruptions. Therefore, structural controllability[41,42] and network robustness[43–45] can be used to identify driver nodes and exploit whether the network withstands failures or attacks.

[1]Department of Engineering, Università Campus Bio-Medico di Roma, Via Álvaro del Portillo 21, Rome 00128, Italy. [2]Center for Life Nano- & Neuro-Science, Italian Institute of Technology, Viale Regina Elena 291, Rome 00161, Italy. [3]Computer and Information Sciences, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, United Kingdom. [4]Department of Medicine and Surgery, Università Campus Bio-Medico di Roma, Via Álvaro del Portillo 21, Rome 00128, Italy. [5]National Institute of Optics, National Research Council, Largo Enrico Fermi 6, Florence 50125, Italy. [6]International Center for Relativistic Astrophysics Network, Piazza della Repubblica 10, Pescara 65122, Italy. ✉email: n.luchetti@unicampus.it; s.filippi@unicampus.it
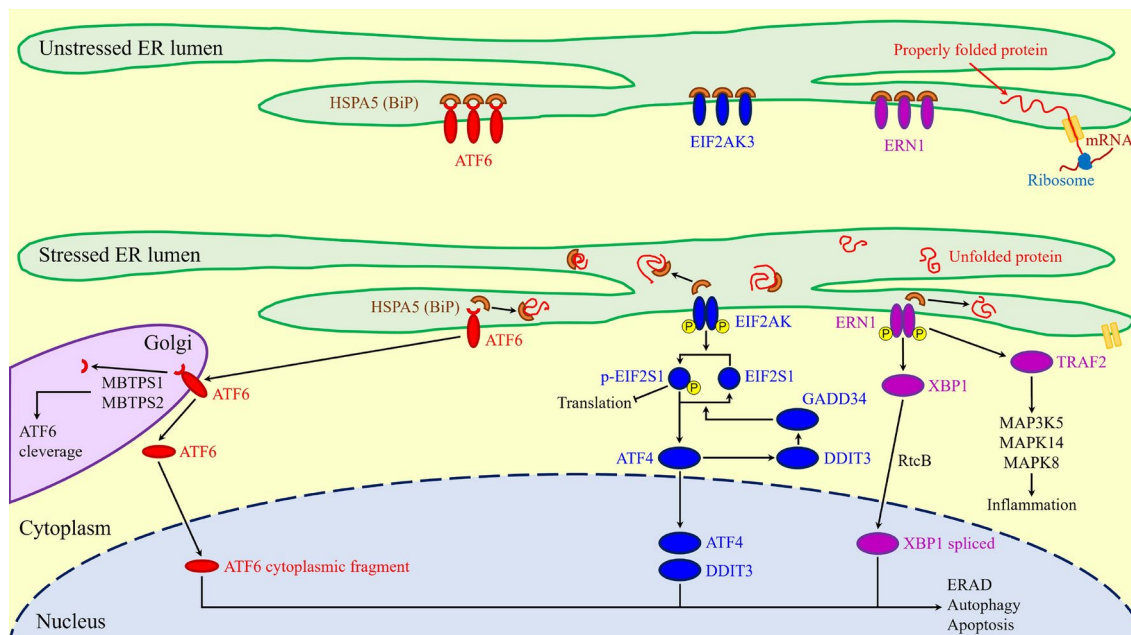
Fig. 1. Visual representation of main UPR signaling pathways during ER stress in vertebrates. The cellular process of post-translational modification and protein folding becomes strained, leading to the buildup of non-properly folded proteins. This accumulation can eventually trigger cell death. The cell initiates a cascade of signaling pathways to counteract ER stress and regain homeostasis. These pathways enhance the production of proteins involved in proper protein folding or facilitate the removal of misfolded proteins through Endoplasmic-reticulum-associated protein degradation (ERAD). (Reproduction under the Creative Commons Attribution License).

Term/metric	Definition
Barycenter	The node with the lowest value of eccentricity (for us, the absolute center of the networks)
Betweenness centrality[46]	Measure of how often a node occurs on the shortest paths between other nodes
Closeness centrality	$C(x) = (\sum_y d(x,y))^{-1}$ is a measure of how close a node is to all other nodes in the network
Clustering coefficient	Proportion of edges between the nodes within the ith neighborhood divided by the number of links that could exist between them
Average clustering coefficient	$CC = (\sum_i CC_i/n)$ is the arithmetic mean of the clustering coefficient of all the nodes
Density[47]	$D = 2M/[N \cdot (N-1)]$, where M is the total number of connections in an N nodes network
Degree	Number of edges of one node
Average Degree	Arithmetic mean of degrees of all network nodes
Diameter	It is defined as the eccentricity of a node with the maximum distance to the other nodes
Edges	Physical or functional connections between pairs of proteins
Modules or Communities	Sub networks that include a high number of inside-sub network edges and a low number of between-sub network edges
Modularity	A measure of network tendencies to divide in communities
Nodes	Proteins composing the network
Shortest path length	Number of edges needed to connect every pair of nodes through their shortest path

Table 1. Definition of quantities used to describe PPI networks.

Here we study the properties of the ER stress response in twelve different organisms, to determine if network analysis methods can provide insight into characteristics of PPI networks. Specifically, we want to identify and analyze the factors that impact the "strength" of various UPR networks and their resistance to potential alterations. This includes looking at random-based and various metric-based attack strategies and identifying similarities between different organism models. Our findings indicate that the several adopted methods can uncover different network characteristics, such as phylogenetic similarities[1], distinguishing mammals and their animal models, and identifying relevant molecular pathways within the UPR mechanism across organisms. Thus, we hypothesize that these network methods can be widely applied to characterize unknown PPI networks in silico.

Results

The results are presented in three subsections, each elucidating the potential descriptive and predictive power of the methods employed. These methods allow the association of network properties with phylogenetic analogies and assist in identifying biological weaknesses through advanced network descriptors. These subsections correspond to the methods tested in this study: i) standard network descriptors[48], as well as topological analysis[49], ii) robustness[43–45], and iii) structural controllability[41,42]. The methodological pipeline starts by establishing a native UPR model network for each organism, as detailed in Table 2, utilizing PPI data sourced from public databases. We then create configuration models by randomizing connections while preserving the same number of connections per protein. Well-established network theory measures and advanced network methods are then applied to evaluate the networks for each organism and model. In Table 2 we also report the average value of network features for each native model.

Network topology analysis

Standard network characteristics

Barycenter

We find that the barycenter of all network models corresponds to the Binding Immunoglobulin Protein (KAR2 and BIP2/BIP3 proteins in *Saccharomyces cerevisiae* and *Arabidopsis thaliana* are homologous to vertebrates HSPA5), apart from *Caenorhabditis elegans*, for which the Heat shock protein 90 (hsp-90) results as the key protein (Table 2). Indeed, literature shows that in *Caenorhabditis elegans*, hsp-90 plays a crucial role in the chemotaxis to non-volatile and volatile attractants detected by AWC sensory neurons[76,77].

Density

Arabidopsis thaliana and *Drosophila melanogaster* are the most densely connected networks, with values of 0.314 and 0.238, respectively (Table 2). The less densely connected network is found for the invertebrates, except for *Homo sapiens*, with values < 0.100.

Average degree

The highest average degree value of the native models is observed for *Homo sapiens* (25) followed by *Mus musculus* (21), *Arabidopsis thaliana*, and *Saccharomyces cerevisiae* with a value of an average degree of 19. The lowest average degree value is found for *Gallus gallus* and *Caenorhabditis elegans*, with a numeric value of 11, followed by *Macaca fascicularis* with a value of 13.

Closeness centrality

Arabidopsis thaliana has the highest closeness, with a value of $9.2 \cdot 10^{-3}$, followed by *Drosophila melanogaster* ($8.9 \cdot 10^{-3}$). Vertebrates show similar values, in the range $[1.8 \cdot 10^{-3}, 2.3 \cdot 10^{-3}]$. Betweenness centrality. Vertebrates show larger values of betweenness centrality, with the lower end for *Homo sapiens* (145.5) and the upper one for *Danio rerio* (181.3), compared to the remaining organisms.

Clustering coefficient

The highest clustering coefficient value is observed for *Arabidopsis thaliana* (0.680) followed by *Saccharomyces cerevisiae* (0.607). All the remaining organisms show a clustering coefficient < 0.600 suggesting that, in the smaller networks, proteins are directly connected with their neighbors.

Modularity and communities

The modularity and number of communities provide a different description of alteration in a network since it is an evaluation based on its configuration models. In Table 2 we report the number of communities and values of modularity calculated with the Louvain algorithm. *Drosophila melanogaster* and *Arabidopsis thaliana* provide the same lowest number of communities (3), probably because the sizes of the networks are quite small and comparable among them, concerning the other organisms (57 and 62 nodes). This is also reflected in the smallest diameter of the graph (5 for both). *Rattus norvegicus*, *Bos taurus*, and *Danio rerio* provide the same number of communities (6) and modularity value (0.374/0.374/0.375). A graph representation of the main communities of native models is shown in Fig. 2.

Correlation analysis

Organism	Barycenter	#nodes	#edges	Density	Diameter	Degree	Closeness	Betweenness	Cust. coeff.	Modularity	Communities
Homo sapiens[2,50]	HSPA5	216	5286	0.114	7	25	0.0020	145.5	0.522	0.311	5
Rattus norvegicus[51,52]	Hspa5	210	3862	0.088	6	18	0.0020	161.2	0.496	0.374	6
Mus musculus[53,54]	Hspa5	231	4900	0.092	6	21	0.0018	170.7	0.538	0.354	6
Macaca fascicularis[55]	HSPA5	186	2468	0.072	6	13	0.0020	160.5	0.437	0.422	7
Bos taurus[56]	HSPA5	205	3292	0.079	8	16	0.0019	172.6	0.447	0.374	6
Oryctologus cuniculis[57,58]	HSPA5	174	2598	0.086	6	15	0.0023	134.7	0.464	0.367	5
Gallus gallus[59,60]	HSPA5	179	2048	0.064	8	11	0.0020	173.5	0.395	0.385	7
Danio rerio[61–63]	hspa5	214	3240	0.071	8	15	0.0018	181.3	0.413	0.375	6
Drosophila melanogaster[64–66]	Hsc70-3	57	760	0.238	5	14	0.0089	29.9	0.560	0.324	3
Caenorhabditis elegans[67–69]	hsp-90	109	1212	0.103	7	11	0.0034	98.9	0.594	0.450	6
Saccharomyces cerevisiae[16,70,71]	KAR2	150	2834	0.127	6	19	0.0028	108.0	0.607	0.352	6
Arabidopsis thaliana[72–75]	BIP2; BIP3	62	1188	0.314	5	19	0.0092	25.6	0.680	0.245	3

Table 2. Average values of common network features for all native models.
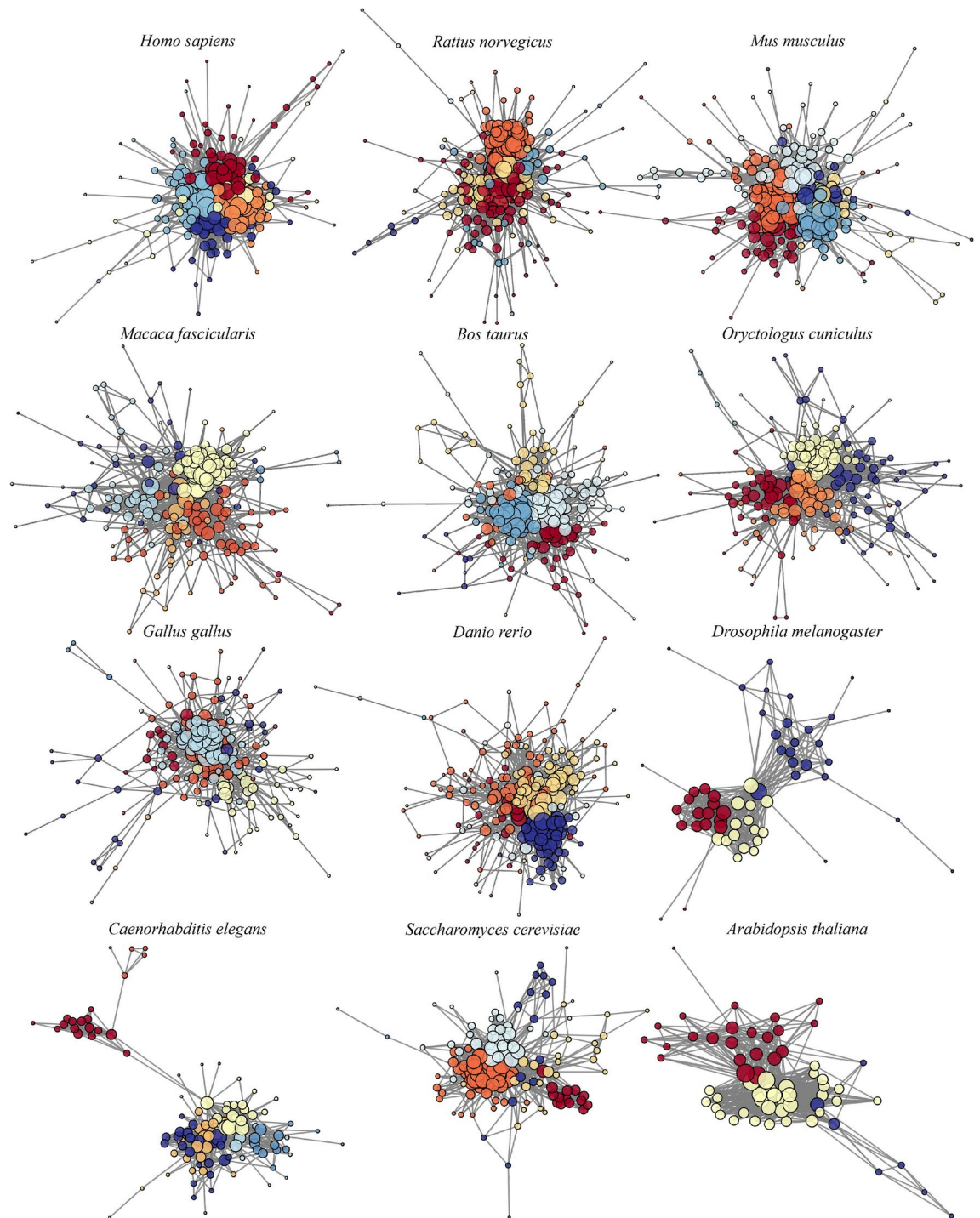
Fig. 2. Graph representation of native UPR models across species. Distinct colors identify communities, and the size of the nodes is related to their degree.

Pairwise comparisons for degree, closeness, and betweenness are evaluated via the Pearson correlation coefficient. This analysis is performed only across these quantities because they are the more common centrality measures. Degree and closeness centrality show a little negative non-significant correlation ($r = -0.04$, $p = 0.91$), while degree and betweenness centrality show a little positive non-significant correlation ($r = 0.006$, $p = 0.98$). Finally, the two centrality measures show a strong significant negative correlation ($r = -0.94$, $p = 4 \cdot 10^{-6}$).

Normalized metrics characteristics

The native UPR models for the different organisms are characterized by different sizes that may affect the standard measures reported in the previous paragraph. Thus, when analyzing normalized metrics (based on the network size), analogies and differences between organisms become clearer (Table 3).

Degree

Except for *Homo sapiens*, the other vertebrates share comparable values of normalized average degree (all with < 0.100). *Drosophila melanogaster* and *Arabidopsis thaliana* have the highest values (> 0.200).

Closeness centrality

Homo sapiens, *Rattus norvegicus*, *Mus musculus*, and *Saccharomyces cerevisiae* share similar values of normalized average closeness centrality (between 0.404 and 0.440). *Macaca fascicularis*, *Bos taurus*, *Danio rerio*, and *Caenorhabditis elegans* models also provide similar values (between 0.368 and 0.386). *Arabidopsis thaliana* and *Drosophila melanogaster* have the highest closeness values (0.563 and 0.501).

Betweenness centrality

Except for *Arabidopsis thaliana* and *Saccharomyces cerevisiae*, which also provide the highest values (0.135 and 0.092), all the other organisms have comparable values of normalized average betweenness centrality, especially the two murine organisms (0.042 and 0.040), and *Macaca fascicularis*, *Gallus gallus*, *Danio rerio* and *Drosophila melanogaster* (0.074/0.073/0.072/0.070)

Configuration models and z-score distributions

We use the configuration models to assess concrete differences between organism models. For each real-world UPR model, we reconstruct 10 configuration models, to compute the z-scores of closeness centrality, betweenness centrality, and clustering coefficient, reported in Table 3. *Rattus norvegicus* and *Mus musculus* show comparable values of z-scores for all three quantities ([-3.1, -3.1], [-1.8, -1.0], [3.1, 3.1] respectively). The *Drosophila melanogaster* model shows a significant difference from all the other organisms regarding the closeness centrality (the only model with a z-score of -2.0). Overall, z-scores of closeness and clustering coefficient are included in small variation ranges ([-3.0, -2.0] and [3.0, 3.1] respectively). Moreover, the z-score for betweenness centrality shows significant differences among organisms (from -2.7 for *Caenorhabditis elegans* to 1.9 for *Homo sapiens*), suggesting that betweenness centrality provides a potentially useful tool for identifying similarity or differences between organisms regarding this specific mechanism.

Configuration models and statistical analysis

Table 3 also shows the nonparametric signed-rank test results for closeness and betweenness centralities. Regarding closeness centrality, we obtain that the native and the configuration models are significantly different for all organisms. All the p-values for closeness are smaller (or much less) than 0.05, so the test rejects the null hypothesis of zero medians at the 5% significance level. On the other hand, for the betweenness centrality distribution, the test provides a non-significant difference for *Drosophila melanogaster* and *Arabidopsis thaliana*. Overall, the configuration models alter the network features regarding closeness and only partially regarding the betweenness centrality.

Highest metrics nodes

To more precisely relate the evaluated network metrics with the biological content, we also analyze the role of protein within the UPR pathways. Few specific genes or their homologues appear in all the sets across organisms, associated with high degree, closeness, and betweenness centralities. In Fig. 3 we show the nodes with the highest values of the three principal metrics, with the values in the round brackets.

A percentage varying between 5% and 30% of the total number of proteins in each native network is represented by heat shock cognate proteins (HSC), which are members of the heat shock protein family (HSP), one of the most ubiquitous and conserved protein families across organisms[78–81]. They are fundamental in the

Organism	Degree	Closeness	Closeness	Closeness	Betweenness	Betweenness	Betweenness	Clust. Coeff.
	Norm. value	Norm. value	z-score	p-value	Norm. value	z-score	p-value	z-score
Homo sapiens	0.113	0.440	-3.1	≪0.05	0.063	1.9	< 0.05	3.1
Rattus norvegicus	0.088	0.408	-3.1	≪0.05	0.042	-1.0	≪0.05	3.1
Mus musculus	0.092	0.417	-3.0	≪0.05	0.040	-1.8	≪0.05	3.1
Macaca fascicularis	0.071	0.377	-3.1	≪0.05	0.074	-0.3	≪0.05	3.1
Bos taurus	0.078	0.386	-3.1	≪0.05	0.053	0.1	≪0.05	3.1
Oryctologus cuniculis	0.086	0.404	-3.0	≪0.05	0.065	1.1	≪0.05	3.0
Gallus gallus	0.064	0.353	-3.1	≪0.05	0.073	0.6	≪0.05	3.1
Danio rerio	0.071	0.383	-3.1	≪0.05	0.072	1.5	≪0.05	3.0
Drosophila melanogaster	0.234	0.501	-2.0	< 0.05	0.070	-1.6	> 0.50	3.0
Caenorhabditis elegans	0.102	0.368	-3.1	≪0.05	0.048	-2.7	< 0.05	3.1
Saccharomyces cerevisiae	0.126	0.422	-3.1	≪0.05	0.092	-1.3	< 0.05	3.1
Arabidopsis thaliana	0.309	0.563	-3.1	≪0.05	0.135	0.7	> 0.50	3.0

Table 3. Normalized values and statistical analyses of common network features for native models. p-values are from the non-parametric signed-rank test applied to paired normalized distributions of native and configuration model metrics.

Organism	Degree	Closeness (·10⁻³)	Betweenness
Homo sapiens	HSPA5 (122), HSPA4 (100), HSP90AA1 (97), HSP90AB1 (93), HSP90B1 (90), HSPA8 (89)	HSPA5 (3.1), HSPA4 (2.9), HSP90AA1 (2.8), HSP90AB1 (2.8), HSP90B1 (2.8), HSPA8 (2.8)	HSPA5 (2309), VCP (1391), HSPA4 (1384), ERN1 (1003), HSPA8 (965), HSP90B1 (945)
Rattus norvegicus	Hspa5 (110), Hsp90b1 (82), Hspa4 (81), Hspa1b (74), Hsp90aa1 (74)	Hspa5 (3.1), Hspa4 (2.8), Hsp90b1 (2.7), Hspa1b (2.7), Hsp90aa1 (2.6), Xbp1 (2.6), Atf6 (2.5), Vcp (2.5)	Hspa5 (3882), Vcp (1662), Hsp90b1 (1313), Hspa4 (1243), Xbp1 (1205), Hsp90aa1 (1113)
Mus musculus	Hspa5 (123), Hsp90b1 (97), Hspa1b (92), Hsp90aa1 (90), Hsp90ab1 (88)	Hspa5 (2.9), Hsp90b1 (2.6), Hspa1b (2.5), Hsp90aa1 (2.5), Xbp1 (2.4)	Hspa5 (4230), Vcp (1960), Hspa1b (1457), Hsp90b1 (1406), Xbp1 (1405), Ern1 (1043), Hsp90aa1 (1028)
Macaca fascicularis	HSPA5 (67), HSP90B1 (55), HSPA4 (50), HSP90AB1 (47), HSP90AA1 (46)	HSPA5 (3.0), HSP901B (2.9), HSPA4 (2.7), XBP1 (2.7), HSP90AB1 (2.7), HSP90AA1 (2.7), DNAJC3 (2.7), HSPA1L (2.6), VCP (2.6)	VCP (2166), HSPA5 (2065), HSP90B1 (1268), XBP1 (1121), CALR (909)
Bos taurus	HSPA5 (94), HSPA1A (70), HSP90B1 (68), HSPA8 (56), HSP90AA1 (56)	HSPA5 (3.0), HSP90B1 (2.6), HSPA1A (2.6), XBP1 (2.5), DNAJC3 (2.5), HSPA8 (2.5), ATF6 (2.5)	HSPA5 (3278), VCP (1453), HSPA1A (1326), ATF6 (1241), XBP1 (1046)
Oryctologus cuniculus	HSPA5 (81), HSP90B1 (69), HSPA9 (51), HSPA8 (50), HSPA4 (49), HSP90AA1 (48)	HSPA5 (3.5), HSP90B1 (3.3), XBP1 (3.1), HSPA4 (3.0), HYOU1 (3.0), HSP90AA1 (3.05), HSPA8 (3.0), DNAJB9 (3.0), HSP90AB1 (3.0)	HSPA5 (2062), VCP (1351), JUN (1182), HSP90B1 (1127)
Gallus gallus	HSPA5 (64), HSP90B1 (50), HSPA8 (48), HSPA9 (45), HSP90AA1 (44), HSP90BA1 (43), HSPA4 (42)	HSPA5 (3.0), HSPA8 (2.8), HSP90B1 (2.7), HSP90AB1 (2.7), HSP90AA1 (2.6), HSPA9 (2.6), XBP1 (2.6), HSPA4 (2.6)	HSPA5 (2383), VCP (1274), CDH1 (1009), HSP90B1 (995), XBP1 (921)
Danio rerio	hspa5 (86), hsp90b1 (70), hsp90aa1.1 (61), hspa9 (59), hsp70.1 (59), xbp1 (54), hspa8 (54)	hspa5 (2.6), hsp90b1 (2.6), xbp1 (2.5), hsp70.1 (2.5), hsp90aa1.1 (2.4), hspa8 (2.4), atf6 (2.4), hspa9 (2.4), dnajc3a (2.3)	hspa5 (2516), xbp1 (2117), hsp90b1 (1274), atf6 (1198), vcp (1151), sec63 (1132), jun (1045)
Drosophila melanogaster	Hsc70-3 (40), Gp93 (33), Hsp83 (31), Hsp70Bb (26), Hsc70-5 (26), Hsp70Ab (26), Droj2 (26)	Hsc70-3 (13.7), Gp93 (12.3), Hsp83 (11.8), Droj2 (11.1), CG2918 (10.8), Hsp70Bb (10.5), Hsc70-5 (10.5), Hsp70Ab (10.5), Xbp1 (10.4)	Hsc70-3 (428), Xbp1 (218), Gp93 (192), Hsp83 (175)
Caenorhabditis elegans	hsp-90 (43), hsp-70 (41), hsp-4 (38), hsp-3 (36), hsp-6 (32), enpl-1 (31)	hsp-90 (5.0), hsp-70 (4.9), xbp-1 (4.9), hsp-4 (4.8), hsp-3 (4.7), hsp-6 (4.6), enpl-1 (4.6)	xbp-1 (2042), abu-1 (1474), hsp-90 (766), hsp-70 (506)
Saccharomyces cerevisiae	KAR2 (67), HSC82 (62), HSP82 (62), SSA1 (52), SSA2 (50), LHS1 (50)	KAR2 (4.0), HSC82 (3.9), HSP82 (3.9), SSA1 (3.8), SSB1 (3.7), HSP104 (3.7)	KAR2 (1180), ATG8 (900), HSC82 (619), SSA1 (570), SSA2 (547), HSP82 (543), SSB1 (532)
Arabidopsis thaliana	BIP3 (49), BIP2 (48), HSP70-4 (47), HSP90-7 (46), HSP90-1 (36), HSP90-3 (35), HSP90-2 (35)	BIP3 (13.3), BIP2 (13.3), HSP70-4 (13.2), HSP90-7 (12.8), HSP90-1 (11.4), HSP90-2 (11.2), HSP90-3 (11.2), HSP90-6 (11.1), HSP90-5 (11.1)	BIP3 (190), BIP2 (172), HSP70-4 (132), HSP90-7 (119), CPN60 (105), CCT6A (99), BZIP60 (93)

Fig. 3. Protein sets with highest values of network metrics, for native models. Values are reported in round brackets after the protein name.
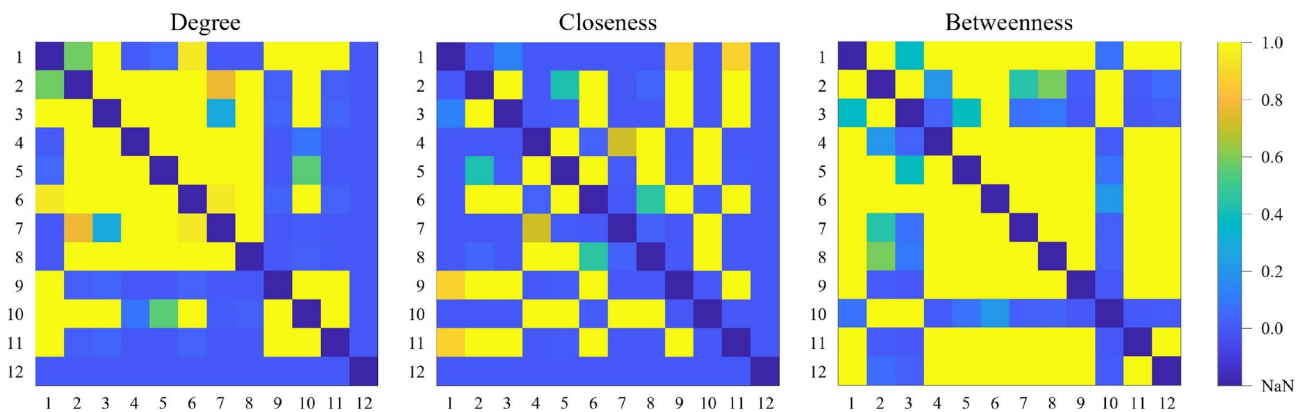
Fig. 4. Correlation matrices (p-value) of normalized metrics distributions. Matrix indices represent the organisms as listed in the text. The color scale is the same for all the matrices (color bar on the right). NaN elements identify the diagonal of the matrices (same organism).

correct functioning of cells, maintaining cellular proteostasis, and protecting cells from induced stresses[82]. Their genes are associated with the highest values of the main three network features, across all the twelve species. They result in relevant nodes and hubs in various diseases, like cancers and strokes[83–85]. HSP/HSC proteins for each organism belong to the same module, associated with the largest node size (Fig. 2). Other proteins related to the highest values of the three metrics for the various networks are i) X-box binding protein 1 (XBP1), which is an important initiator and modulator factor of ER stress response[86,87], ii) ATF6 and iii) ERN1, both of which are ER membrane receptors (together with EIF2AK3)[9,88–90], in charge of initiating and regulating the stress response after the activation promoted by HSPA5/BiP[91–93]. The most relevant pathway in the UPR mechanism for *Saccharomyces cerevisiae* and *Arabidopsis thaliana* is related to the IRE1/ERN1 signaling cascade[94–99], also notable from the fact that we cannot find homologs for the other two ER stress sensors. Finally, a recurrent protein among vertebrates is the Valosin-containing protein (VCP) which is one of the most abundant cytoplasmic proteins in eukaryotic cells[100]; its main function is to mediate protein quality control processes to maintain cell homeostasis, like ERAD[101].

Multi-comparison test In Fig. 4 we report the correlation matrices related to multiple comparison tests with the Bonferroni correction applied to normalized metrics distributions (degree, closeness centrality, and betweenness centrality) cross-species. The representation of metrics distributions is shown in Fig. S5 of Supplementary material.

Degree − We observe that normalized degree strong similarities among vertebrates (big yellow square between 2−*Mus musculus* and 8−*Danio rerio*). Interestingly, *Homo sapiens* distribution is similar, besides the

two murine models and the rabbit, to *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*. Nicely, *Caenorhabditis elegans* distribution is similar to 7 out of 12 organism models.

Closeness centrality − This feature distributions show a strong overlap between *Rattus norvegicus* and *Mus musculus*. Interestingly, these two models together with the human and the rabbit produce a similarity with *Drosophila melanogaster* and *Saccharomyces cerevisiae*, while *Macaca fascicularis*, *Bos taurus Gallus gallus* and *Danio rerio* models produce a similarity with *Caenorhabditis elegans*.

Betweenness centrality − On the contrary, the correlation matrix of betweenness centrality shows that most organisms have identical normalized distributions (most matrix elements are closer to 1), so this metric does not discriminate among all organisms.

We can conclude that statistical and correlation analyses show that degree and closeness centrality result in better network features to discriminate across organisms. Fig. S6 of Supplementary material shows the dendrograms of hierarchical analysis of normalized average degree and centralities. The degree and closeness trees identify similarities among vertebrates, and nicely with *Caenorhabditis elegans* and *Saccharomyces cerevisiae* (animal models). In all cases, the two murine models are among the most similar. Fig. S7 of Supplementary material shows the dendrograms of hierarchical analysis of normalized median of degree and centralities distributions (Fig. S5). The situation slightly changes from the analysis on average values. Again, *Rattus norvegicus* and *Mus musculus* are grouped as similar. Nevertheless, closeness centrality confirms the phylogenetic hierarchy obtained with average values; the degree groups the murine models with fruit fly and worm models, while *Arabidopsis thaliana* is grouped with the rest of the invertebrates. Betweenness centrality now provides a situation analog to the hierarchical organization of the degree for average values.

Topological analysis

A different type of information comes from the topological analysis of adjacency matrices, evaluated with the Generalized Hamming Distance (GHD, Table 4). It provides a degree of difference between two $N \times N$ matrices, by comparing paired matrix elements[49]. In Table 4 we report the average value of GHD calculated for the native network concerning all the 10 associated configuration models. The most different models are provided by *Drosophila melanogaster* (0.219), *Arabidopsis thaliana* (0.198), and *Caenorhabditis elegans* (0.135); this can be rationalized because the size of the network is small compared to the other organisms, and we have poor biological information about the nodes, so the null models generate quite different related networks. All vertebrate models provide small and comparable GHD values (< 0.100), especially *Mus musculus* (0.051), *Homo sapiens* (0.057), *Danio rerio* (0.059), and *Rattus norvegicus* (0.060).

To sum up, degree and closeness centrality are convenient network features since they discriminate across different organisms and identify native or null networks within the same organism. Summarizing the results for all the above quantities, the vertebrates (*Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, and *Bos taurus*) and their biological models (*Danio rerio*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*) share similar values, in particular of closeness, and have similar connections, as shown by the correlation matrices and the GHD values. Overall, the description of UPR networks via standard network quantities and GHD allows us to identify phylogenetic similarities and characterize the networks of vertebrates concerning other phyla.
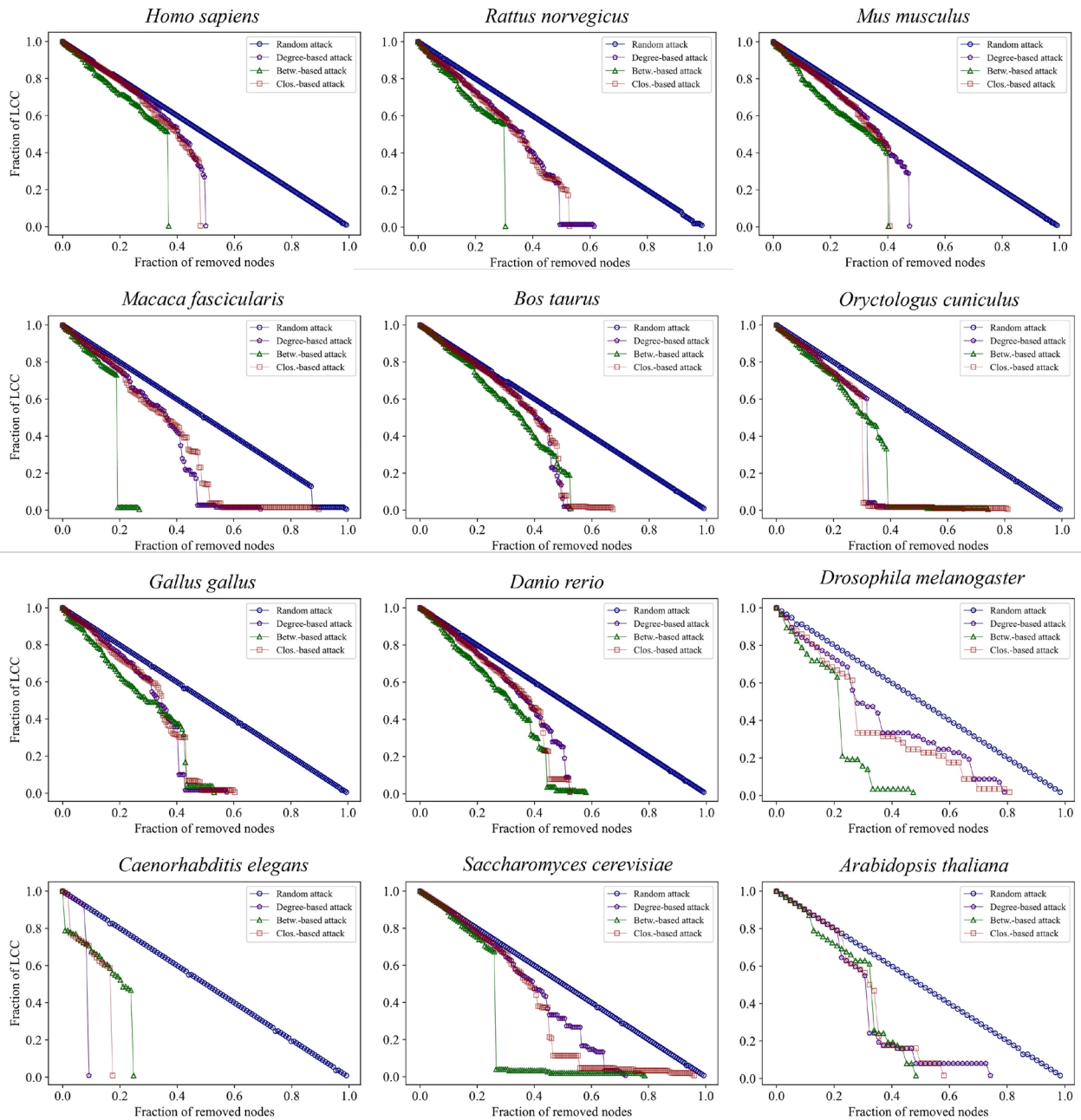
Network robustness

The network robustness (Fig. 5) has been tested with random- and metrics-based target attacks on the native networks. Table 5 reports normalized average path length and efficiency values for all investigated models. The average shortest path length values are comparable across organisms (from the lowest value for *Drosophila melanogaster* (2.068) to the largest value for *Gallus gallus* (2.949)). Murine models have the highest efficiency values (> 0.600), which also share comparable values of average path length.

The evolution of the largest connected component (LCC) in each network is evaluated by removing at each step one node from the network based on ascending index (random)[102], degree[103], and centralities[104] attacks, i.e. the least "important" nodes are removed first. The LCC identifies a connected component of a given graph that contains a significant fraction of the entire graph's vertices. The removal based on random node choice provides a linear trend for all organisms (Fig. 5). A random attack strategy (based on random node removal) requires the removal of many nodes to decrease the potency significantly, so targeted attacks are more efficient in degrading the network[105,106]. In most cases, the betweenness-based attack strongly affects the behavior of the network, except for *Caenorhabditis elegans*, for which the opposite is observed. In general, degree-based and closeness-based attacks show the same behavior of LCC degradation. This finding is also supported by sets of nodes with the highest values of metrics (see Fig. 3), where it is shown that sets for degree and closeness are more similar compared to sets obtained for betweenness centrality, as also confirmed by the computed correlation coefficients between network metrics (see above). Furthermore, vertebrate models show similar behaviors to all attacks.

By removing nodes with increasing metrics values, we expect a slower degradation of the network integrity; the results could be explained by considering a larger number of "external" and less important nodes within the networks (low values of network metrics). Attacking the networks based on decreasing metrics values provides a random-like trend in network robustness. The sudden jumps in the robustness can be associated not only with a precise role of a specific protein removal, because their removal at the beginning of the process could not produce the same results, as in a purely star network. Instead, it can also be the result of a cumulative effect,

H. sapiens	R. norvegicus	M. musculus	M. fascicularis	B. taurus	O. cuniculis	G. gallus	D. rerio	D. melanogaster	C. elegans	S. cerevisiae	A. thaliana
0.057	0.060	0.051	0.076	0.064	0.076	0.071	0.059	0.219	0.135	0.095	0.198

Table 4. Average Generalized Hamming Distance between native and configuration models.

Fig. 5. Network robustness evaluated using random attack and target attacks on native models. Different colors identify attack strategy trends.

because the overall removal of all the previous proteins builds a star-like network, that collapses upon removal of a specific protein.

However, going through the principal functions and pathways that are involved in the UPR and arising from the network robustness, across all, or most of, the organisms, we highlight that those proteins with the role of chaperones, i.e. initiating the signaling pathways and quality control of protein folding, are present in all considered organisms (except *Homo sapiens*) as the ones that induce a sudden collapse of metrics robustness. Also, proteins involved in pro-apoptosis or ERAD mechanisms emerge from the robustness analysis as present in most organisms. Other mechanisms, with corresponding proteins obtained from robustness stress analysis, are listed in Table 6.

We report the network robustness analysis applied to the configuration models in Fig. S8 of Supplementary material. We choose the nearest and the farthest models from the native models based on the network topology (lowest and biggest dGHD values) to investigate the behavior concerning the randomized networks. The nearest topology network of *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Gallus gallus*, and *Saccharomyces cerevisiae* shows a stronger resistance to fails than the farthest topology network (similar to the native models − Fig. 5). In addition, there is a (nearly) complete overlap of the three targeted attack trends for all organisms.

Organism	Average path length	Efficiency
Homo sapiens	2.354	0.488
Rattus norvegicus	2.543	0.638
Mus musculus	2.485	0.681
Macaca fascicularis	2.735	0.446
Bos taurus	2.693	0.586
Oryctologus cuniculis	2.557	0.593
Gallus gallus	2.949	0.504
Danio rerio	2.702	0.551
Drosophila melanogaster	2.068	0.571
Caenorhabditis elegans	2.832	0.429
Saccharomyces cerevisiae	2.450	0.478

Table 5. Values of average path length and efficiency of all native models.

Organism	Chaperone	MP,R,&T	Kinase	R&M	ERAD/Autoph./Pro-survival	Apopt.	I&T factors	Protein folding
Homo sapiens	–	–	MAPK8	MAPK8	TMBIM6, THBS4	MAPK8	–	–
Rattus norvegicus	CCt5, Dnajb2	Ep300, Dnajb2	–	Dnajb2	Dnajb2	–	–	CCt5
Mus musculus	Bag3	–	–	–	Bag3, Creb3, Manf	–	Creb3	–
Macaca fascicularis	HSPA9	–	–	DDIT3	–	HSPA9, DDIT3,BCL2L11	DDIT3	HSPA9
Bos taurus	DNAJA1, HSPA4L	–	–	HSF1	–	DNAJA1, HSF1	HSF1	–
Oryctologus cuniculis	DNAJC10	VAPB, ERO1B	–	–	DNAJC10, VAPB	DNAJC10	–	ERO1B
Gallus gallus	HSPD1, HSPA4, PDIA6	–	–	HSPA4	PDIA6	–	–	HSPD1, ERP44, HSPA4, PDIA6
Danio rerio	bag6, hyou1, cryaa	–	–	hyou1	bag6, hyou1, cryaa	bag6	–	hyou1, cryaa
Drosophila melanogaster	Gp93, DnaJ-1, Calr, Grp170	Calr	–	Dad1	Dad1, Der-2	–	–	Gp93, DnaJ-1, Calr, Grp170
Caenorhabditis elegans	enpl-1	col-109, abu-6, F38B6.6, abu-12	–	–	col-109, abu-6, enpl-1, abu-12	F38B6.6	pqn-90, enpl-1	–
Saccharomyces cerevisiae	CCT3, HSP10	BUD27	–	–	EDE1	–	–	BUD27, CCT3, HSP10
Arabidopsis thaliana	HSP90-3, HSP90-3, BIP3, HSP70-7, HSP70-10	–	–	–	HSP70-7, HSP70-10, BIP3	–	–	HSP70-7, HSP70-10, BIP3

Table 6. Proteins and functions related to jumps in the robustness trends. MP,R,&T = membrane proteins, receptors, and transfers; R&M = regulators and messengers; I&T factor = initiation and transcription factors. Autoph./Aptopt. = autophagy/apoptosis. Jump-related proteins are considered when the LCC dimension decreases by at least 10%.

Structural controllability

The structural controllability allows for determining the number and identifying the minimum driver nodes (Table 7). For all organisms, the study of the native model may produce different driver node sets; however, some proteins recur in sets across organisms, suggesting their importance within the networks and in the mechanism. Unfortunately, the algorithm cannot find any driver node set for the bull model, as it is.

The algorithm finds a unique set only for *Homo sapiens* and *Drosophila melanogaster* models. RPAP2 is associated with transcription and RNA processing, connects to RUBVL2, and has minimal network metrics[107]. In metrics terms, this node is characterized by 1 degree (connected to RUBVL2), $1.48 \cdot 10^{-3}$ closeness centrality, and 0 betweenness centrality (we can define it as an "external node"). In fruit flies, Hsp70Ab, which belongs to the Hsp70 family, stabilizes non-native protein conformations[108,109]. The L(2)efl protein[110] induces phosphorylation of Eif2α, playing a role in aging[111–113].

The FIC domain protein adenylyltransferase (Ficd) is found in *Mus musculus* and *Oryctologus cuniculis*. Ficd is associated with various cellular pathways, particularly the ATF6 and EIF2AK3 branches of the UPR pathway, which regulate ER homeostasis. In humans, FICD is typically present at low basal levels in most cell types, and its expression is tightly regulated[114,115].

Nicely, the Ufm1-specific ligase 1 (Ufl1)[116,117] is found in *Rattus norvegicus*, *Macaca fascicularis*, and *Gallus gallus*. The Ufl1 protein is essential for activating the stress sensor EIF2AK3 and helps protect cardiomyocytes from cell death triggered by ER stress.

Kinesins (Krt4, Krt14, and Krt16) are found in *Rattus norvegicus* and *Gallus gallus*[118,119].

Organism	Driver nodes	# possible sets
Homo sapiens	1 (RPAP2)	1
Rattus norvegicus	3 (Ufl1, Rhbdd2, Krt14/Krt16)	2
Mus musculus	4 (Ficd, Parp8, Crebzf/Crebrf, Il24/Igtp)	4
Macaca fascicularis	4 (TOR1B, SIAH2, HSP1/HSPA4L, UFL1/DDRGK1)	4
Bos taurus	–	–
Oryctologus cuniculis	3 (FAP/LOXL2, FICD/SERPINH1, PTPN2/TNIK)	8
Gallus gallus	3 (KRT6A, KRT4/KRT14, CDKSRAP3/UFL1)	4
Danio rerio	2 (tnika, parp6b/parp8)	2
Drosophila melanogaster	2 (Hsp70Ab, L(2)efl)	1
Caenorhabditis elegant	3 (abu-12, abu-5/pqn-79, pqn-70/pqn-90)	3
Saccharomyces cerevisiae	4 (TSR4/SYO1, NUP42/ASM4, NUP1/NUP60, NUP145/NUP159)	16
Arabidopsis thaliana	9 (DJA6, GFA2, T14C9.70/T13C7.14/T17F15.190/F14N23.23/F23H11.4/T10O8.100/T16O11.15/DJA5)	8

Table 7. Structural controllability results for all species models (DNs (set dimension) and number of possible node sets). Identified proteins for each model are listed in the round brackets.

The gene family activated in blocked UPR (abu) in *Caenorhabditis elegans*is implicated in the UPR regulation in response to the ER stress. They help maintain protein-folding homeostasis and manage the accumulation of misfolded proteins in the ER[69]. This class is a subset of the pqn family, prion-like Q/N proteins[120].

Interestingly, many resulting proteins are chaperones, such as heat shock proteins (HSP) and DnaJ homologues (i.e., all *Arabidopsis thaliana* proteins). Their chaperone and protein-folding regulation activity is well known.

Discussion

Unfolded Protein Response is one of the most conserved fundamental biological mechanisms through organisms[15,16,75,121], occurring primarily in the endoplasmic reticulum (ER). Its main function is to restore cell homeostasis after a pathological accumulation of non-properly folded proteins[6,7,10,11,122,123]. Any inefficiency in the adaptive response to ER stress can lead to unfolded or misfolded protein accumulation at different levels. These proteins tend to aggregate, posing a threat to cellular and tissue integrity and serving as a primary driver for the onset of amyloidosis and neurodegenerative diseases[25–27,124].

A key signaling pathway governing the UPR, originally discovered in *Saccharomyces cerevisiae*during the 1970s[16,50,125,126], is characterized by a single transmembrane protein, ERN1, responsible for the detection of ER stress provoked by over-accumulation of unfolded/misfolded proteins[94–96]. The major ER chaperone BiP triggers the dimerization of ERN1, which leads to its subsequent autophosphorylation and the activation of its signaling cascade[9]. This pathway reinforces the ER function and is conserved across eukaryotes[127]. Indeed, the basic features of the UPR mechanism result in being highly preserved throughout metazoans; most species have homologues of the three main stress sensors, ERN1, ATF6, and EIF2AK3. Signaling pathways of the stress sensors cooperate to restore and/or bolster ER function, primarily through the upregulation of many components of the protein folding machinery (as the action of XBP1 on the regulation of BiP chaperone[86,87]within the ERN1 pathway) and the quality control machinery within the ER. Additionally, these signaling pathways limit ER stress by dampening the translation attenuation and potentially engaging the regulated ERN1-dependent decay (known as RIDD)[128–130]. A network-based description of cell mechanisms, using protein-protein interactions (PPI) networks, offers a valuable tool for comprehending the behavior of complex systems. Some points must be considered before discussing the analysis results. PPI network models are built utilizing biological data from publicly available PPI databases. These databases collect various types of data, from thousands of experimental works. Despite their large size, the databases are not exhaustive, because only part of the molecular pathways have been completely understood and characterized. The absence or bias in information stored in interaction databases must be considered when creating and analyzing biological networks. PPI databases must contain sufficient information for a specific pathway to yield qualitative accurate results from network analysis techniques.

With these premises, we apply a classical network metrics investigation to assess the similarities and differences among UPR models from twelve different organisms, considering the existing phylogenetic pathway in all the analyzed models. As model organisms, we choose for our investigation *Homo sapiens*, *Rattus norvegicus*, and *Mus musculus*due to the quite complete genomic accordance with human[131–136] and other model organisms, as *Macaca fascicularis*[55], *Bos taurus*[56], *Oryctologus cuniculis*[57,58], *Gallus gallus*[59–63], *Drosophila melanogaster*[64–66], *Caenorhabditis elegans*[67–69], *Saccharomyces cerevisiae*[16,70,71], and *Arabidopsis thaliana*[72–75].

Comparisons made for common metrics provide that network measures are sensitive to the network features (Table 2, and Figs. S2, S3, S4 from Supplementary material); however, from a biological point of view, the network analysis shows that there are strong similarities among the model organisms (Table 3), in particular among vertebrates. Moreover, highly conserved genes and pathways across species arise (Fig. 3). Chaperones belonging to the heat shock protein (HSP) family are among the most central proteins regarding network metrics, as the densest connected, and they also play a fundamental role in the correct functioning of cells[78–82]. This result can be interpreted as support for the significance of these proteins, as it implies a correlation between the structure and the biological properties of protein networks. In summary, network theory and statistical mechanics confirm that it is possible to identify similarities among organisms phylogenetically related via their PPI networks. In

addition, we find that some network metrics are better at discriminating among organisms; they are useful for identifying similarities and differences, such as closeness centrality and node degree.

Additionally, in this study, we employ a network scientific methodology to investigate the tolerance levels of multiple systems when subjected to external perturbations. We achieve this by adopting a measure of network robustness[137–140], to characterize the potential resilience of several PPI networks. The robustness of the twelve organism models is evaluated by removing nodes, and consequently altering the network integrity, adopting different network-based metrics target strategies (Fig. 5). Obviously, in these cases, network robustness is influenced by network features and the degree of accuracy of biological information accessible online, but similarities in resilience behavior arise for organisms that are phylogenetically closer, for example, *Homo sapiens* and murine species, despite the difference between network features. This result can be supported by the closeness between these organisms in terms of phylogeny since 99% of the genome is conserved between human and murine species[131–136]. A detailed analysis of genes related to jumps in the network robustness highlights that, despite different proteins being involved in the organisms, there is a recurrence of pathways across most species. In particular, the sensing role of the chaperones is fundamental, together with the apoptosis and endoplasmic-reticulum-associated protein degradation (ERAD) functions (Table 6). Lastly, the minimum drive nodes methodology, based on the structural controllability analysis[41,42] and employing Kalman's rank condition[141], yields significant biological insights into the proteins involved in specific mechanisms. It becomes feasible to pinpoint the key nodes that exert control over the entire network[142]. We apply structural controllability to all organism models investigated in this work. Driver node sets are related to the topological structure of the network; as for the other network methods applied, results are sensitive to the biological information available online (also in the high variability of the possible DNs sets − several sets for the same organism in Table 7). Attributing a biological significance to the set of driver nodes is not straightforward. When analyzing networks with numerous nodes and edges, the possibility of losing biological information in the models is due to the consequent increase of missing information stored in the databases. However, some specific mechanisms and genes, already identified via robustness and network statistics, also arise through structural controllability.

This study represents a comprehensive and innovative analysis of the biological behavior and characteristics of a fundamental cellular control mechanism, the Unfolded Protein Response (UPR). By modeling the UPR as a protein-protein interaction (PPI) network, it uses standard and advanced techniques to extract meaningful information from the network. The study establishes a direct correlation between specific network features and biological components. It finds that the three methods employed − standard network metrics and topology, network robustness, and network control theory − offer complementary and non-conflicting characterizations of the systems studied.

The first class of methods offers a comprehensive portrayal of the PPI networks and discerns phylogenetic similarities, distinguishing vertebrates from organisms of other phyla. Additionally, it identifies key genes (network nodes) that significantly influence the overall UPR mechanism. Network robustness utilizes metrics to assess the network's ability to withstand node removal, identify pivotal proteins and sub-pathways, and simulate disease onset. This robustness analysis also highlights phylogenetic similarities among organisms.

Among the various UPR sub-mechanisms, chaperone activity, apoptosis, and ERAD are identified as relevant. Though less distinctly, control theory also pinpoints proteins with central roles, particularly chaperones, transcription factors, and ERAD proteins. Using network models for molecular and cellular pathways is a powerful yet underutilized approach. This study's combination of diverse network descriptors and methods provides profound insights into complex mechanisms. It emphasizes revising and enriching PPI databases to create increasingly accurate biological models.

Materials and methods

Here we present a study of the unfolded protein response mechanism among twelve organisms. Below we report general information of the network models investigated in this study and the description of the performed analyses. We perform calculations of i) network descriptors, ii) modularity and communities, iii) topological distance, iv) network robustness, and v) structural controllability. We perform comparisons among native models, with the help of the configuration models (degree-based reconstruction). Native models are built using Python v. 3.11 programming language and NetworkX Python library[143], based on information on the paired connection between couples of nodes, and configuration models are built using MATLAB v. R2023a programming language. Modularity, community detection, and robustness analyses are performed using the NetworkX library, while network descriptors analysis and structural controllability[144,145] have been implemented in MATLAB.

Unfolded protein response network models

Here we present the UPR network models proposed in this work. Investigated organisms are *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Macaca fascicularis*, *Bos taurus*, *Oryctologus cuniculis*, *Gallus gallus*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana*. We combine information stored in the UniProt database[146–148] and the String protein-protein interaction database[149,150] to identify proteins involved in the mechanism and build undirected network models. Since the available information in the UniProt database is poor for the vertebrates (old network features of organisms models are reported in Table S1 of Supplementary material), we start from the proteins found for *Homo sapiens*, which results as the most characterized organism, to reconstruct the PPI networks for *Rattus norvegicus*, *Mus musculus*, *Macaca fascicularis*, *Bos taurus*, *Oryctologus cuniculis*, *Gallus gallus*, and *Danio rerio*. We therefore combine the proteins listed in the UniProt database for the specific organism with the human list. In addition, we manually curate the literature to include the proteins (chaperone BiP and three stress sensors with their pathways) from previous work[151] in the vertebrate models. We impose a minimum required interaction score of medium confidence (0.400, i.e., useful interactions to build the networks must have scores ≥ 0.400), and

we construct the set of connections among nodes considering experimental evidence, curated databases, text mining, and co-expression associations. To analyze the networks, we download the lists of paired interactions as Tab Separated Values (TSV) extension files from the STRING website. In this framework, original models, also named native, are built directly from biological information obtained from the databases, and connections in configuration models are semi-randomly built based on the total degree of each node in the original models. The matrix representation of native UPR models is shown in Fig. S1 of Supplementary material.

Analysis of network models using network metrics

We first characterize the models using the usual metrics of network descriptors, (i) to highlight differences and similarities between models related to different organisms and (ii) to evaluate the influence of constructing a network with random connections starting from the degree of nodes of original models. For each model, we calculate (i) the total degree of the nodes, (ii) the betweenness and closeness centralities, which are measures of how often each graph node appears on the shortest path between two nodes in the graph, (iii) local clustering coefficient, and (iv) the modularity and number of communities, as measures of the structure of networks to evaluate the strength of division into different modules, calculated using the Louvain community detection Algorithm. As a definition, networks with high modularity have dense connections within communities. In addition, we also include normalized values of degree ($\text{norm(deg)} = 2M/N$, where $2M$ is the total degree of a node in a Nnodes network)[152], betweenness ($\text{norm(bet)} = (\text{bet}-\text{min(bet)})/(\text{max(bet)}-\text{min(bet)})$) and closeness centrality ($\text{norm(clos)} = (N-1) \cdot \text{clos}$)[153] (Table 3). All measures are normalized by removing the dependence from the network dimensionality (normalized distributions shown in Figs. S2, S3, S4 of Supplementary material). We compute z-scores of averaged normalized metrics distributions, defined as $(x - \mu)/\sigma$, where x is the variable value, μ and σ are the population's mean value and standard deviation, respectively. Highest metrics (degree and centralities) proteins are extracted directly from metrics distributions.

Topological differences between native and configuration models

As well explained in our previous work[151], we use the algorithm relies on the Generalized Hamming Distance (GHD)[49], which can be used for assigning a "weight" to the topological difference between networks and evaluating its statistical significance, based on comparison between matrix elements. We apply this theory to assess the difference between the original models (built from the databases) and the configuration models (re-created from the original models). If we consider two distinct networks, labeled X and Y, with the same number of nodes (N), we can calculate the distance dGHD between the two networks as follows:

$$\text{dGHD}(X, Y) = \frac{1}{N \cdot (N-1)} \sum_{i \neq j} (x'_{ij} - y'_{ij}) \tag{1}$$

where x'_{ij} and y'_{ij} are mean-centered edge-weights, and depend on the network topology, providing a measure of connectivity between every pair of i$^{\text{th}}$ and j$^{\text{th}}$ nodes in X and Y, respectively.

Statistical analysis of network models

The graph analysis and network robustness are performed using Python 3.11[154]. The statistical analysis, the implementation of structural controllability, and the GHD algorithm are performed using MATLAB R2023a[155]. We apply non-parametric tests because our variables are not normally distributed. For paired comparisons between the centralities distributions of native and null models considering one specific organism we use the Wilcoxon signed-rank test, and for non-paired comparisons across different organisms we use a multi-comparison test computing values with the Bonferroni method[156,157], on the results of a one-way Kruskal Wallis analysis of variance (shown in Fig. S5 of Supplementary material). If $p < 0.05$, the results are considered as statistically significant.

Network robustness exploitation

Considering a network X, composed of N nodes set denoted as $V = \{v_1, v_2, \ldots, v_N\}$, interconnected by M links represented by $E = \{(v_i, v_j) : v_i, v_j \in V\}$, the robustness R of the Xnetwork is defined by the ratio[43,158]

$$R = \frac{1}{N} \sum_{i=1}^{N} G_i \tag{2}$$

where $G_i = n_i/N$ is the size of the large connected component after the i$^{\text{th}}$ node removal; normalization factor N^{-1} is useful for comparing networks of different sizes. Some metrics can be used to quantify a network's robustness. The average path length $l$can provide a quantification of network robustness since large values mean that nodes are farther apart from each other, and the removal of a node can significantly increase the average paths between nodes, decreasing the robustness of the network[159]:

$$l = \frac{1}{N \cdot (N-1)} \sum_{i \neq j} d_{ij} \tag{3}$$

where the sum of all possible paired-node distances is normalized over all the possible couples of N nodes. Another useful metric to quantify network robustness is the variation of the efficiency ΔEdepending on an

increasing number of removed nodes[43]. A robust network would have a small drop in network efficiency with node removal.

$$\Delta E_i = \frac{E - E_i}{E} \tag{4}$$

where the efficiency is defined as $E = (N \cdot (N-1))^{-1} \sum_{i \neq j} d_{ij}^{-1}$. In the investigation presented here, the theory of network robustness is employed to assess the capability of networks to deliver and maintain an acceptable level of service in the presence of faults, as outlined by our models. Our analysis involves subjecting each network to random and targeted attacks, following specific strategies (degree-, closeness-, and betweenness-based).

Structural controllability and minimum driver node identification

Lastly, we employ the structural controllability theory to assess the nodes' ranking using Kalman's rank condition for continuous linear time-invariant systems[141]. In addition, we also implement the Minimum Driver Nodes (MDN) algorithm[144,145], proposed by Liu et al., which is based on the minimal set of input signals required to control the network, and the MDN selection algorithm used by Zhang et al[142]., which can be used to identify the driver nodes − the nodes on which an input signal must be injected to obtain full control of the network. Generally, the time-evolution of a network system consisting of N nodes and M input signals, with $M \leq N$, can be described with the following linear differential equation

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{5}$$

where $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_N)^{\mathrm{T}}$ is the state vector for the N nodes system and $\mathbf{u} = (u_1, u_2, u_3, \ldots, u_M)^{\mathrm{T}}$ is the control vector. $\mathbf{A}$ is the $N \times N$ state matrix, in which each element a_{ij} identifies the connection between the i$^{\text{th}}$ and j$^{\text{th}}$ nodes. $\mathbf{B}$ is the $M \times N$ control matrix, whose dimension M depends on the number of input signals:

$$\mathbf{B} = (e_1^{\mathrm{T}} \; e_2^{\mathrm{T}} \; e_3^{\mathrm{T}} \; \ldots \; e_M^{\mathrm{T}}) \tag{6}$$

where $\{e_1, e_2, e_3, ..., e_M\}$ are the vectors of the canonical base. Given $\mathbf{A}$ and $\mathbf{B}$, it is possible to assemble the controllability matrix $\mathbf{C}$:

$$\mathbf{C} = (\mathbf{B}, \; \mathbf{AB}, \; \mathbf{A}^2\mathbf{B}, \; \mathbf{A}^3\mathbf{B}, \; \ldots, \; \mathbf{A}^{N-1}\mathbf{B}) \tag{7}$$

The network is fully controllable if the controllability matrix has a full rank, i.e., $\mathrm{rank}(\mathbf{C}) = N$. Theory and algorithms are well-explained in Ref[151].

Data availability

The datasets generated during and analyzed during the current study are available from the corresponding author upon reasonable request. Source code (Python and MATLAB scripts) used during the current study are available at https://github.com/NLuchetti/StatMech_of_UPR.git.

References

1. Schröder, M. & Kaufman, R. J. The mammalian unfolded protein response. *Annual Review of Biochemistry* **74**, 739–789 (2005).
2. Malhotra, J. D. & Kaufman, R. J. The endoplasmic reticulum and the unfolded protein response. *Seminars in Cell & Developmental Biology* **18**, 716–731 (2007).
3. Määttänen, P., Gehring, K., Bergeron, J. J. & Thomas, D. Y. Protein quality control in the er: The recognition of misfolded proteins. *Seminars in Cell & Developmental Biology* **21**, 500–511 (2010).
4. Stolz, A. & Wolf, D. H. Endoplasmic reticulum associated protein degradation: A chaperone assisted journey to hell. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1803**, 694–705 (2010).
5. Uversky, V. N. Natively unfolded proteins: A point where biology waits for physics. *Protein Science* **11**, 739–756 (2002).
6. Hetz, C. & Papa, F. R. The unfolded protein response and cell fate control. *Molecular Cell* **69**, 169–181 (2018).
7. Hetz, C. The unfolded protein response: controlling cell fate decisions under er stress and beyond. *Nature Reviews Molecular Cell Biology* **13**, 89–102 (2012).
8. Wang, J., Lee, J., Liem, D. & Ping, P. Hspa5 gene encoding hsp70 chaperone bip in the endoplasmic reticulum. *Gene* **618**, 14–23 (2017).
9. Kopp, M. C., Larburu, N., Durairaj, V., Adams, C. J. & Ali, M. M. U. Upr proteins ire1 and perk switch bip from chaperone to er stress sensor. *Nature Structural & Molecular Biology* **26**, 1053–1062 (2019).
10. Shen, J., Chen, X., Hendershot, L. & Prywes, R. Er stress regulation of atf6 localization by dissociation of bip/grp78 binding and unmasking of golgi localization signals. *Developmental Cell* **3**, 99–111 (2002).
11. Shen, J., Snapp, E. L., Lippincott-Schwartz, J. & Prywes, R. Stable binding of atf6 to bip in the endoplasmic reticulum stress response. *Molecular and Cellular Biology* **25**, 921–932 (2005).
12. Sharma, P. et al. Autophagy, apoptosis, the unfolded protein response, and lung function in idiopathic pulmonary fibrosis. *Cells* **10**, 1642 (2021).
13. Brown, M. K. & Naidoo, N. The endoplasmic reticulum stress response in aging and age-related diseases. *Frontiers in Physiology* **3** (2012).

14. Brown, M. K. et al. Aging-induced endoplasmic reticulum stress alters sleep and sleep homeostasis. *Neurobiology of Aging* **35**, 1431–1441 (2014).
15. Bommiasamy, H. & Popko, B. *Animal Models in the Study of the Unfolded Protein Response* (2011).
16. Hollien, J. Evolution of the unfolded protein response. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1833**, 2458–2463 (2013).
17. Mori, K. Evolutionary aspects of the unfolded protein response. *Cold Spring Harbor Perspectives in Biology* a041262 (2022).
18. Krishnan, K. & Askew, D. S. *The fungal upr. Virulence* **5**, 334–340 (2014).
19. Moreno, A. A. & Orellana, A. The physiological role of the unfolded protein response in plants. *Biological Research* **44**, 75–80 (2011).
20. Samperna, S. et al. Arabidopsis defense against the pathogenic fungus drechslera gigantea is dependent on the integrity of the unfolded protein response. *Biomolecules* **11**, 240 (2021).
21. Raman, K. Construction and analysis of protein-protein interaction networks. *Automated Experimentation* **2**, 2 (2010).
22. Tomkins, J. E. & Manzoni, C. Advances in protein-protein interaction network analysis for parkinson's disease. *Neurobiology of Disease* **155**, 105395 (2021).
23. Safari-Alighiarloo, N., Taghizadeh, M., Rezaei-Tavirani, M., Goliaei, B. & Peyvandi, A. A. Protein-protein interaction networks (ppi) and complex diseases. *Gastroenterology and Hepatology from bed to bench* **7**, 17 (2014).
24. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* **5**, 101–113 (2004).
25. Scheper, W. & Hoozemans, J. J. M. The unfolded protein response in neurodegenerative diseases: a neuropathological perspective. *Acta Neuropathologica* **130**, 315–331 (2015).
26. van Ziel, A. M. & Scheper, W. The upr in neurodegenerative disease: Not just an inside job. *Biomolecules* **10**, 1090 (2020).
27. Ghemrawi, R. & Khair, M. Endoplasmic reticulum stress and unfolded protein response in neurodegenerative diseases. *International Journal of Molecular Sciences* **21**, 6127 (2020).
28. Glover, K. & Silverman, L. Characterization of structural controllability. *IEEE Transactions on Automatic Control* **21**, 534–537 (1976).
29. Shields, R. & Pearson, J. Structural controllability of multiinput linear systems. *IEEE Transactions on Automatic Control* **21**, 203–212 (1976).
30. Lin, C.-T. Structural controllability. *IEEE Transactions on Automatic Control* **19**, 201–208 (1974).
31. Uhart, M., Flores, G. & Bustos, D. M. Controllability of protein-protein interaction phosphorylation-based networks: Participation of the hub 14-3-3 protein family. *Scientific Reports* **6**, 26234 (2016).
32. Abdallah, C. T. Mathematical controllability of genomic networks. *Proceedings of the National Academy of Sciences* **108**, 17243–17244 (2011).
33. Vinayagam, A. et al. Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. *Proceedings of the National Academy of Sciences* **113**, 4976–4981 (2016).
34. Liu, S., Xu, Q., Chen, A. & Wang, P. Structural controllability of dynamic transcriptional regulatory networks for saccharomyces cerevisiae. *Physica A: Statistical Mechanics and its Applications* **537**, 122772 (2020).
35. Ackerman, E. E., Alcorn, J. F., Hase, T. & Shoemaker, J. E. A dual controllability analysis of influenza virus-host protein-protein interaction networks for antiviral drug target discovery. *BMC Bioinformatics* **20**, 297 (2019).
36. Kanhaiya, K., Czeizler, E., Gratie, C. & Petre, I. Controlling directed protein interaction networks in cancer. *Scientific Reports* **7**, 10327 (2017).
37. Gonzalez, O. *Protein-Protein Interaction Databases* (Springer, New York, 2013).
38. Bajpai, A. K. et al. Systematic comparison of the protein-protein interaction databases from a user's perspective. *Journal of Biomedical Informatics* **103**, 103380 (2020).
39. Nakajima, N., Akutsu, T. & Nakato, R. *Databases for Protein-Protein Interactions* (2021).
40. Lehne, B. & Schlitt, T. Protein-protein interaction databases: keeping up with growing interactomes. *Human Genomics* **3**, 291 (2009).
41. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
42. Liu, Y.-Y. & Barabási, A.-L. Control principles of complex systems. *Reviews of Modern Physics* **88**, 035006 (2016).
43. Liu, J., Zhou, M., Wang, S. & Liu, P. A comparative study of network robustness measures. *Frontiers of Computer Science* **11**, 568–584 (2017).
44. Oehlers, M. & Fabian, B. Graph metrics for network robustness-a survey. *Mathematics* **9**, 895 (2021).
45. Artime, O. *et al.* Robustness and resilience of complex networks. *Nature Reviews Physics* (2024).
46. Barthelemy, M. Betweenness centrality in large complex networks. *The European physical journal B* **38**, 163–168 (2004).
47. Smith, K. M. & Escudero, J. Normalised degree variance. Applied Network. *Science* **5**, 32 (2020).
48. Zhang, P. et al. A protein network descriptor server and its use in studying protein, disease, metabolic and drug targeted networks. *Briefings in bioinformatics* **18**, 1057–1070 (2017).
49. Bookstein, A., Kulyukin, V. A. & Raita, T. Generalized hamming distance. *Information Retrieval* **5**, 353–375 (2002).
50. Hetz, C., Zhang, K. & Kaufman, R. J. Mechanisms, regulation and functions of the unfolded protein response. *Nature Reviews Molecular Cell Biology* **21**, 421–438 (2020).
51. Du, Z. et al. Probing the unfolded protein response in long-lived naked mole-rats. *Biochemical and Biophysical Research Communications* **529**, 1151–1157 (2020).
52. Timberlake, M., Prall, K., Roy, B. & Dwivedi, Y. Unfolded protein response and associated alterations in toll-like receptor expression and interaction in the hippocampus of restraint rats. *Psychoneuroendocrinology* **89**, 185–193 (2018).
53. Naidoo, N. *The Unfolded Protein Response in Mouse Cerebral Cortex* (2011).
54. Rana, T. et al. An activated unfolded protein response promotes retinal degeneration and triggers an inflammatory response in the mouse retina. *Cell Death & Disease* **5**, e1578–e1578 (2014).
55. Umeda, S. et al. Molecular composition of drusen and possible involvement of anti-retinal autoimmunity in two different forms of macular degeneration in cynomolgus monkey (macaca fascicularis). *The FASEB Journal* **19**, 1683–1685 (2005).
56. Yonekura, S. et al. The unfolded protein response is involved in both differentiation and apoptosis of bovine mammary epithelial cells. *Journal of Dairy Science* **101**, 3568–3578 (2018).
57. Ghribi, O., Herman, M. M., Pramoonjago, P. & Savory, J. Mpp sup + /sup induces the endoplasmic reticulum stress response in rabbit brain involving activation of the atf-6 and nf-kb signaling pathways. *Journal of Neuropathology & Experimental Neurology* **62**, 1144–1153 (2003).
58. Kruzliak, P., Sabo, J. & Zulli, A. Endothelial endoplasmic reticulum and nitrative stress in endothelial dysfunction in the atherogenic rabbit model. *Acta Histochemica* **117**, 762–766 (2015).
59. Huo, Y. et al. Exposure to copper activates mitophagy and endoplasmic reticulum stress-mediated apoptosis in chicken (gallus gallus) cerebrum. *Environmental Toxicology* **38**, 392–402 (2023).
60. Gao, P.-C. et al. Selenium alleviates endoplasmic reticulum calcium depletion-induced endoplasmic reticulum stress and apoptosis in chicken myocardium after mercuric chloride exposure. *Environmental Science and Pollution Research* **30**, 51531–51541 (2023).
61. Lin, Y.-F., Sam, J. & Evans, T. Sirt1 promotes tissue regeneration in zebrafish through regulating the mitochondrial unfolded protein response. *iScience* **24**, 103118 (2021).

62. Vacaru, A. M. et al. Molecularly defined unfolded protein response subclasses have distinct correlations with fatty liver disease in zebrafish. *Disease Models & Mechanisms* **7**, 823–835 (2014).
63. Li, J. et al. A transgenic zebrafish model for monitoring xbp1 splicing and endoplasmic reticulum stress in vivo. *Mechanisms of Development* **137**, 33–44 (2015).
64. Ryoo, H. D. Drosophila as a model for unfolded protein response research. *BMB Reports* **48**, 445–453 (2015).
65. Katow, H., Vasudevan, D. & Ryoo, H. D. *Drosophila Unfolded Protein Response (UPR) Assays In Vitro and In Vivo* (2022).
66. Demay, Y., Perochon, J., Szuplewski, S., Mignotte, B. & Gaumer, S. The perk pathway independently triggers apoptosis and a rac1/slpr/jnk/dilp8 signaling favoring tissue homeostasis in a chronic er stress drosophila model. *Cell Death & Disease* **5**, e1452–e1452 (2014).
67. Shen, X. *et al.* Complementary signaling pathways regulate the unfolded protein response and are required for c. elegans development. *Cell* **107**, 893–903 (2001).
68. Beaudoin-Chabot, C. *et al.* The unfolded protein response reverses the effects of glucose on lifespan in chemically-sterilized c. elegans. *Nature Communications* **13**, 5889 (2022).
69. Urano, F. et al. A survival pathway for caenorhabditis elegans with a blocked unfolded protein response. *The Journal of Cell Biology* **158**, 639–646 (2002).
70. Kimata, Y., Ishiwata-Kimata, Y., Yamada, S. & Kohno, K. Yeast unfolded protein response pathway regulates expression of genes for anti-oxidative stress and for cell surface proteins. *Genes to Cells* **11**, 59–69 (2006).
71. Nguyen, P. T. M., Ishiwata-Kimata, Y. & Kimata, Y. Fast-growing saccharomyces cerevisiae cells with a constitutive unfolded protein response and their potential for lipidic molecule production. *Applied and Environmental Microbiology* **88** (2022).
72. Ruberti, C. & Brandizzi, F. Unfolded protein response in arabidopsis. *Methods in molecular biology (Clifton, N.J.)* **1691**, 231–238 (2018).
73. Manghwar, H. & Li, J. Endoplasmic reticulum stress and unfolded protein response signaling in plants. *International journal of molecular sciences* **23** (2022).
74. Kamauchi, S., Nakatani, H., Nakano, C. & Urade, R. Gene expression in response to endoplasmic reticulum stress in arabidopsis thaliana. *The FEBS Journal* **272**, 3461–3476 (2005).
75. Alcântara, A., Seitner, D., Navarrete, F. & Djamei, A. A high-throughput screening method to identify proteins involved in unfolded protein response of the endoplasmic reticulum in plants. *Plant Methods* **16**, 4 (2020).
76. Vowels, J. J. & Thomas, J. H. Multiple chemosensory defects in daf-11 and daf-21 mutants of caenorhabditis elegans. *Genetics* **138**, 303–316. https://doi.org/10.1093/genetics/138.2.303 (1994).
77. Birnby, D. A. et al. A transmembrane guanylyl cyclase (daf-11) and hsp90 (daf-21) regulate a common set of chemosensory behaviors in caenorhabditis elegans. *Genetics* **155**, 85–104. https://doi.org/10.1093/genetics/155.1.85 (2000).
78. Rehman, S. U. et al. Genomic identification, evolution and sequence analysis of the heat-shock protein gene family in buffalo. *Genes* **11**, 1388 (2020).
79. Storey, J. M. & Storey, K. B. Chaperone proteins: universal roles in surviving environmental stress. *Cell Stress & Chaperones* **28**, 455–466 (2023).
80. Whitley, D., Goldberg, S. P. & Jordan, W. D. Heat shock proteins: a review of the molecular chaperones. *Journal of Vascular Surgery* **29**, 748–751 (1999).
81. Turan, M. Genome-wide analysis and characterization of HSP gene families (HSP20, HSP40, HSP60, HSP70, HSP90) in the yellow fever mosquito (Aedes aegypti) (Diptera: Culicidae). *Journal of Insect Science* **23**, 27. https://doi.org/10.1093/jisesa/iead114 (2023).
82. Hu, C. et al. Heat shock proteins: Biological functions, pathological roles, and therapeutic opportunities. *MedComm* **3**, e161 (2022).
83. Sun, T. et al. A bottom-up coarse-grained model for nucleosome-nucleosome interactions with explicit ions. *Journal of Chemical Theory and Computation* **18**, 3948–3960 (2022).
84. Liu, H.-Y. et al. Network analysis identifies an hsp90-central hub susceptible in ovarian cancer. *Clinical Cancer Research* **19**, 5053–5067 (2013).
85. Manikandan, P. et al. Exploring the biological behavior of heat shock proteins (hsps) for understanding the anti-ischemic stroke in humans. *Journal of Infection and Public Health* **15**, 379–388 (2022).
86. Park, S.-M., Kang, T.-I. & So, J.-S. Roles of xbp1s in transcriptional regulation of target genes. *Biomedicines* **9**, 791 (2021).
87. Wang, F.-M., Chen, Y.-J. & Ouyang, H.-J. Regulation of unfolded protein response modulator xbp1s by acetylation and deacetylation. *Biochemical Journal* **433**, 245–252 (2011).
88. Hillary, R. F. & FitzGerald, U. A lifetime of stress: Atf6 in development and homeostasis. *Journal of Biomedical Science* **25**, 48 (2018).
89. Tam, A. B. et al. The upr activator atf6 responds to proteotoxic and lipotoxic stress by distinct mechanisms. *Developmental Cell* **46**, 327-343.e7 (2018).
90. Gardner, B. M., Pincus, D., Gotthardt, K., Gallagher, C. M. & Walter, P. Endoplasmic reticulum stress sensing in the unfolded protein response. *Cold Spring Harbor Perspectives in Biology* **5**, a013169–a013169 (2013).
91. Nowakowska, M. et al. Profiling the expression of endoplasmic reticulum stress associated heat shock proteins in animal epilepsy models. *Neuroscience* **429**, 156–172 (2020).
92. Chen, Y. et al. Dihydroartemisinin-induced unfolded protein response feedback attenuates ferroptosis via perk/atf4/hspa5 pathway in glioma cells. *Journal of Experimental & Clinical Cancer Research* **38**, 402 (2019).
93. Chen, X., Shi, C., He, M., Xiong, S. & Xia, X. Endoplasmic reticulum stress: molecular mechanism and therapeutic targets. *Signal Transduction and Targeted Therapy* **8**, 352 (2023).
94. Cox, J. S., Shamu, C. E. & Walter, P. Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase. *Cell* **73**, 1197–1206 (1993).
95. Mori, K., Ma, W., Gething, M.-J. & Sambrook, J. A transmembrane protein with a cdc2+/cdc28-related kinase activity is required for signaling from the er to the nucleus. *Cell* **74**, 743–756 (1993).
96. Riaz, T. A. et al. Role of endoplasmic reticulum stress sensor ire1α in cellular physiology, calcium, ros signaling, and metaflammation. *Cells* **9**, 1160 (2020).
97. Trentmann, S. M. Ern1, a novel ethylene-regulated nuclear protein of arabidopsis. *Plant Molecular Biology* **44**, 11–25 (2000).
98. Angelos, E. & Brandizzi, F. The upr regulator ire1 promotes balanced organ development by restricting tor-dependent control of cellular differentiation in arabidopsis. *The Plant Journal* **109**, 1229–1248 (2022).
99. Ruberti, C., Kim, S.-J., Stefano, G. & Brandizzi, F. Unfolded protein response in plants: one master, many questions. *Current Opinion in Plant Biology* **27**, 59–66 (2015).
100. Meyer, H., Bug, M. & Bremer, S. Emerging functions of the vcp/p97 aaa-atpase in the ubiquitin system. *Nature Cell Biology* **14**, 117–123. https://doi.org/10.1038/ncb2407 (2012).
101. Christianson, J. C. & Ye, Y. Cleaning up in the endoplasmic reticulum: ubiquitin in charge. *Nature Structural & Molecular Biology* **21**, 325–335 (2014).
102. Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the internet to random breakdowns. *Physical review letters* **85**, 4626 (2000).
103. Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Breakdown of the internet under intentional attack. *Physical review letters* **86**, 3682 (2001).

104. Holme, P., Kim, B. J., Yoon, C. N. & Han, S. K. Attack vulnerability of complex networks. *Physical review E* **65**, 056109 (2002).
105. Ruj, S. & Pal, A. Analyzing cascading failures in smart grids under random and targeted attacks (2014). arXiv:1402.6809.
106. Callaway, D. S., Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.* **85**, 5468–5471 (2000).
107. Jeronimo, C. et al. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7sk capping enzyme. *Molecular Cell* **27**, 262–274 (2007).
108. Imai, J., Yashiroda, H., Maruya, M., Yahara, I. & Tanaka, K. Proteasomes and molecular chaperones: cellular machinery responsible for folding and destruction of unfolded proteins. *Cell Cycle* **2**, 584–588 (2003).
109. Larburu, N., Adams, C. J., Chen, C.-S., Nowak, P. R. & Ali, M. M. U. Mechanism of hsp70 specialized interactions in protein translocation and the unfolded protein response. *Open Biology* **10**, 200089 (2020) (**Epub 2020 Aug 19**).
110. Runtuwene, L. R. et al. The lethal (2)-essential-for-life [l (2) efl] gene family modulates dengue virus infection in aedes aegypti. *International journal of molecular sciences* **21**, 7520 (2020).
111. Tower, J. Heat shock proteins and drosophila aging. *Experimental gerontology* **46**, 355–362 (2011).
112. Morrow, G., Le Pécheur, M. & Tanguay, R. M. Drosophila melanogaster mitochondrial hsp22: a role in resistance to oxidative stress, aging and the mitochondrial unfolding protein response. *Biogerontology* **17**, 61–70 (2016).
113. Yu, G. & Hyun, S. Proteostasis-associated aging: lessons from a drosophila model. *Genes & genomics* **43**, 1–9 (2021).
114. Bunney, T. D. et al. Crystal structure of the human, fic-domain containing protein hype and implications for its functions. *Structure* **22**, 1831–1843 (2014).
115. Sanyal, A. et al. A novel link between fic (filamentation induced by camp)-mediated adenylation/ampylation and the unfolded protein response. *Journal of Biological Chemistry* **290**, 8482–8499 (2015).
116. Li, J. et al. Ufm1-specific ligase ufl1 regulates endoplasmic reticulum homeostasis and protects against heart failure. *Circ Heart Fail* **11**, e004917 (2018).
117. Jiang, Q. et al. Ufl1, a ufmylation e3 ligase, plays a crucial role in multiple cellular stress responses. *Frontiers in Endocrinology* **14**, 1123124 (2023).
118. Evtushenko, N. A., Beilin, A. K., Kosykh, A. V., Vorotelyak, E. A. & Gurskaya, N. G. Keratins as an inflammation trigger point in epidermolysis bullosa simplex. *International Journal of Molecular Sciences* **22**, 12446 (2021).
119. Evstafieva, A. G., Kovaleva, I. E., Shoshinova, M. S., Budanov, A. V. & Chumakov, P. M. Implication of krt16, fam129a and hkdc1 genes as atf4 regulated components of the integrated stress response. *PloS one* **13**, e0191107 (2018).
120. Viswanathan, M. *et al.* A role for sir-2.1 regulation of er stress response genes in determining c. elegans life span. *Developmental Cell* **9**, 605–615 (2005).
121. Tran, H. C. & Aken, O. V. Mitochondrial unfolded protein-related responses across kingdoms: similar problems, different regulators. *Mitochondrion* **53**, 166–177 (2020).
122. Walter, P. & Ron, D. The unfolded protein response: From stress pathway to homeostatic regulation. *Science* **334**, 1081–1086 (2011).
123. Almanza, A. et al. Endoplasmic reticulum stress signaling - from basic mechanisms to clinical applications. *The FEBS Journal* **286**, 241–278 (2019).
124. Vidal, R. L. & Hetz, C. Crosstalk between the upr and autophagy pathway contributes to handling cellular stress in neurodegenerative disease. *Autophagy* **8**, 970–972 (2012).
125. Mori, K. The unfolded protein response: The dawn of a new field. *Proceedings of the Japan Academy, Series B, Physical and Biological Sciences* **91**, 469–480. https://doi.org/10.2183/pjab.91.469 (2015).
126. Ma, Y. & Hendershot, L. M. The unfolding tale of the unfolded protein response. *Cell* **107**, 827–830. https://doi.org/10.1016/s0092-8674(01)00623-7 (2001).
127. Zhang, L., Zhang, C. & Wang, A. Divergence and conservation of the major upr branch ire1-bzip signaling pathway across eukaryotes. *Scientific Reports* **6**, 27362. https://doi.org/10.1038/srep27362 (2016).
128. Gómora-García, J. C., Gerónimo-Olvera, C., Pérez-Martínez, X. & Massieu, L. Ire1α ridd activity induced under er stress drives neuronal death by the degradation of 14-3-3 θ mrna in cortical neurons during glucose deprivation. *Cell Death Discovery* **7**, 131 (2021).
129. Maurel, M., Chevet, E., Tavernier, J. & Gerlo, S. Getting ridd of rna: Ire1 in cell fate regulation. *Trends in biochemical sciences* **39**, 245–254 (2014).
130. Hollien, J. et al. Regulated ire1-dependent decay of messenger rnas in mammalian cells. *Journal of Cell Biology* **186**, 323–331 (2009).
131. Zheng, Z. et al. Transcriptional signatures of unfolded protein response implicate the limitation of animal models in pathophysiological studies. *Environmental disease* **1**, 24–30 (2016).
132. Breschi, A., Gingeras, T. R. & Guigó, R. Comparative transcriptomics in human and mouse. *Nature Reviews Genetics* **18**, 425–440 (2017).
133. Cheng, Y. et al. Principles of regulatory information conservation between mouse and human. *Nature* **515**, 371–375 (2014).
134. Lin, S. et al. Comparison of the transcriptional landscapes between human and mouse tissues. *Proceedings of the National Academy of Sciences* **111**, 17224–17229 (2014).
135. Zhu, F., Nair, R. R., Fisher, E. M. & Cunningham, T. J. Humanising the mouse genome piece by piece. *Nature communications* **10**, 1845 (2019).
136. Monaco, G., van Dam, S., Casal Novo Ribeiro, J. L., Larbi, A. & de Magalhães, J. P. A comparison of human and mouse gene co-expression networks reveals conservation and divergence at the tissue, pathway, and disease levels. *BMC evolutionary biology* **15**, 1–14 (2015).
137. Klein, B. et al. A computational exploration of resilience and evolvability of protein-protein interaction networks. *Communications Biology* **4**, 1352 (2021).
138. Ortiz-Vilchis, P., la Cruz-García, J.-S.D. & Ramirez-Arellano, A. Identification of relevant protein interactions with partial knowledge: A complex network and deep learning approach. *Biology* **12**, 140 (2023).
139. Wagner, A. Robustness and evolvability: a paradox resolved. *Proceedings of the Royal Society B: Biological Sciences* **275**, 91–100 (2008).
140. Zitnik, M., Sosič, R., Feldman, M. W. & Leskovec, J. Evolution of resilience in protein interactomes across the tree of life. *Proceedings of the National Academy of Sciences* **116**, 4426–4433 (2019).
141. Kalman, R. E. Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics Series A Control* **1**, 152–192 (1963).
142. *Minimum driver nodes selection in complex networks* (IEEE, 2017).
143. Varoquaux, G., Vaught, T. & Millman, J. (eds). *Exploring network structure, dynamics, and function using NetworkX* (2008).
144. Yuan, Z., Zhao, C., Di, Z., Wang, W.-X. & Lai, Y.-C. Exact controllability of complex networks. *Nature Communications* **4**, 2447 (2013).
145. Patel, T. P., Man, K., Firestein, B. L. & Meaney, D. F. Automated quantification of neuronal networks and single-cell calcium dynamics using calcium imaging. *Journal of Neuroscience Methods* **243**, 26–38 (2015).
146. Bateman, A. et al. Uniprot: the universal protein knowledgebase in 2023. *Nucleic Acids Research* **51**, D523–D531 (2023).
147. Consortium, T. U,. Uniprot: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**, D506–D515 (2019).
148. Wang, Y. et al. A crowdsourcing open platform for literature curation in uniprot. *PLOS Biology* **19**, e3001464 (2021).

149. v. Mering, C. String: a database of predicted functional associations between proteins. *Nucleic Acids Research* **31**, 258–261 (2003).
150. Szklarczyk, D. et al. The string database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research* **51**, D638–D646 (2023).
151. Luchetti, N., Loppini, A., Matarrese, M. A., Chiodo, L. & Filippi, S. Structural controllability to unveil hidden regulation mechanisms in unfolded protein response: The role of network models. *Physica A: Statistical Mechanics and its Applications* **617**, 128671 (2023).
152. *On estimating the average degree.*
153. Chen, H.-H. & Dietrich, U. Normalized closeness centrality of urban networks: impact of the location of the catchment area and evaluation based on an idealized network. *Applied Network Science* **8**, 60 (2023).
154. Rossum, G. V. & Drake, F. L. *Python 3 Reference Manual* (CreateSpace, 2009).
155. Inc., T. M. Matlab version: 9.14.0.2254940 (r2023a) update 2 (2023).
156. Bretz, F. et al. Graphical approaches for multiple comparison procedures using weighted bonferroni, simes, or parametric tests. *Biometrical journal. Biometrische Zeitschrift* **53**, 894–913 (2011).
157. Armstrong, R. A. When to use the bonferroni correction. *Ophthalmic & physiological optics: the journal of the British College of Ophthalmic Opticians (Optometrists)* **34**, 502–508 (2014).
158. Iyer, S., Killingback, T., Sundaram, B. & Wang, Z. Attack robustness and centrality of complex networks. *PloS one* **8**, e59613 (2013).
159. Kang, K. Analysis of evolutionary process of fog computing system based on ba and er network hybrid model. *Evolutionary Intelligence* **13**, 33–38 (2020).

Acknowledgements

Author contributions

Conceptualization: N.L., K.M.S., S.F. and L.C. Methodology: N.L., K.M.S., M.A.G.M., A.L. and L.C. Formal analysis: N.L. Investigation: N.L. Writing − Original draft: N.L., M.A.G.M., and L.C. Writing − Review and editing: all authors Visualization: N.L. and M.A.G.M. Supervision: K.M.S., S.F., and L.C.

Declaration

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-79086-8.

Correspondence and requests for materials should be addressed to N.L. or S.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.αα