Intrinsic disorder codes for leaps of protein expression

Chi-Ning Chuang[1], Tai-Ting Woo[1], Shih-Ying Tsai[2], Wan-Chen Li[1], Chia-Ling Chen[1], Hou-Cheng Liu[1], Chiung-Ya Chen[1], Yi-Ping Hsueh[1], Ting-Fang Wang[1,*]

[1]Institute of Molecular Biology, Academia Sinica, Taipei 115, Taiwan

[2]Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan

*Corresponding author

E-mail: tfwang@gate.sinica.edu.tw

Running title: N-terminal IDR promotes protein expression

The SI Appendix includes six supplemental tables, three supplemental figures, and two datasets in Excel format.

Authors' email addresses

Chi-Ning Chuang email: chining@gate.sinica.edu.tw

Tai-Ting Woo email: asanwoo@gmail.com

Shih-Ying Tsai email: wiccesibyl@gmail.com

Chia-Ling Chen email: chialing.chen1118@gmail.com

Wan-Chen Li email: wanwan9121@gmail.com

Hou-Cheng Liu email: hc666@gate.sinica.edu.tw

Chiung-Ya Chen email: chiungya@gate.sinica.edu.tw

Yi-Ping Hsueh email: yph@gate.sinica.edu.tw

Ting-Fang Wang email: tfwang@gate.sinica.edu.tw

Abstract

Intrinsically disordered regions (IDRs) are protein sequences lacking fixed or ordered three-dimensional structures. Many IDRs are endowed with important molecular functions such as physical interactions, posttranslational modifications or solubility enhancement. We reveal that several biologically important IDRs can act as N-terminal fusion carriers to promote target protein folding or protein quality control, thereby enhancing protein expression. This nanny function has a reasonably strong correlation with high S/T/Q/N amino acid content in IDRs and it is tunable (e.g., via phosphorylation) to regulate protein homeostasis. We propose a hypothesis that "N-terminal intrinsic disorder facilitates abundance" (NIDFA) to explain how some yeast proteins use their N-terminal IDRs (N-IDRs) to generate high levels of protein product. These N-IDRs are versatile toolkits for functional divergence in signaling and evolution.

Significance

Disorder within an otherwise well-structured protein is mostly found in intrinsically disordered regions (IDRs). IDRs can provide many advantages to proteins, including: (1) mediating protein-protein or protein-peptide interactions by adopting different conformations; (2) facilitating protein regulation via diverse posttranslational modifications; and (3) regulating the half-lives of proteins that have been targeted for proteasomal degradation. Here, we report that several biologically important IDRs in *S. cerevisiae* can act as N-terminal fusion carriers to promote target protein folding or protein quality control, thereby enhancing protein expression. We demonstrate by genetic and bioinformatic analyses that this nanny function is well correlated with high content of serine, threonine, glutamine and asparagine in IDRs and is tunable (e.g., via phosphorylation) to regulate protein homeostasis.

Introduction

Intrinsically disordered regions (IDRs) are protein sequences without stable secondary and/or tertiary structure (1). Several studies have revealed that high content of Ser (S), Thr (T), Gln (Q), Asn (N), Pro, Gly or charged amino acids [Arg (R), Lys (K) and His (H)] is a common feature of many IDRs (2-4). Functionally, IDRs are key components of subcellular machineries and signaling pathways because they have the potential to associate with many partners due to their multiple possible metastable conformations. Many IDRs are regulated by alternative splicing and post-translational modifications. Some IDRs are involved in the formation of various membraneless organelles via intracellular liquid-liquid phase separation (5, 6). It has also been reported that highly-charged IDRs can act as entropic bristles that, when translationally fused to their partner proteins, only enhanced water solubility but not the steady-state levels of the partner proteins (7).

Intriguingly, intrinsic disorder gain or loss via gene duplication or deletion, respectively, is a mechanism by which a genome can diverge and innovate (8, 9). For example, the genome of the budding yeast *Saccharomyces cerevisiae* has undergone whole genome duplication (WGD) relative to the pre-WGD outgroup species *Saccharomyces kluyveri* (also known as *Lachancea kluyveri*) (10). WGD results in a pair of paralogous genes called ohnologs. Subsequently, it was reported that there is a preference for retaining IDRs during gene duplication, i.e., there is more intrinsic disorder in *S. cerevisiae* duplicates than in their *S. kluyveri* orthologs or singletons. For many *S. cerevisiae* duplicated gene pairs, the number of physical interactions is significantly greater for the more intrinsically disordered ohnolog (8). Accordingly, "IDRs" have been evolutionarily retained as toolkits to create complexity in regulatory networks and they provide multiple mechanisms for cell-type or tissue-specific signaling, e.g., solubility, physical interactions, post-translational modifications or liquid-liquid phase transition.

It is important to note that the abundance of proteins with IDRs in cells is considered to be erratic because improperly folded proteins are often proteolytically degraded or otherwise form abnormal aggregates that can give rise to disease-causing prions. Two interesting hypothetical models have been proposed for this important biological scenario. Firstly, as described above, the IDRs in some proteins can adopt a specific conformation upon the protein binding to a partner protein (5, 11-13), with the potential for a given protein to adopt distinct conformations depending on the involvement of different binding partners, chaperones (to support protein folding or degradation) or post-translational modifications. Secondly, the "N-terminal folding nucleation" (NFN) hypothesis illustrates that some proteins use their structured N-terminal domains (SNTDs) to organize the remaining protein chain by means of intramolecular interactions (14). In brief, to reduce the risk of protein degradation or aggregation, the SNTD folds spontaneously during protein translation and then serves as a nucleation point for the as yet unstructured amino acid chain. Then, the SNTD-docked regions and free loops create more compact shapes not only to coordinate assembly of sub-complexes in defined loop-sections, but also to enable novel regulatory mechanisms such as via posttranslational modifications of docked regions (14). Although both these hypotheses are applicable to many IDR proteins, neither can account for the functions of proteins that harbor an N-terminal IDR (N-IDR) but lack binding partners or that fold prior to protein-protein interaction.

In this study, we report that several IDRs with high STQN amino acid contents exhibit autonomous expression-enhancing activity for high-level production of native protein and when fused to exogenous target proteins, e.g., β-galactosidase (LacZ) *in vivo*. Interestingly, most (if not all) IDRs we describe herein possess one or more ATR^{Mec1}/ATM^{Tel1} DNA damage checkpoint kinase phosphorylation sites, suggesting their functions and/or stability may be tunable by ATR^{Mec1} or ATM^{Tel1} via protein phosphorylation. We also reveal structural and

genetic requirements for the nanny function of N-IDRs in regulating protein homeostasis. Taken together, we propose another mechanism by which N-IDRs might facilitate high-levels of protein expression *in vivo*.

Results

S/T-Q cluster domains (SCDs) and prion domains of prion-causing proteins are IDRs with high STQN amino acid contents

SCDs are concentrated ATM/ATR kinase phosphorylation sites in DNA damage response (DDR) proteins. $Mec1^{ATR}$ and $Tel1^{ATM}$ are yeast orthologs of mammalian DNA damage sensor kinases ATM (ataxia-telangiectasia mutated) and ATR (RAD3-related), respectively (15). They preferentially phosphorylate S/T-Q motifs, i.e., S and T that are followed by Q (16). A unifying definition of an SCD is having ≥3 S/T-Q sites within a stretch of 50 or 100 amino acids in *S. cerevisiae* (17) or mammals (18), respectively. One of the best understood mechanisms of SCD phosphorylation involves the association of SCDs with their binding partners containing a forkhead-associated (FHA) domain. For example, $Mec1^{ATR}$– and $Tel1^{ATM}$–dependent phosphorylation of Rad53-SCD1 (residues 1-29) and Hop1-SCD (residues 258-324) [*Supplemental Information (SI) Appendix*, Table S1] specifically recruits and activates their downstream kinases Dun1 and Mek1, respectively (19-21).

$Mec1^{ATR}$ and $Tel1^{ATM}$ also exhibit an essential function in regulating protein homeostasis (22, 23). In agreement with this function, we reported recently that the N-terminal domain (NTD; residues 1–66) of Rad51 contains three SQ motifs (S^2Q, $S^{12}Q$ and $S^{30}Q$), the phosphorylation of which is dependent on $Mec1^{ATR}$ and $Tel1^{ATM}$ protein kinases (24). This clustering of three SQ motifs within a stretch of 30 amino acids in Rad51-NTD fulfills the criteria to define an SCD. Intriguingly, Rad51-NTD has dual functions for repairing DSBs.

Firstly, Rad51-NTD autonomously promotes high-level production of native Rad51 and its COOH-terminal fusion protein LacZ *in vivo*. Secondly, three SQ motifs of Rad51-NTD are phosphorylated in a Mec1^{ATR}– and Tel1^{ATM}–dependent manner during vegetative growth and meiosis. Mec1/Tel1–dependent phosphorylation antagonizes Rad51 degradation via the proteasomal pathway, increasing the half-life of Rad51 from ~30 min to ≥180 min (24). It is noteworthy that the nanny function of Rad51-NTD in promoting high-level protein production can occur regardless of the presence or absence of Mec1^{ATR}– or Tel1^{ATM}–dependent phosphorylation (24). However, Mec1^{ATR}– or Tel1^{ATM}–dependent phosphorylation may function post-translationally to further stabilize Rad51 or Rad51-NTD fusion proteins by preventing proteasomal degradation. Further investigations are needed to reveal the molecular mechanism underlying this nanny function.

A hallmark of DDR-SCDs is their high STQN amino acid content, suggesting that DDR-SCDs are intrinsically disordered regions (IDRs) in proteins (2-4). We also noticed that the prion (nucleation) domains of three yeast prion-causing proteins (Sup35, Ure2 and New1) share two common structural features with DDR-SCDs. They not only have high S/T/Q/N amino acid contents, but also contain at least one S/T-Q site that might be phosphorylated by Mec1^{ATR} or Tel1^{ATM} (*SI Appendix*, Table S1). The prion-forming domain (PFD; residues 1-114) of Sup35 drives the structural conversion to amyloid and the prion nucleation domain (PND; residues 1-39) of Sup35 has critical functions in promoting [*PSI*+] prion nucleation (25). The STQN content of Sup35-PFD is ≥49% and for Sup35-PND it is ≥61% (*SI Appendix*, Table S1). Although Sup35-PND is not ascribed as an SCD, it does contain an S^{17}Q motif that can be phosphorylated during DDR (see below). Similarly, the Ure2 prion domain (UPD) (residues 1-91) of the Ure2 nitrogen catabolite repression transcriptional regulator is the basis of the prion [*URE3*$^{+}$] (26, 27). UPD is critical for Ure2's *in vivo* function because removal of UPD from Ure2 results in reduced protein stability and steady-state protein levels (but not transcript

levels) of the corresponding Ure2-ΔUPD mutants (28). Like Sup35-PND, Ure2-UPD has a high STQN content (>63%; 10 serines, 5 threonines, 10 glutamines, 33 asparagines in 91 amino acids) and adopts a completely disordered structure (29). Moreover, Ure2-UPD also possesses an $S^{68}Q$ motif. Finally, the New1 prion domain (NPD) supports [NU^{+}] and is susceptible to [*PSI*+] prion induction (30, 31). New1 is a non-essential ATP-binding cassette type F protein that fine-tunes the efficiency of translation termination or ribosome recycling (32). New1-NPD also has a high STQN content (>44%; 19 serines, 8 threonines, 14 glutamines, 26 asparagines in 156 amino acids) and an $S^{145}Q$ motif.

Four DDR-SCDs and three prion domains can all promote target protein expression

As reported recently (24), removal of the NTD from Rad51 reduced by ~97% the protein levels of corresponding Rad51-ΔN proteins relative to wild-type (Fig. 1*A*), leading to yeast hypersensitivity to the DNA damage agent methyl methanesulfonate (MMS) (Fig. 1*B*). This nanny function of Rad51-NTD could be rescued in Rad51-ΔN by N-terminal fusion of Rad53-SCD1, Rad53-SCD1-5STA (all five S/T-Q motifs changed to AQs) or Sup35-PND (Fig. 1 and *SI Appendix*, Table S1), respectively.

Next, we expressed different X-LacZ-NVH fusion proteins using a *CEN-ARS* plasmid (low-copy number) under the control of the native *RAD51* gene promoter (P_{RAD51}), including Rad51-NTD, Rad53-SCD1, Hop1-SCD (residues 258-324), Sml1-NTD (residues 1-27), Sml1-NTD (residues 1-50), Sup35-PFD, Sup35-PND, Sup35-PND-$S^{17}A$, Ure2-UPD and New1-NPD (Fig. 1*C*, and *SI Appendix*, Table S1 and Table S2). The NVH tag contains an SV40 nuclear localization signal (NLS) peptide preceding a V5 epitope tag and a hexahistidine (His_6) affinity tag (24). The $S^{17}Q$ motif was changed to AQ in Sup35-PND-$S^{17}A$. Sml1 is a potent inhibitor of ribonucleotide reductase and, in the SK1 *S. cerevisiae* strain, it harbors three S/T-Q motifs ($S^{4}Q$, $S^{14}Q$ and $T^{47}Q$). We found that the β-galactosidase activities of all X-LacZ-

NVH fusion proteins were 4.1-11.6-fold higher than that of LacZ-NVH (Fig. 1*C*). Interestingly, using anti-phosphorylated Rad51-S^{12}Q, Rad51-S^{30}Q, Hop1-T^{318}Q, and Sup35-S^{17}Q antisera, respectively, we found that Rad51-NTD-LacZ-NVH, Hop1-SCD-LacZ-NVH (*SI Appendix*, Table S2 and Fig. S1*A*), and Sup35-PND-LacZ-NVH, but not Sup35-PND-S^{17}A-LacZ-NVH (*SI Appendix*, Fig. S1*C*), were phosphorylated in response to MMS treatment. As described recently (24), the nanny function of these IDRs can occur regardless of the presence or absence of an MMS-induced DNA damage response.

It is also noteworthy that the nanny function of DDR–SCDs and prion–IDRs are not likely to be relevant to the N-end rule, which links the *in vivo* half-life of a protein to the identity of its N-terminal residues. In *S. cerevisiae*, the N-end rule pathway operates as part of the ubiquitin system and comprises two pathways. Firstly, the Arg/N-end rule pathway, involving a single N-terminal amidohydrolase Nta1, mediates deamidation of N-terminal Asn (N) and Gln (Q) into Asp (D) and Glu (E), which in turn are arginylated by a single Ate1 R-transferase, generating the Arg/N-degron. N-terminal R and other primary degrons are recognized by a single N-recognin Ubr1 in concert with the Ubiquitin-conjugating Ubc2/Rad6. Ubr1 can also recognize several other N-terminal residues, including Lys (K), His (H), Phe (F), Trp (W), Leu (L) and Ile (I) (33-35). Secondly, the Ac/N-end rule pathway targets proteins containing N-terminally acetylated (Ac) residues. Prior to acetylation, Met-aminopeptides (MetAPs) catalyze removal of the first amino acid Met (M) unless a residue at position 2 is non-permissive (too large) for MetAPs. If the (retained) N-terminal M or N-terminal Val (V), Cys (C), Ala (A), Ser (S) and Thr (T) are followed by residues that allow N-terminal acetylation (Ac), proteins containing ^{Ac}N-degrons are targeted for ubiquitylation and proteasome-mediated degradation by the Doa10 E3 ligase (36). Since the first two amino acid residues of Rad51-NTD, Hop1-SCD, Rad53-SCD1, Sup35-PND, Rad51-ΔN and LacZ-NVH are MS, ME, ME, MS, ME and MI, respectively, and Sml1-NTD behaved similarly to these N-terminal fusion

tags despite its MQ amino acid signature in the N-terminus, we infer that the protein expression-enhancing activities of these IDRs are not likely to arise from counteracting the N-end rule.

Intrinsically disordered N-termini harboring S/T-Q sites that have arisen from WGD promote LacZ expression and activity

In light of our finding that the unique Rad51-NTD in yeast exhibits an expression-enhancing capability, we sought to investigate if it is a common scenario that IDRs that have evolutionarily accumulated S/T-Q motifs at their N-termini behave similarly in promoting protein production. In *S. cerevisiae*, the protein abundance (median molecule number per cell) of 487 duplicated protein (DP) pairs has been determined (37), and 296 of those DPs contain N-terminal sequences different from those of their ohnologs. Among them, distinct occurrence of N-IDRs within the DP pair may implicate their differential gain or loss after WGD, which could be interpreted by analyzing for the presence or absence of the respective IDR in their *L. kluyveri* orthologs (8). We speculate that gain or loss of an N-IDR harboring S/T-Q motifs after WGD may alter DP expression levels. We tested this notion directly in our LacZ-NVH system by examining the protein expression-enhancing capability of the N-IDRs of three DDR-unrelated proteins, Vps64, Ssk2 and Kel1 (described below in Discussion), that have arisen from WGD and all contain at least one SQ motif in their unique N-IDRs (Fig. 1*C* and Dataset S1, Table DS1). The N-IDRs of Vps64, Ssk2 and Kel1 that we used for N-terminal fusion to LacZ-NVH are 152, 191 and 69 amino acids in length and contain three, three and one S/T-Q motif(s), respectively. As expected, all three N-IDR-LacZ-NVH exhibited higher LacZ activities (6.3-10.6-fold) relative to that of LacZ-NVH alone (Fig. 1*C*), agreeing well with previous proteome-wide studies that have revealed higher median protein abundance of these

three duplicated proteins (Vps64, Ssk2 and Kel1) compared to those of their ohnolog proteins in vegetatively growing cells (Dataset S1, Tables DS1) (37-39).

Proteome-wide analysis reveals an obscure propensity for N-IDRs to promote expression of duplicated proteins

Among the 296 DP pairs whose N-terminal sequences vary between ohnologs, 152 DPs contain longer N-termini that are predicted by IUPred (40, 41) to be intrinsically disordered and are thus referred to as N-IDR$^+$ DPs (Dataset S1, Tables DS1). Based on the assumption that a functional protein fragment should reach a certain length threshold to acquire its physiological properties, we deemed putative IDRs of ≤10 amino acids in length to be non-disordered, as suggested by a previous study that analyzed unique IDRs that have arisen from WGD (8). From a consolidated analysis in which we curated the datasets of protein abundance per cell (37) and our list of N-IDR$^+$ DPs (Dataset S1, Tables DS1), we determined the effect of N-IDRs on protein level control by comparing the median protein abundance per cell (i.e., number of molecules per vegetative cell) of the DPs containing longer N-fragments divided by those of their respective ohnologs. We found an interesting trend approaching conventional statistical significance (P = 0.058) in that the presence of at least two S/T-Q motifs in the N-IDRs (N-IDR$^+$-SQ/TQ$^+$ DPs, n = 24) is likely correlated with higher protein abundance relative to corresponding ohnologs when they are compared to DPs with longer N-fragments that are not predicted as IDRs (N-IDR$^-$ DPs, n = 144), but not to N-IDR$^+$ DPs that harbor no or only one S/T-Q motif in their N-IDRs (N-IDR$^+$-SQ/TQ$^-$ DPs, n = 128) (Fig. 2).

It is noteworthy that the altered abundance we determined using the ratio of protein abundance of long-NH_2 DPs over that of their ohnologs was from large-scale data-mining, but it may not truly reflect the scenario of all DP pairs. For example, Sml1 has two S/T-Q motifs (in S288c) in the N-IDR and its protein abundance is higher (~9-fold) than that of its ohnolog,

Dif1, but the N-IDR of Sml1 is shorter. Therefore, this DP pair (Sml1/Dif1) falls into the latter category so that the median protein abundance of Dif1 (an N-IDR^{+} DP that does not contain S/T-Q motifs) divided by that of Sml1 is used in our analysis (Dataset S1, Tables DS1).

Taken together, we propose that gain or loss of an N-IDR after WGD often alters DP expression levels as suggested by our LacZ-NVH assay, but the effects could be buffered by co-evolution of their regulation at the transcriptional level or protein turnover control as shown in the proteome-wide study. This protein expression-enhancement effect would be more apparent only when proteins with N-IDRs harboring at least two S/T-Q motifs are compared with those without N-IDRs. This observation might be relevant to the previous finding showing that Mec1^{ATR}– and Tel1^{ATM}–dependent phosphorylation antagonizes proteolytic degradation (24).

Quaternary structures and intermolecular phase transition

We found that N-terminal fusion of Rad53-SCD1 to four different NVH-tagged target proteins all resulted in higher production (Fig. 3); LacZ is a tetrameric protein, GST is dimeric, and non-dimerizing GST (GSTnd) and GFP are monomeric proteins. There is growing interest in the topic of intermolecular phase transition in IDRs (6), but our results imply that the expression-enhancing function of IDRs might not be mediated by a protein's phase behavior. Firstly, the β-galactosidase activities of Rad51-NTD-LacZ-NVH were higher than those of LacZ-NVH in both wild-type and *hsp104*Δ vegetative cells (Fig. 4), with the Hsp104 disaggregase that could break polymeric Sup35 into oligomeric "seeds" being required for propagation of [*PSI*+] prion (26, 42). Secondly, both the β-galactosidase activities and steady-state levels of Sup35-PND-LacZ-NVH proteins were not significantly affected by 1,6-hexanediol (1,6-HD) treatment (*SI Appendix*, Fig. S2). 1,6-HD is a mild apolar solvent that has been used extensively to modify phase separation *in vitro* and *in vivo* (6). Taken together, our

results suggest that target protein quaternary structures and intermolecular phase transition are unlikely relevant to IDR's nanny function.

High STQN content is critical for IDR's nanny functions

A common feature of DDR-SCDs, Sup35-PND, Ure2-UPD and New1-NPD is their high STQN amino acid contents (Fig. 5*A*). We wondered if the high STQN content of IDRs is critical for promoting high-level protein expression. We employed an alanine-scanning mutagenesis approach to test this hypothesis. Alanine scanning by definition increases the relative content of a given sequence in alanine residues that are known to favor the formation of alpha helices (43) and hence to favor a transition from disorder to order.

Our results reveal a positive correlation between relative β-galactosidase activities or LacZ-NVH fusion protein levels and overall STQN amino acid percentages in IDRs (Fig. 5 and *SI Appendix*, Fig. S3). For each individual IDR that we examined, we observed a linear relationship ($R^2 = 0.86 – 0.94$) between its STQN amino acid content and its LacZ activity (Fig. 5 *B-D*). Although the overall lengths of these IDRs differ, our results indicate that in order to attain optimal LacZ activity of greater than 5-fold relative to wild-type, the STQN amino acid content of IDRs should be at least ≥30%.

It is important to note that the threshold of STQN content varies in the three cases presented here (Fig. 5 *B-D*), indicating that the fraction of STQN residues in IDRs is not the only factor contributing to protein expression level. Thus, we carried out more in-depth analysis of the amino acid compositions of all the N-IDRs we analyzed here. However, we did not find any other amino acid residues as critical as STQN in terms of promoting protein expression.

The IDRs' nanny function is likely controlled after protein translation

The nanny function of Rad51-NTD in promoting high protein production (Fig. 6*A*) is not likely to operate at the transcriptional level, as revealed by genomic and reverse-transcription quantitative polymerase chain reactions (i.e., g-qPCR or RT-qPCR). We found that the addition of WT and mutant Rad51-NTD to LacZ-NVH not only did not affect the average copy number of the corresponding *CEN-ARS* plasmids in exponentially growing *S. cerevisiae* cells (Fig. 6*B*), but also even decreased the steady-state transcript levels of the corresponding LacZ-NVH fusion protein genes (Fig. 6*C*). Therefore, addition of Rad51-NTD to LacZ-NVH did not result in a significant increase in plasmid copy number or its transcription.

The nanny function of N-IDRs is unlikely to be employed during protein translation. The translation initiation and elongation rates for 1106 *S. cerevisiae* open reading frames (ORFs) in exponentially growing wild-type yeast cells were reported recently by measuring protein synthesis rates in conjunction with ribosome footprinting (44). Here, we applied IUPred (https://iupred.elte.hu/) (40, 41, 45) to determine if the N-fragments of these ORFs harbor N-IDRs. We did not find evidence of either a higher "overall protein synthesis rate" or greater "protein-synthesis rate per mRNA" associated with N-IDRs (*SI Appendix*, Table S3).

Next, we applied “FOLD-RATE” (https://www.iitm.ac.in/bioinfo/fold-rate/) (46), a statistical method based on multiple regression, to predict protein folding rates using amino acid composition and properties. This method clusters proteins based on their secondary structure (α-helices only, β-sheets or mixed), and the correlations between the predicted log folding rate and the actual log folding rate ranging from 0.90 and to 0.99 (46). Using the “mixed” mode of “FOLD-RATE”, we found that N-terminal fusion of most (if not all) IDRs described in this study (Fig. 1 and Fig. 6) increases the predicted protein folding rates of LacZ-NVH, GFP-NVH, GST-NVH, GST(nd)-NVH and Rad51-ΔN (*SI Appendix*, Table S4). Our results also reveal that the enhanced levels of protein expression do not always increase proportionally with the predicted protein folding rates. One possibility for that outcome is that the predictive

power of "FOLD-RATE" may be overstated due to large fluctuations and overfitting of the estimates (47).

To further investigate if the STQN amino acid contents of N-IDRs play critical roles in promoting protein folding, we also applied "FOLD-RATE" and "IUPred" to analyze all (5917) *S. cerevisiae* ORFs in the *Saccharomyces* Genome Database (SGD; https://www.yeastgenome.org). There are 3817 ORFs with an ordered NTD, 963 ORFs having an N-IDR with <30% STQN amino acid content, and 1137 ORFs having an N-IDR with ≥30% STQN amino acid content (Dataset S2, Tables DS2-DS4). Interestingly, we found statistically significant trends that N-IDRs with ≥30% STQN amino acid content outperformed N-IDRs with <30% STQN amino acid content or ordered NTDs in promoting protein folding (Fig. 7 and Dataset S2, Tables DS2-DS4). These results are highly consistent with our finding that N-terminal addition of IDRs (≥30% STQN amino acid content) to LacZ results in >5-fold greater LacZ activity relative to LacZ alone (Fig. 5).

We also determined if the IDRs' nanny function is regulated by protein quality control (48), a mechanism by which a cell monitors proteins to ensure that they are appropriately folded, by comparing the β-galactosidase activities of Rad51-NTD-LacZ-NVH and LacZ-NVH in wild-type (WT), *hsp104*Δ, *new1*Δ, *doa1*Δ, *doa4*Δ, *san1*Δ and *oaz1*Δ lines. It is noteworthy that the protein products encoded by each of these six genes are all functionally relevant to protein homeostasis or prion propagation. Hsp104 is a heat-shock protein with disaggregase activities that disrupt aggregation of proteins. New1 is a translation factor that fine-tunes ribosome recycling and the efficiency of translation termination (32). Doa1 (also called Ufd3) is a ubiquitin- and Cdc48-binding protein with a role in ubiquitin homeostasis and/or protein degradation (49, 50). The *doa1*Δ mutant exhibits diminished formation of [*PSI*+] prion (51). Doa4 is a deubiquitinating enzyme required for recycling ubiquitin from proteasome-bound ubiquitinated intermediates (52). The *doa4*Δ mutant exhibits increased sensitivity to the protein

synthesis inhibitor cycloheximide (53). San1 is a ubiquitin-protein ligase that targets highly aggregation-prone proteins (54, 55). Oaz1 (ornithine decarboxylase antizyme) stimulates ubiquitin-independent degradation of Spe1 ornithine decarboxylase by the proteasome (56).

Our results reveal that the β-galactosidase activities of Rad51-NTD-LacZ-NVH in WT and all six gene-knockout strains we examined here are 10-29-fold higher than those of LacZ-NVH (Fig. 4*A*). Intriguingly, the β-galactosidase activities of LacZ-NVH in the six gene-knockout mutants are all lower (30-70%) than those in WT (Fig. 4*B*). In contrast, the β-galactosidase activities of Rad51-NTD-LacZ-NVH in WT are slightly higher or lower than those in the six null mutants (Fig. 4*C*). These results indicate that addition of Rad51-NTD to LacZ-NVH can lessen the protein homeostasis defects caused by loss of each of these six genes. For example, Rad51-NTD might compensate the ribosome assembly and translation defects in *new1*Δ (32), as well as the cycloheximide-hypersensitive phenotype in *doa4*Δ (53). Accordingly, the β-galactosidase activities of Rad51-NTD-LacZ-NVH in *new1*Δ and *doa4*Δ are higher than those in WT, respectively. In contrast, those of LacZ-NVH in *new1*Δ and *doa4*Δ are lower, respectively, than those in WT. Finally, although the *doa1*Δ mutant is defective in the formation of [*PSI*+] prion (51), the steady-state levels of Rad51-NTD-LacZ-NVH in *doa1*Δ are also slightly higher than those in WT. These results suggest that the nanny function of Rad51-NTD in promoting protein expression might be functionally distinct from prion nucleation or propagation.

Taken together, these findings indicate that N-IDRs with high STQN amino acid content likely exert their nanny function by promoting protein folding and/or protein quality control.

Discussion

The IDRs we have described in this study possess two previously unappreciated functions. First, they all can act as N-terminal fusion carriers to promote expression of target proteins by increasing protein folding rates and this nanny function can occur independently of ATR^{Mec1}– or ATM^{Tel1}–mediated protein phosphorylation. Second, some IDRs of non-DDR proteins may activate the cellular protein quality control machinery mediated by ATR^{Mec1} or ATM^{Tel1}. Interestingly, we reported recently that the stability (or half-lives) of N-IDR fusion proteins (e.g., Rad51 and Rad51-NTD-LacZ-NVH) (24) is tunable by ATR^{Mec1}– or ATM^{Tel1}–dependent protein phosphorylation, explaining why $Mec1^{ATR}$ and $Tel1^{ATM}$ also exhibit a DDR-independent function in regulating protein homeostasis or proteostasis (22). It is also noteworthy that the three prion-causing proteins (Sup35, Ure2 and New1) and the three N-IDR^{+} DP proteins (Vps64, Ssk2 and Kel1) we described in this study have no known DDR-related functions. Vps64 (also called Far9) is required for cytoplasm-to-vacuole targeting of proteins (57). Intriguingly, the *vps64*Δ mutant shows increased aneuploidy tolerance (58). Kel1 (Kelch repeat 1) is required for proper cell fusion and cell morphology, and functions with Kel2 (the ohnolog of Kel1) to negatively regulate mitotic exit (59). Ssk2 (suppression of sensor kinase 2) is a mitogen-activated protein kinase (MAPK) kinase of the Hog1 osmoregulatory signaling pathway (60). Activation of Ssk2 mediates actin cytoskeleton recovery from osmotic stress (61). Accordingly, our results are in agreement with an intriguing hypothesis that SCDs may possess broad and overarching roles in eukaryotic cells given that in *S. cerevisiae* and mammals there is an unexpected abundance of SCD-containing proteins belonging to functional regimes not related to DDR, such as endocytosis, microtubules, and actin cytoskeleton, as well as in pathways related to nervous system development (17, 62). Thus, we suggest that these IDRs are versatile toolkits for functional divergence in signaling and evolution. Since all the IDRs we described herein share the common feature of promoting

target protein expression, they exhibit useful potential for applications in both basic research (e.g., synthetic biology) and biotechnology.

Here, we propose a hypothetical model by which N-terminal IDRs promote high protein expression and stability *in vivo*. This hypothesis of "N-terminal intrinsic disorder facilitates abundance" (NIDFA) (Fig. 8, upper panel) provides a conceptual framework that could explain how N-terminal IDRs might fulfill their functions in promoting high-levels of protein expression either by facilitating protein folding and/or by preventing proteasomal degradation via their likely overrepresentation of $Mec1^{ATR}$ and $Tel1^{ATM}$ phosphorylation sites. This hypothesis further encompasses our recent finding that $Mec1^{ATR}$– and $Tel1^{ATM}$–dependent IDR (e.g., Rad51-NTD) phosphorylation can prevent target protein (e.g., Rad51) degradation (24), thereby explaining why these two evolutionarily conserved protein kinases exhibit an essential function in regulating protein homeostasis (22, 23).

It is important to note that there is an essential distinction between our NIDFA hypothesis (Fig. 8, top) and the previously reported NFN hypothesis (Fig. 8, bottom) (14). The latter implies that structured N-terminal domains (SNTDs) not only serve as a nucleation core for additional and specific intramolecular protein chain contacts during protein folding, but also constitute more compact modules to avoid proteolysis or aggregation. In contrast, the NIDFA hypothesis proposes that N-terminal IDRs with high STQN contents promote protein folding as well as protein quality control after protein translation. This supposition is consistent with our finding (Fig. 4) that N-terminal addition of Rad51-NTD to LacZ-NVH could compensate the ribosome assembly and translation defects in the *new1*Δ mutant, as well as defective ubiquitin recycling from proteasome-bound ubiquitinated intermediates exhibited by *doa4*Δ. It will be interesting to test if the conceptual framework provided by the NIDFA hypothesis is applicable to other N-terminal IDR proteins in yeast and in other model organisms.

Moreover, further computational and single-molecule biophysical studies on the properties of N-terminal IDRs are warranted.

Materials and Methods

Miscellaneous

All yeast plasmids and strains used in this study are listed in *SI Appendix*, Table S2 and Table S5, respectively. Guinea pig antisera against Rad51, and rabbit antisera against phosphorylated Rad51-S^{12}Q, phosphorylated Rad51-S^{30}Q, and phosphorylated Hop1-T^{318}Q were described previously (21, 24). The mouse anti-V5 antibody, anti-GFP antibody and anti-GST antibody were purchased from Bio-Rad (CA, USA, MCA1360), Clontech (SF, USA, 632381) and GenScript (NJ, USA, A00865), respectively. The rabbit anti-Sml1 antibody was purchased from Agrisera (Vännäs, SE, AS10847). The rabbit anti-Hsp104 antiserum was kindly provided by Chung Wang (Academia Sinica, Taiwan). Rabbit antisera against phosphorylated Sup35-S^{17}Q were raised using the synthetic phosphopeptide N^{12}YQQYS$^{[P]}$QNGNQQQGNNR^{28} as an antigen, where $S^{[P]}$ is phosphorylated serine. Phosphopeptide synthesis and animal immunization were conducted by LTK BioLaboratories, Taiwan. Western blotting analyses were performed as described (21, 24). Quantitative β-galactosidase activity assays were carried out as previously described (24, 63, 64).

Isolation of yeast total DNA and total RNA

Yeast cells (OD_{600} ~10) were harvested and frozen in liquid nitrogen. Total DNA (i.e., genomic DNA and plasmid DNA) was extracted by the phenol/chloroform extraction method (65). The total RNA was extracted with TRIzol Reagent solution (Invitrogen, Waltham, USA) and the RNeasy Mini Kit (Qiagen, Hilden, DE). Frozen yeast cells were disrupted by glass bead beating

with 1 mL of TRIzol Reagent solution, then mixed with 0.2 mL chloroform. After centrifugation, the aqueous phase was transferred into a new microcentrifuge tube and mixed with an equal volume of 100% ethanol to precipitate the RNA. The solution containing the precipitates was loaded onto an RNeasy spin column. Subsequent procedures followed the manufacturer's protocol, including on-column DNase I digestion with the RNase-Free DNase set (Qiagen, Hilden, DE). The quality of extracted RNA was further analyzed with the RNA 6000 Nano kit using an Agilent 2100 Bioanalyzer system (Agilent Technologies, CA, USA).

g-qPCR and RT-qPCR

g-qPCR was used to determine relative plasmid copy number. Two sets of primers were designed. The first set, *ACT1* (*SI Appendix*, Table S6), was used to amplify a single copy gene encoding actin in the budding yeast genome (66). The second set, LacZ, was targeted to amplify the β-galactosidase gene in the expression vector (67). For RT-qPCR, total RNA was converted into cDNA using oligo$(dT)_{20}$ primers and the SuperScript III First-Strand Synthesis System (Invitrogen, Waltham, USA). The cDNA samples and total DNA samples were diluted, and then mixed with gene-specific primers (*SI Appendix*, Table S6) and Fast SYBR Green Master Mix (Applied Biosystems, Waltham, USA) according to the manufacturer's recommendations. *ACT1* was also used for normalization of RT-qPCR data (66). Quantifications of g-qPCR and RT-qPCR datasets were performed in a QuantStudio™ 12K Flex Real-Time PCR System with default settings. The normalized gene expression levels were analyzed by the $2^{-\Delta\Delta C}T$ method (68) using QuantStudio™ 12K Software v1.3 (Thermo Fisher Scientific, Waltham, USA)

Acknowledgements

We thank John O′Brien for English editing, Rita Cha (Bangor University, UK) for sharing and discussion of unpublished results on Sml1 in regulating protein homeostasis, and Yi-Fang Cheng for constructing five N-IDR-LacZ-NVH expression vectors.

Author contributions

Conceptualization, design and writing of the original manuscript: TFW; Performance of experiments, data analysis and interpretation: CNC, TTW, SYT, WCL, CLC, HCL, CYC, YPH, and TFW.

Funding

T.-F.W. was supported in this work by Academia Sinica (AS108-TP-B07) and the Ministry of Science and Technology (MOST109-2311-B-001-008-MY3), Taiwan, Republic of China.

Competing interests

TFW, CNC and TTW applied for a USA provisional patent “METHODS AND VECTORS FOR ENHANCING PROTEIN EXPRESSION" on June 9, 2020.

Supporting Information File #1: six tables (S1-S6) and three figures (S1-S3)

Table S1. Sequences of IDRs used in this study.

Table S2. Plasmids used in this study.

Table S3. Results of protein synthesis rate and IUPred prediction

Table S4. Predicted protein folding rates.

Table S5. *S. cerevisiae* strains used in this study.

Table S6. Primers used for g-qPCR and RT-qPCR.

Figure S1. Western blots for visualization of LacZ-NVH fusion proteins.

Figure S2. 1,6-HD does not affect the expression-promoting function of IDRs.

Figure S3. Alanine scanning mutagenesis of IDRs.

Dataset S1

Table DS1. Protein abundance ratio and N-terminal folding prediction of duplicated proteins that differ in N-termini.

Dataset S2

Table DS2. ORFs with an ordered N-terminal domain (NTD).

Table DS3. ORFs having an N-IDR with <30% STQN amino acid content.

Table DS4. ORFs having an N-IDR with ≥30% STQN amino acid content.

References

1. V. N. Uversky, J. R. Gillespie, A. L. Fink, Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**, 415-427 (2000).
2. P. Romero *et al.*, Sequence complexity of disordered protein. *Proteins* **42**, 38-48 (2001).
3. V. N. Uversky, Intrinsically disordered proteins and their "mysterious" (meta)physics. *Frontiers in Physics* **7** (2019).
4. M. Macossay-Castillo *et al.*, The balancing act of intrinsically disordered proteins: enabling functional diversity while minimizing promiscuity. *J Mol Biol* **431**, 1650-1670 (2019).
5. P. E. Wright, H. J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**, 321-331 (1999).
6. A. E. Posey, A. S. Holehouse, R. V. Pappu, Phase separation of intrinsically disordered proteins. *Methods Enzymol* **611**, 1-30 (2018).
7. A. A. Santner *et al.*, Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. *Biochemistry* **51**, 7250-7262 (2012).
8. F. Montanari, D. C. Shields, N. Khaldi, Differences in the number of intrinsically disordered regions between yeast duplicated proteins, and their relationship with functional divergence. *PLoS One* **6**, e24989 (2011).
9. R. Pancsa, P. Tompa, Coding regions of intrinsic disorder accommodate parallel functions. *Trends Biochem Sci* **41**, 898-906 (2016).
10. K. H. Wolfe, D. C. Shields, Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708-713 (1997).
11. H. J. Dyson, P. E. Wright, Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**, 54-60 (2002).

12. C. J. Oldfield *et al.*, Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **9 Suppl 1**, S1 (2008)

13. P. Tompa *et al.*, Close encounters of the third kind: disordered domains and the interactions of proteins. *Bioessays* **31**, 328-335 (2009).

14. P. C. Simister, F. Schaper, N. O'Reilly, S. McGowan, S. M. Feller, Self-organization and regulation of intrinsically disordered proteins with folded N-termini. *PLoS Biol* **9**, e1000591 (2011).

15. R. J. Craven, P. W. Greenwell, M. Dominska, T. D. Petes, Regulation of genome stability by *TEL1* and *MEC1*, yeast homologs of the mammalian ATM and ATR genes. *Genetics* **161**, 493-507 (2002).

16. S. T. Kim, D. S. Lim, C. E. Canman, M. B. Kastan, Substrate specificities and identification of putative substrates of ATM kinase family members. *J Biol Chem* **274**, 37538-37543 (1999).

17. H. C. Cheung *et al.*, An S/T-Q cluster domain census unveils new putative targets under Tel1/Mec1 control. *BMC Genomics* **13**, 664 (2012).

18. A. Traven, J. Heierhorst, SQ/TQ cluster domains: concentrated ATM/ATR kinase phosphorylation site regions in DNA-damage-response proteins. *Bioessays* **27**, 397-407 (2005).

19. H. Lee *et al.*, Diphosphothreonine-specific interaction between an SQ/TQ cluster and an FHA domain in the Rad53-Dun1 kinase cascade. *Mol Cell* **30**, 767-778 (2008).

20. J. A. Carballo, A. L. Johnson, S. G. Sedgwick, R. S. Cha, Phosphorylation of the axial element protein Hop1 by Mec1/Tel1 ensures meiotic interhomolog recombination. *Cell* **132**, 758-770 (2008).

21. C. N. Chuang, Y. H. Cheng, T. F. Wang, Mek1 stabilizes Hop1-Thr318 phosphorylation to promote interhomolog recombination and checkpoint responses during yeast meiosis. *Nucleic Acids Res* **40**, 11416-11427 (2012).

22. I. Corcoles-Saez *et al.*, Essential function of Mec1, the budding yeast ATM/ATR checkpoint-response kinase, in protein homeostasis. *Dev Cell* **46**, 495-503 e492 (2018).

23. I. Corcoles-Saez, K. Dong, R. S. Cha, Versatility of the Mec1(ATM/ATR) signaling network in mediating resistance to replication, genotoxic, and proteotoxic stresses. *Curr Genet* 10.1007/s00294-018-0920-y (2019).

24. T. T. Woo, C. N. Chuang, M. Higashide, A. Shinohara, T. F. Wang, Dual roles of yeast Rad51 N-terminal domain in repairing DNA double-strand breaks. *Nucleic Acids Res* **48**, 8474-8489 (2020).

25. J. A. Toombs, N. M. Liss, K. R. Cobble, Z. Ben-Musa, E. D. Ross, [*PSI*+] maintenance is dependent on the composition, not primary sequence, of the oligopeptide repeat domain. *PLoS One* **6**, e21953 (2011).

26. R. B. Wickner *et al.*, Prions: proteins as genes and infectious entities. *Genes Dev* **18**, 470-485 (2004).

27. R. B. Wickner, [*URE3*] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* **264**, 566-569 (1994).

28. F. Shewmaker, L. Mull, T. Nakayashiki, D. C. Masison, R. B. Wickner, Ure2p function is enhanced by its prion domain in *Saccharomyces cerevisiae*. *Genetics* **176**, 1557-1565 (2007).

29. S. Ngo, V. Chiang, E. Ho, L. Le, Z. Guo, Prion domain of yeast Ure2 protein adopts a completely disordered structure: a solid-support EPR study. *PLoS One* **7**, e47248 (2012).

30. A. Santoso, P. Chien, L. Z. Osherovich, J. S. Weissman, Molecular basis of a yeast prion species barrier. *Cell* **100**, 277-288 (2000).

31. L. Z. Osherovich, J. S. Weissman, Multiple Gln/Asn-rich prion domains confer susceptibility to induction of the yeast [*PSI*(+)] prion. *Cell* **106**, 183-194 (2001).

32. V. Kasari *et al.*, A role for the *Saccharomyces cerevisiae* ABCF protein New1 in translation termination/recycling. *Nucleic Acids Res* **47**, 8807-8820 (2019)

33. A. Bachmair, D. Finley, A. Varshavsky, In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179-186 (1986).

34. T. Tasaki, S. M. Sriram, K. S. Park, Y. T. Kwon, The N-end rule pathway. *Annu Rev Biochem* **81**, 261-289 (2012).

35. A. Varshavsky, N-degron and C-degron pathways of protein degradation. *Proc Natl Acad Sci U S A* **116**, 358-366 (2019).

36. C. S. Hwang, A. Shemorry, D. Auerbach, A. Varshavsky, The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. *Nat Cell Biol* **12**, 1177-1185 (2010).

37. B. Ho, A. Baryshnikova, G. W. Brown, Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* Proteome. *Cell Systems* **6**, 192-205 (2018).

38. M. Breker, M. Gymrek, M. Schuldiner, A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. *J Cell Biol* **200**, 839-850 (2013).

39. Y. T. Chong *et al.*, Yeast proteome dynamics from single cell imaging and automated analysis. *Cell* **161**, 1413-1424 (2015).

40. B. Meszaros, G. Erdos, Z. Dosztanyi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* **46**, W329-W337 (2018).

41. G. Erdos, Z. Dosztanyi, Analyzing protein disorder with IUPred2A. *Curr Protoc Bioinformatics* **70**, e99 (2020).

42. F. Ness, P. Ferreira, B. S. Cox, M. F. Tuite, Guanidine hydrochloride inhibits the generation of prion "seeds" but not prion protein aggregation in yeast. *Mol Cell Biol* **22**, 5593-5605 (2002)

43. S. Marqusee, V. H. Robbins, R. L. Baldwin, Unusually stable helix formation in short alanine-based peptides. *Proc Natl Acad Sci U S A* **86**, 5286-5290 (1989).

44. A. Riba *et al.*, Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc Natl Acad Sci U S A* **116**, 15023-15032 (2019).

45. Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434 (2005).

46. M. M. Gromiha, A. M. Thangakani, S. Selvaraj, FOLD-RATE: prediction of protein folding rates from amino acid sequence. *Nucleic Acids Res* **34**, W70-74 (2006).

47. M. Corrales *et al.*, Machine learning: How much does it tell about protein folding rates? *PLoS One* **10**, e0143166 (2015)

48. B. Chen, M. Retzlaff, T. Roos, J. Frydman, Cellular strategies of protein quality control. *Cold Spring Harb Perspect Biol* **3**, a004374 (2011).

49. J. E. Mullally, T. Chernova, K. D. Wilkinson, Doa1 is a Cdc48 adapter that possesses a novel ubiquitin binding domain. *Mol Cell Biol* **26**, 822-830 (2006).

50. G. Zhao, G. Li, H. Schindelin, W. J. Lennarz, An Armadillo motif in Ufd3 interacts with Cdc48 and is involved in ubiquitin homeostasis and protein degradation. *Proc Natl Acad Sci U S A* **106**, 16197-16202 (2009).

51. J. Tyedmers, M. L. Madariaga, S. Lindquist, Prion switching in response to environmental stress. *PLoS Biol* **6**, e294 (2008).

52. S. Swaminathan, A. Y. Amerik, M. Hochstrasser, The Doa4 deubiquitinating enzyme is required for ubiquitin homeostasis in yeast. *Mol Biol Cell* **10**, 2583-2594 (1999).

53. A. M. Dudley, D. M. Janse, A. Tanay, R. Shamir, G. M. Church, A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol Syst Biol* **1**, 2005 0001 (2005).

54. A. Dasgupta, K. L. Ramsey, J. S. Smith, D. T. Auble, Sir Antagonist 1 (San1) is a ubiquitin ligase. *J Biol Chem* **279**, 26830-26838 (2004).

55. E. K. Fredrickson, P. S. Gallagher, S. V. Clowes Candadai, R. G. Gardner, Substrate recognition in nuclear protein quality control degradation is governed by exposed hydrophobicity that correlates with aggregation and insolubility. *J Biol Chem* **288**, 6130-6139 (2013).

56. Z. Porat *et al.*, Yeast antizyme mediates degradation of yeast ornithine decarboxylase by yeast but not by mammalian proteasome: new insights on yeast antizyme. *J Biol Chem* **283**, 4528-4534 (2008).

57. C. J. Bonangelino, E. M. Chavez, J. S. Bonifacino, Genomic screen for vacuolar protein sorting genes in *Saccharomyces cerevisiae*. *Mol Biol Cell* **13**, 2486-2501 (2002).

58. E. M. Torres *et al.*, Identification of aneuploidy-tolerating mutations. *Cell* **143**, 71-83 (2010).

59. J. Philips, I. Herskowitz, Identification of Kel1p, a kelch domain-containing protein involved in cell fusion and morphology in *Saccharomyces cerevisiae*. *J Cell Biol* **143**, 375-389 (1998)

60. T. Maeda, M. Takekawa, H. Saito, Activation of yeast PBS2 MAPKK by MAPKKKs or by binding of an SH3-containing osmosensor. *Science* **269**, 554-558 (1995).

61. T. Yuzyuk, M. Foehr, D. C. Amberg, The MEK kinase Ssk2p promotes actin cytoskeleton recovery after osmotic stress. *Mol Biol Cell* **13**, 2869-2880 (2002).

62. L. Cara, M. Baitemirova, J. Follis, M. Larios-Sanz, A. Ribes-Zamora, The ATM- and ATR-related SCD domain is over-represented in proteins involved in nervous system development. *Sci Rep* **6**, 19050 (2016).

63. F. M. Lin, Y. J. Lai, H. J. Shen, Y. H. Cheng, T. F. Wang, Yeast axial-element protein, Red1, binds SUMO chains to promote meiotic interhomologue recombination and chromosome synapsis. *EMBO J* **29**, 586-596 (2010).

64. M. Niethammer, M. Sheng, Identification of ion channel-associated proteins using the yeast two-hybrid system. *Methods Enzymol* **293**, 104-122 (1998).

65. C. S. Hoffman, F. Winston, A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* **57**, 267-272 (1987)

66. C. Zhang *et al.*, Transcriptomic profiling of chemical exposure reveals roles of Yap1 in protecting yeast cells from oxidative and other types of stresses. *Yeast* **33**, 5-19 (2016).

67. R. A. Elbarbary, W. Li, B. Tian, L. E. Maquat, STAU1 binding 3' UTR IRAlus complements nuclear retention to protect cells from PKR-mediated translational shutdown. *Genes Dev* **27**, 1495-1510 (2013).

68. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method. *Methods* **25**, 402-408 (2001).

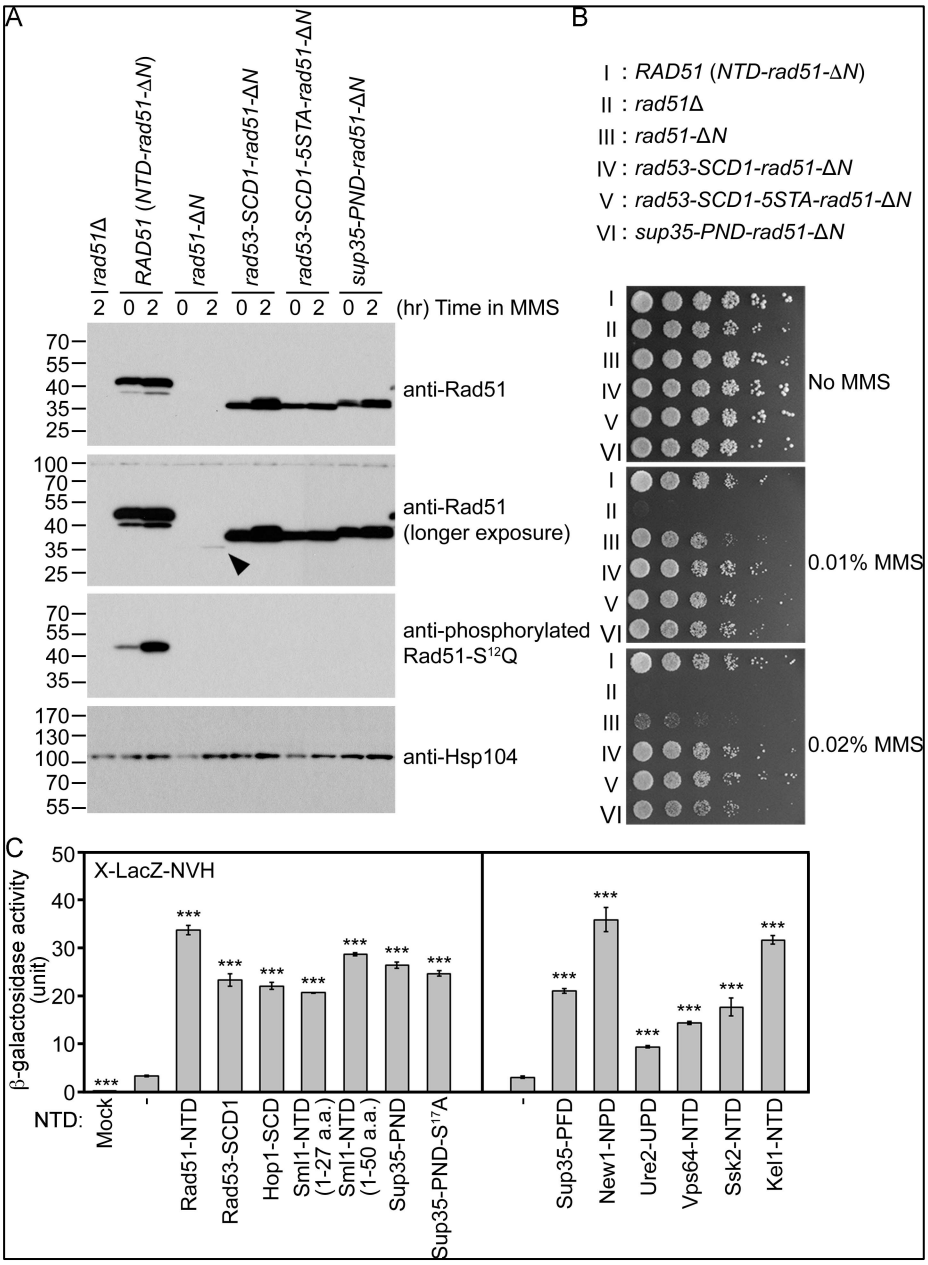

Figure 1. N-terminal fusion of SCD or the prion (nucleation) domain promotes high-level expression of target proteins. (A) Visualization of native Rad51 (NTD-Rad51-ΔN), Rad51-ΔN, and the Rad51-ΔN fusion proteins by immunoblotting. Hsp104 was used as a loading control. Size in kilodaltons of standard protein markers is labeled to the left of the blots. The black arrowhead indicates the protein band of Rad51-ΔN. (B) MMS sensitivity. Spot assay showing five-fold serial dilutions of indicated strains grown on YPD plates with or without MMS at the indicated concentrations (w/v). (C) Quantitative β-galactosidase assays. Error bars indicate standard deviation between experiments ($n \geq 3$). Asterisks indicate significant differences with P values calculated using a two-tailed T-test (***, P value <0.001).

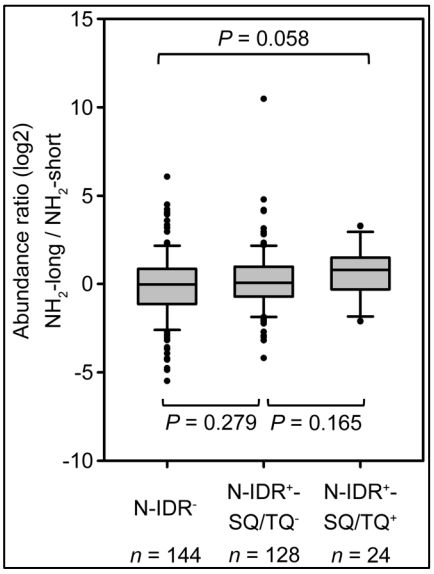

Figure 2. Gain of N-IDRs with SQ/TQ motifs after WGD results in a moderate tendency toward higher proteome-wide protein expression. The likely propensity of increasing protein abundance by the presence of at least two S/T-Q motifs in N-IDRs. Box-plot displaying the distribution of protein abundance ratio for DP pairs belonging to the indicated groups. The *S. cerevisiae* ohnolog lists (547 pairs) were downloaded from YGOB (Yeast Gene Order Browser, http://ygob.ucd.ie/browser/ohnologs.html), and 296 DP pairs having different N-terminal lengths and available abundance data were analyzed. Y-axis represents the log2 abundance ratio of median molecule number per cell (37), after GFP autofluorescence filtering of DPs having longer N-termini (NH_2-long) divided by those of their corresponding ohnologs (NH_2-short). Center line, median; box limits, upper and lower quartiles; whiskers, the 5th and 95th percentiles (outliers shown as black dots). The boxplot was generated using SigmaPlot (v. 13.0.0.83, Systat Software, CA, USA) with *P* values between groups calculated by Mann-Whitney Rank Sum Tests. The homologous region shared by the DP pair was determined by protein sequence alignments performed in MAFFT (v.7.215), and the starting positions in each protein are shown in (Dataset S1, Tables DS1). N-IDRs were determined by IUPred software using the "structured regions" parameter to predict globular regions, with the non-globular N-termini >10 amino acids in length specific to each of the longer ohnologs being determined as N-IDRs (8). Protein abundance values are shown as the median number of molecules per cell.

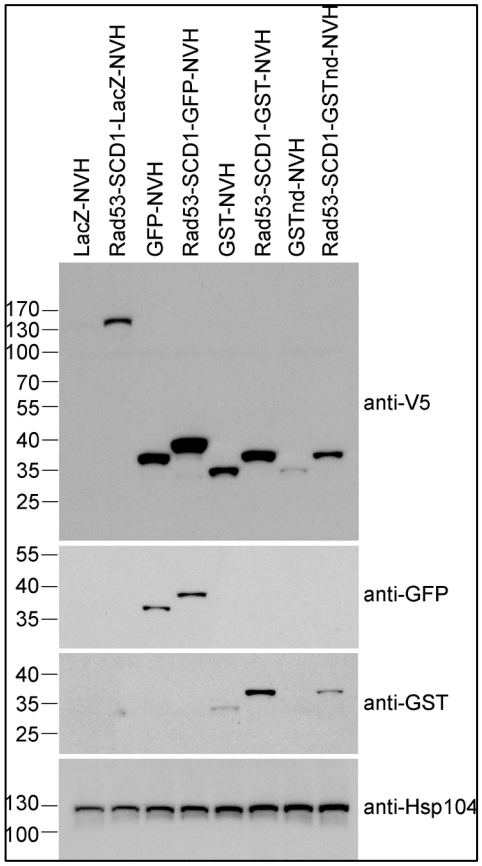

Figure 3. Rad53-SCD1 can be used as an N-terminal fusion tag to enhance production of four different target proteins: LacZ-NVH, GST-NVH, GSTnd-NVH and GFP-NVH.

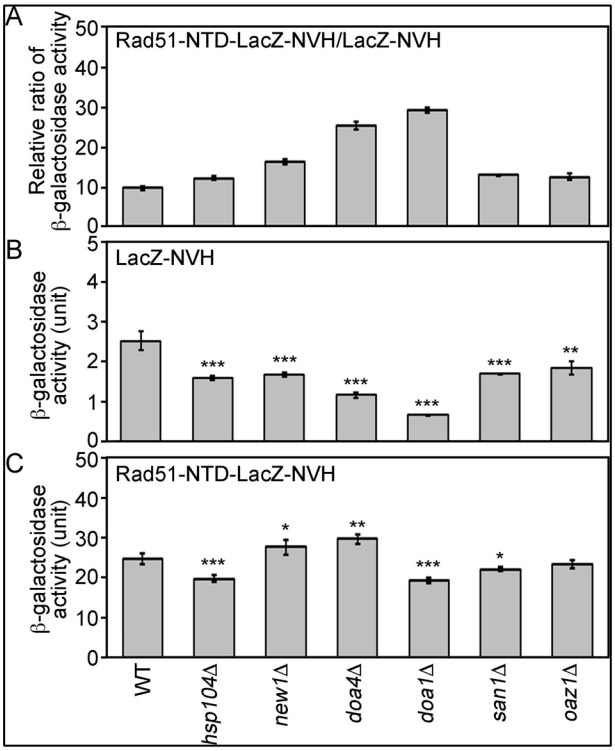

Figure 4. The nanny function of Rad51-NTD is controlled during protein translation and does not affect ubiquitin-mediated protein degradation. (A) The impacts of six protein homeostasis genes on the β-galactosidase activity ratios of Rad51-NTD-LacZ-NVH to LacZ-NVH. (B, C) The β-galactosidase activities of LacZ-NVH (B) and Rad51-NTD-LacZ-NVH (C) in WT and the six gene knockout mutants. Asterisks indicate significant differences with *P* values calculated using a two-tailed *T*-test (***, *P* value <0.001; **, *P* value <0.01; *, *P* value <0.05).

A

Name of SCDs or IDRs	Length	S	T	Q	N	STQN(%)	Relative LacZ activity
Rad51-NTD	66	9	2	9	4	36.4	10.65±0.27
Rad53-SCD1	29	2	4	7	1	48.3	7.35±0.27
Hop1-SCD	67	6	6	8	9	43.3	7.00±0.19
Sml1-NTD(1-27 a.a.)	27	3	3	7	2	55.6	6.52±0.12
Sml1-NTD(1-50 a.a.)	50	5	4	8	3	40	9.05±0.18
Sup35-PND	39	3	0	12	9	61.5	8.34±0.20
Sup35-PND-S^{17}A	39	2	0	12	9	59.0	7.80±0.18
Sup35-PFD	114	5	0	32	19	49.1	7.12±0.26
New1-NPD	156	19	8	14	28	44.2	12.19±0.65
Ure2-UPD	91	10	5	10	33	63.7	3.16±0.12
Vps64-NTD	152	21	10	15	8	35.53	4.86±0.18
Ssk2-NTD	191	40	13	11	13	40.31	5.99±0.43
Kel1-NTD	69	11	6	5	3	36.23	10.75±0.40
Rad51-NTD	66	9	2	9	4	36.4	11.27±0.21
Rad51-NTD-3SA	66	6	2	9	4	31.8	10.22±0.24
Rad51-NTD-6SQA	66	6	2	6	4	27.3	7.66±0.05
Rad51-NTD-9SQA	66	6	2	3	4	22.7	2.54±0.01
Rad51-NTD-12SQA	66	6	2	0	4	18.2	1.25±0.01
Sup35-PND	39	3	0	12	9	61.5	8.30±0.08
Sup35-PND-S^{17}A	39	2	0	12	9	59.0	7.13±0.06
Sup35-PND-3SA	39	0	0	12	9	53.8	5.18±0.03
Sup35-PND-3QA	39	3	0	9	9	53.8	8.02±0.12
Sup35-PND-5QA	39	3	0	7	9	48.7	6.93±0.11
Sup35-PND-8QA	39	3	0	4	9	41.0	7.09±0.05
Sup35-PND-15SQA	39	0	0	0	9	23.1	3.02±0.03
Sup35-PND-9NA	39	3	0	12	0	38.5	5.14±0.10
Sup35-PND-24SQNA	39	0	0	0	0	0	-0.07±0.00
Rad53-SCD1	29	2	4	7	1	48.3	6.46±0.12
Rad53-SCD1-5STA	29	1	0	7	1	31.0	5.34±0.06
Rad53-SCD1-7QA	29	2	4	0	1	24.1	4.32±0.07
Rad53-SCD1-12STQA	29	1	0	0	1	6.9	1.14±0.01

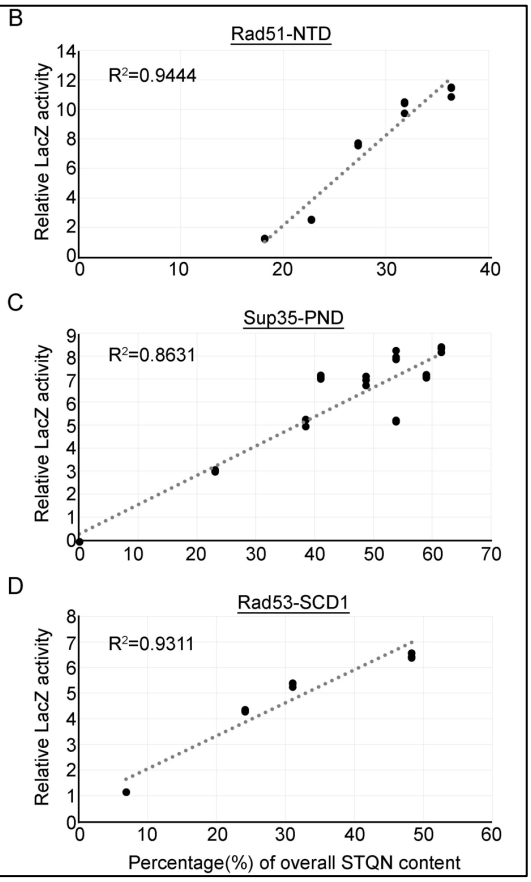

Figure 5. Relative β-galactosidase (LacZ) activities are correlated with the percentage STQN amino acid content of IDRs. (A) List of IDRs with their respective length, numbers of S/T/Q/N amino acids, overall STQN percentage, and relative β-galactosidase activity. (B-D) Linear regressions between relative β-galactosidase activities and overall STQN percentages for Rad51-NTD (B), Sup35-PND (C), and Rad53-SCD1 (D). The coefficients of determination (R^2) are indicated for each simple linear regression.

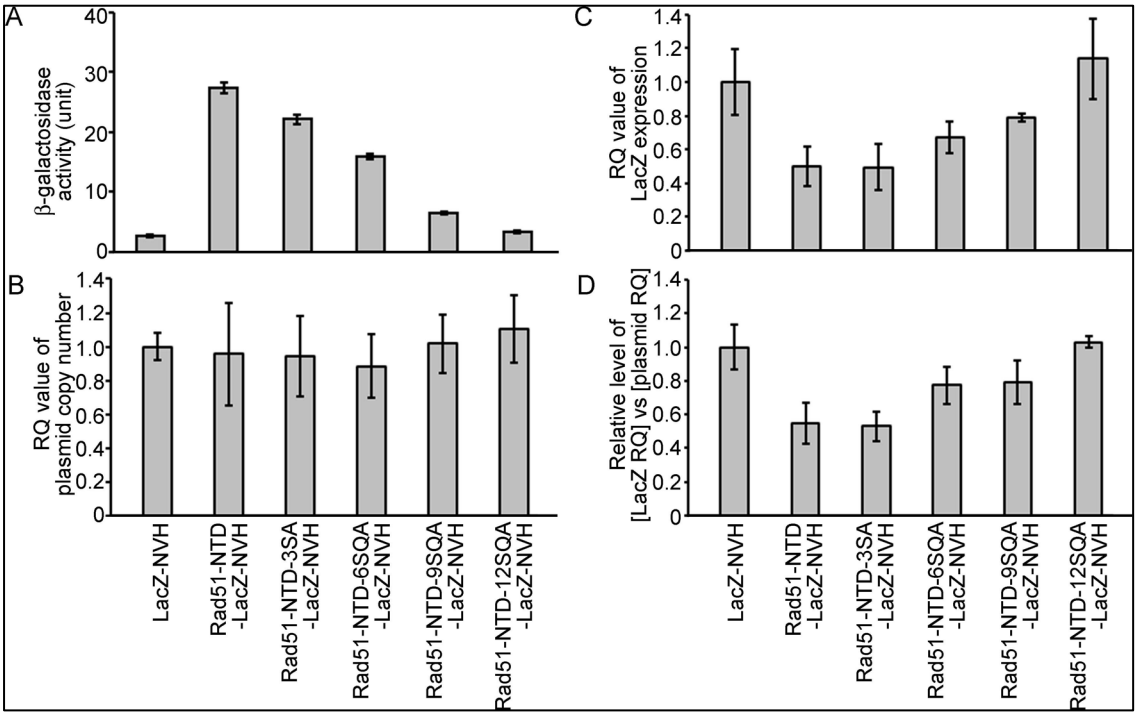

Figure 6. The nanny function of Rad51-NTD is unlikely to be controlled during transcription or simply arise from plasmid copy number differences. The effects of WT and mutant Rad51-NTD on β-galactosidase activities (A), plasmid DNA copy numbers (B), relative steady-state levels of LacZ-NVH mRNA normalized to *ACT1* (actin) mRNA (C), and relative ratios of LacZ-NVH mRNA *versus* plasmid DNA copy number (D). The wild-type yeast cells were transformed with indicated *CEN-ARS* plasmids, respectively, to express WT and mutant Rad51-NTD-LacZ-NVH fusion proteins or LacZ-NVH alone under the control of the native *RAD51* gene promoter (P_{RAD51}). The relative quantification (RQ = $2^{-\Delta\Delta C_T}$) values were determined to reveal the plasmid DNA copy number and steady-state levels of LacZ-NVH mRNA by g-qPCR and RT-qPCR, respectively. LacZ and *ACT1* were selected as target and reference protein-encoding genes, respectively, in both g-qPCR and RT-qPCR. The data shown represent mean ± SD from three independent biological data-points.

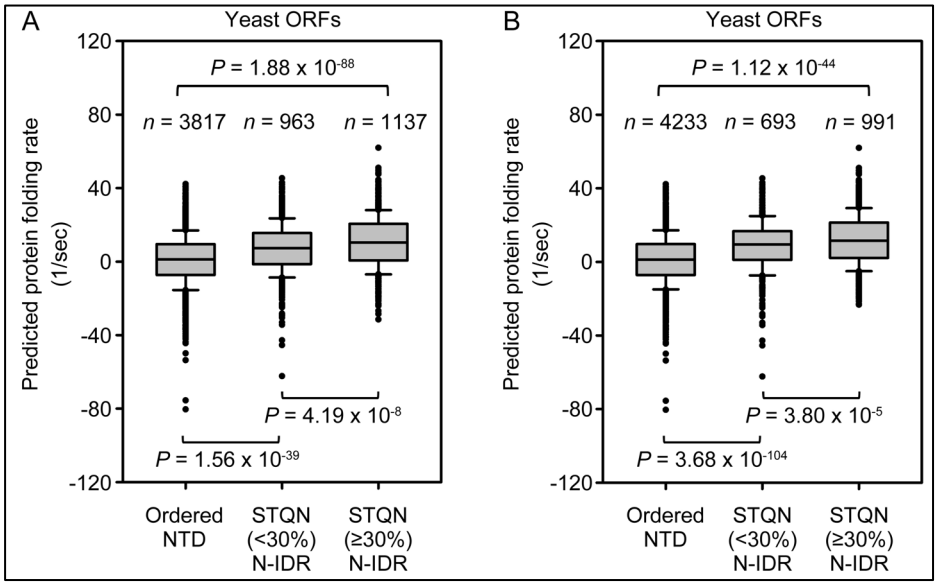

Figure 7. Predicted protein folding rates are higher for N-IDRs with greater STQN amino acid content. The amino acid sequences of all (5,917) *S. cerevisiae* ORFs were downloaded from the *Saccharomyces* Genome Database (SGD) and then analyzed using "FOLD-RATE" (https://www.iitm.ac.in/bioinfo/fold-rate/) (40, 41) and "IUPred structured domains" (https://iupred.elte.hu) (46) to determine protein folding rates (Y-axis) and to characterize their N-IDRs, respectively. The lengths of N-IDRs were determined in IUPred using the "structured regions" parameter, with the non-globular N-termini >10 amino acids in length being determined as N-IDRs (8) (A), or without the amino acid length restriction (B). The STQN amino acid content of N-IDRs were also determined. Center line, median; box limits, upper and lower quartiles; whiskers, the 5th and 95th percentiles (outliers shown as black dots). The boxplot was generated using SigmaPlot (v. 13.0.0.83, Systat Software, CA, USA) with *P* values between groups calculated by Mann-Whitney Rank Sum Tests.

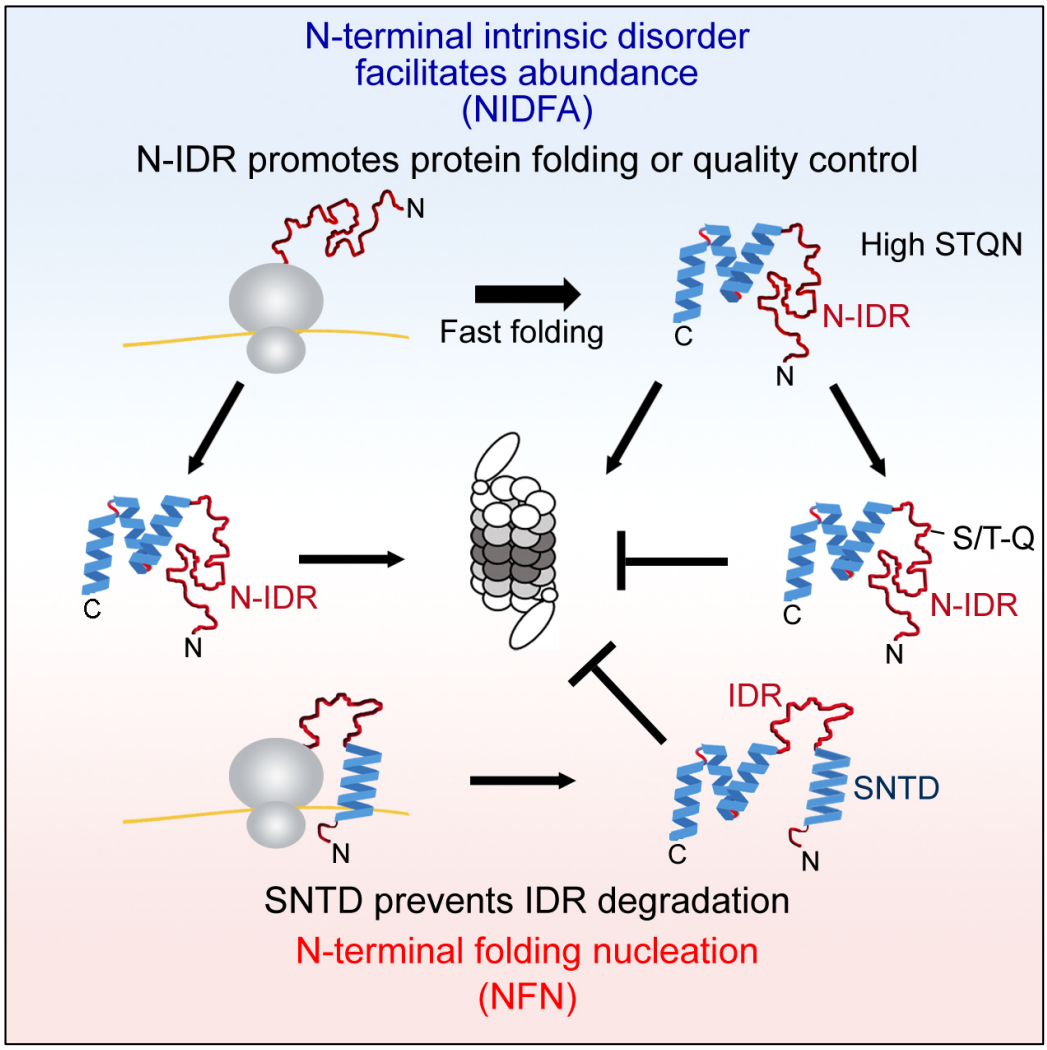

Figure 8. Illustrated comparison of the N-terminal intrinsic disorder facilitates abundance (NIDFA) and the N-terminal folding nucleation (NFN) hypotheses. The NIDFA hypothesis proposes that N-terminal IDRs can promote protein expression by facilitating protein folding or quality control. Moreover, N-terminal IDRs promote folding and/or quality control of the nascent polypeptide chain. This hypothetical model further encompasses our recent finding that Mec1^{ATR}– and Tel1^{ATM}–dependent IDR (e.g., Rad51-NTD) phosphorylation can prevent proteolytic degradation due to the likely overrepresentation of S/T-Q motifs in IDRs. In contrast, the NFN hypothesis (14) proposes that as the nascent polypeptide chain emerges from the ribosome (in grey), the intrinsically structured N-terminal domain (SNTD) folds rapidly and spontaneously, serving as a nucleation core for intramolecular protein chain contacts that generate a more compact protein module. This compaction may help to avoid proteolytic degradation and abnormal aggregation.