Molecular basis of Eukaryotic Unfolded Protein Response (UPR)

Protein folding is strategically important to cellular function. Secreted, membrane-bound and organelle-targeted proteins are typically processed and folded in the endoplasmic reticulum (ER) in eukaryotes (**???**). Intracellular perturbations caused by a variety of stressors disturb the specialized environment of the ER leading to the accumulation of unfolded proteins (**??**). Normally, cells ensure that proteins are correctly folded using a combination of molecular chaperones, foldases and lectins (**?**). However, when proper folding can not be restored, incorrectly folded proteins are targeted to ER Associated Degradation (ERAD) pathways for processing (**?**). If unfolded or misfolded proteins continue to accumulate, eukaryotes induce the unfolded protein response (UPR).

In mammalian cells, UPR is a complex signaling program mediated by three ER transmembrane receptors: activating transcription factor 6 (ATF6), inositol requiring kinase 1 (IRE1) and double-stranded RNA-activated protein kinase (PKR)-like endoplasmic reticulum kinase (PERK). UPR performs three functions, adaptation, alarm and apoptosis. During adaptation, the UPR tries to reestablish folding homeostasis by inducing the expression of chaperones that enhance protein folding. Simultaneously, translation is globally attenuated to reduce the ER folding load while the degradation of unfolded proteins is increased. If these steps fail, the UPR induces a cellular alarm and apoptosis program. The alarm phase involves several signal transduction events, ultimately leading to the removal of the translational block and the down-regulation of the expression and activity of pro-survival factors such as the B-cell lymphoma 2 (Bcl2) protein. After the alarm phase, cells can undergo apoptosis, although ER stress can also initiate autophagy (**???????**). Thus, ER folding homeostasis strongly influences physiology (**?**). Aberrant protein folding and UPR have been implicated in a number of pathologies. For example, the onset of diabetes (**?**) as well as myocardial ischaemia, cardiac hypertrophy, atherosclerosis and heart failure (**?**) have all been linked with aberrant folding or UPR signaling.

The folding cycle, quality control and ER associated degradation (ERAD): Newly synthesized polypeptide chains enter the ER through a peptide translocon in the ER mem-

brane composed of four proteins, Sec61α,β,γ and TRAM (**?**). Upon entering the ER, these nascent chains begin to fold, often as they are being co-translationally modified (**?**). The folding quality of proteins in the ER is maintained by an in-built quality control (QC) system which ensures proteins are in their native folded state before exiting the ER (**??**). A protein is correctly folded, if it has attained its native conformation after required co- or post-translational modifications. On the other hand, exposed hydrophobic regions, unpaired cysteine residues, or aggregation are all markers of an unfolded or misfolded conformation (**?**), which leads to subsequent retro-translocation to the cytosol. Once in the cytosol, these unfolded or misfolded proteins are degraded by the ubiquitin proteasome system (**?**). Hydrophobic unfolded or misfolded queues are recognized in the ER by molecular chaperones which bind these queues and increase the probability of correct folding (**???**). For example, the HSP70 family of chaperones recognize, in an ATP-dependent manner, exposed hydrophobic patches on a broad spectrum of unfolded or misfolded proteins (**?**). Repeated binding and release of HSP70 chaperones ensures that incorrectly folded proteins do not exit the ER (**?**). One critical member of the HSP70 family is BiP or GRP78. BiP consists of an N-terminal ATPase domain and a C-terminal peptide binding domain (**?**). BiP also regulates the activation of the three transmembrane ER stress transducers: PERK, ATF6 and IRE1. Normally, BiP is bound to these ER receptors, blocking their activation. However, in the presence of exposed hydrophobic residues BiP disassociates, allowing PERK, ATF6 and IRE1 activation. Overexpression of BiP leads to reduced activation of IRE1 and PERK (**??**). The PERK and ATF6 branches are thought to be activated before IRE1 (**?**); this ordering is consistent with the signals that each branch transduces. The PERK and ATF6 pathways largely promote ER adaptation to misfolding, while IRE1 has a dual role, transmitting both survival and pro-apoptotic signals.

Double-stranded RNA-activated protein kinase (PKR)-like endoplasmic reticulum kinase (PERK) pathway: The PERK branch of UPR transduces both pro-survival as well as pro-apoptotic signals following the accumulation of unfolded or misfolded protein in the ER. PERK is a type I transmembrane protein, composed of a ER luminal stress

sensor and a cytosolic protein kinase domain. Dissociation of BiP from the N-terminus of PERK initiates dimerization and autophosphorylation of the kinase domain at T981 (**?**). The eIF2α protein, which is composed of three subunits, is critical to translation initiation in eukaryotes, including GTP-dependent start-site recognition (**?**). Activated PERK can phosphorylate eIF2α at S51 (**??**), which leads to three downstream effects. First, phosphorylated eIF2α globally attenuates translation initiation (Not included in the current model). Decreased translation reduces the influx of protein into the ER, hence diminishing the folding load. Translation attenuation is followed by increased clearance of the accumulated proteins from the ER by ERAD and expression of pro-survival genes. For example, PERK activation induces of the expression of cellular inhibitor of apoptosis (cIAP) (**?**). Interestingly, decreased protein translation is not universal; genes with internal ribosome entry site (IRES) sequences in the 5′ untranslated regions bypass the eIF2α translational block (**?**). One of the most well-studied of these, *ATF4*, encodes a cAMP response element-binding transcription factor (C/EBP) (**?**) ATF4 that drives the expression of pro-survival functions such as amino acid transport and synthesis, redox reactions and protein secretion (**?**). Taken together, these effects seem to be largely pro-survival. However, ATF4 can also induce the expression of pro-apoptotic factors. For example, ATF4 induces the expression of the transcription factor C/EBP homologous protein (CHOP), which is associated with apoptotic cell-death. CHOP (also known as GADD153) is 29 kDa protein composed of an N-terminal transcriptional activation domain and a C-terminal basic-leucine zipper (bZIP) domain that is normally present at low levels in mammalian cells (**?**). The transcriptional activator domain is positively regulated by phosphorylation at S78 and S81 by p38 MAPK family members (**??**) while the bZIP domain plays a key role in the homodimerization of the protein (**??**). CHOP activity promotes apoptosis primarily by repression of Bcl2 expression and the sensitization of cells to ER-stress inducing agents (**??**).

Activating transcription factor 6 (ATF6) pathway: ATF6 activation involves a complex series of translocation and irreversible proteolytic processing steps, ultimately leading to

the up-regulation of a pro-survival transcriptional program, in the presence of unfolded or misfolded proteins. ATF6 is a 90 kDa ER transmembrane protein with two homologs: ATF6α (**??**) and ATF6β (**???**). In the current model, only ATF6α is included. Similar to IRE1 and PERK, ER stress leads to the dissociation of BIP from the N-terminus of ATF6, followed by translocation and activation. N-terminal golgi localization sequences (GLS1 and GLS2) seem to be involved with BiP regulation of ATF6. BiP binding to the N-terminal GLS1 promotes the retention of ATF6 in the ER (**?**). On the other hand, the GLS2 domain was required to target ATF6 to the golgi body following BiP dissociation from GLS1 (**?**). Unlike the previous two kinase pathways, ATF6 activation does not involve phosphorylation of a C-terminal kinase domain. Rather, after translocated to the golgi, ATF6 undergoes regulated intramembrane proteolysis (RIP); the luminal domain is first cleaved by serine protease site-1 protease (S1P) followed by metalloprotease site-2 protease (S2P) cleavage (**????**). Cleavage at the juxtamembrane site allows the 50 kDa transcriptional domain of ATF6 to be translocated to the nucleus where it regulates the expression of genes with ATF/cAMP response elements (CREs) (**?**) and ER stress response elements (ERSE) in their promoters (**??**). Cleaved ATF6 induces a gene expression program, in conjunction with other bZIP transcription factors and required co-regulators, such as nuclear factor Y (NF-Y) (**??**), that increases chaperone activity as well as the degradation of unfolded proteins (**??**). For example, ATF6 upregulates BiP, protein disulfide isomerase (PDI) and ER degradation-enhancing alpha-mannosidase-like protein 1 (EDEM1) expression. Additionally, ATF6 induces the expression of the X box-binding protein 1 (XBP1) which, after processing by activated IRE1α, induces the expression of chaperones. The ATF6-induced gene expression program is also cytoprotective. For example, ATF6 induces regulator of calcineurin 1 (RCAN1) expression (**?**). RCAN1 sequesters calcineurin (**?**), a calcium activated protein-phosphotase B, that dephosphorylates Bcl2-antagonist of cell death (BAD) at S75 or S99 (**?**). This leads to sequestering of Bcl2 by Bad, which inhibits its downstream anti-apoptotic activity (**?**).

Inositol-requiring kinase 1 (IRE1) pathway: IRE1 initiates a program with both pro-survival and pro-apoptotic components in the presence of misfolded or unfolded proteins. IRE1 is a 100 kDa type I ER transmembrane protein with both an endoribonuclease and a serine-threonine kinase domain (**?**). IRE1 has two homologs, IRE1α and IRE1β; IRE1α is expressed in a variety of tissues (**?**) while IRE1β is found only in the intestinal epithelia (**??**). In the current model only IRE1α has been considered. The N-terminus of IRE1, located in the ER lumen, senses unfolded or misfolded proteins through its interaction with BiP (**???**). Normally BiP is bound to the N-terminus of IRE1 (**???**). However, in the presence of unfolding queues BiP dissociates and is sequestered by the unfolded or misfolded proteins (**?**). Subsequently, IRE1 is activated by homooligomerization followed by autophosphorylation of the C-terminal kinase domain at S724 (**????**). IRE1 activation enables both its kinase and endoribonuclease activities to transduce signals simultaneously through two distinct signaling axes. The endoribonuclease activity cleaves a 26-nucleotide intron from the XBP1-mRNA (**???**) which generates a 41 kDa frameshift variant (sXBP1) that acts as a potent transcription factor. sXBP1 homodimers, along with co-regulators such as nuclear factor Y (NF-Y), regulate the expression of a variety of ER chaperones and protein degradation related genes (**??**). Cytosolic IRE1α dimers interact with adaptors such as tumor necrosis factor receptor-associated factor 2 (TRAF2) to drive signal-regulating kinase (ASK1) activation and then subsequently cJUN NH_2-terminal kinase (JNK) and p38MAPK activation (**?**). ASK1 activity is regulated by phosphorylation/de-phosphorylation at several sites as well as by physical interaction with other proteins. ASK1 phosphorylates and activates two downstream kinases, MMK4 and MMK3 which in turn activate JNK and p38 MAP kinase, respectively. JNK is activated by dual phosphorylation at T183 and Y185 by MMK4 (**?**). Activated JNK activates the proapoptotic Bcl-2 family member Bim by phosphorylation at S65 (**??**). JNK activation also regulates the activity of anti-apoptotic protein Bcl2 (**??**). Active JNK1 inhibits Bcl2 via phosphorylation at sites T69, S70 and S87 (**?**). Ultimately, inhibition of Bcl2 and the activation of Bim leads to BAX/BAK dependent apoptosis. Thus, signals initiated from the cytosolic kinase domain of IRE1α are largely pro-apoptotic. IRE1α activity is regulated by

protein serine/threonine phosphatase (PTC2P).

ER stress-induced apoptosis: Ultimately, if UPR fails to restore ER homeostasis, cells initiate terminal programs such as apoptosis. A common biomarker of apoptosis is the activation of aspartate-specific proteases, collectively known as caspases (**?**). Caspases rapidly dismantle cell cycle, cytoskeletal and organelle proteins by proteolytic cleavage. There are two pathways that result in caspase activation in response to apoptotic signals; the death-receptor and the stress mediated pathways. The death-receptor pathway is marked by ligand-mediated activation of death receptors on the plasma membrane. The alternative pathway for caspase activation is mediated by cellular stress e.g., ER stress. Caspases are activated from their zymogens (procaspases), in response to various death cues. First, the initiator caspases, caspase-8 and caspase-9, are activated in response to death cues (**?**). This is followed by the activation of executioner caspases, such as caspase-3, caspase-6 and caspase-7. Activated executioner caspases proteolytically process several substrates, facilitating cell death. They also activate initiator caspases, forming a positive feedback loop. Activation of both the PERK and IRE1 pathways modulate stress-induced apoptosis through their regulation of Bcl2 expression and activity. Overall, stress induced apoptosis can occur through both mitochondrial-dependent and independent pathways. Stress signals cause oligomerization of pro-apoptotic proteins, such as Bax and Bak. These proteins are normally sequestered at the mitochondrial outer membrane by the survival protein Bcl2, under non-apoptotic conditions (**?**). Once Bax and Bak oligomerize, they insert into the mitochondrial membrane and breach membrane integrity (**?**). This results in a net efflux of cytochrome-c from the mitochondria to the cytosol and the initiation of the well-studied Apaf-1 mediated caspase-9 activation pathway. Stress induced mitochondrial-independent apoptotic pathways are not well understood. Currently, caspase 12 has been suggested as a possible ER-stress apoptotic mediator (**???**). However, caspase 12 is not expressed in human. Moreover, there is considerable debate about its role in stress-induced apoptotic cell-death (**?**).

Model Building

Estimating a population of Canonical meodels using POETs: Using the multiobjective POETs algorithm was used to generate predictive UPR model populations. Each model family was trained and validated on different experimental data. Starting from an initial best-fit initial parameter set (nominal set), more than 25,000 probable models were estimated by POETs from which we selected N = 100 models (25 from each training family) with a Pareto rank of one or less (from approximately 1200 possible choices) for further study. The nominal, training (75 models), and prediction (25 models) errors were calculated for each objective (Table **??**). Models used for prediction error calculations for a particular objective were *not* trained on that objective. The prediction likelihood was statistically significantly better for 31 of the 33 objective functions at a 95% confidence level, compared with random parameter sets generated from the nominal set (Table **??**).

Strong Pareto fronts identified in POETs suggested an inability to simultaneously model different aspects of the training data as well as experimental artifacts. Negative feedback was considered to lead to conflicting objectives. For example, XBP1 mRNA measurements (O14) conflicted with CHOP protein measurements (O13), even though these data-sets were taken from the same study and were collected in the same cell-line. XBP1 splicing increased BiP levels, which in turn reduced CHOP protein levels, hence the trade-off. Lastly, in addition to fronts, we also observed strong correlation between objectives. For example, models that performed well for the CHOP protein (O11), also performed well against Procaspase-12 (O22) measurements, even though these were not in the same cell-line or from the same study. Both CHOP and Procaspase-12 are downstream of the IRE1/TRAF2/JNK signaling cascade, so these errors were directly correlated (Fig. **??**).

Model training using data from UPR inititation events:

Signal flow, sensitivity, and robustness analysis of UPR network: Simulated KO and OX studies of key protiens provided insight into the signal flow within the UPR network. Interestingly, PERK and ATF4 KO studies revealed a slower and lower amount of

BiP production ($\sim$ 50%) as compared to WT. However, ATF6 or IRE1 KO did not affect BiP regulation as compared to WT. This highlighted the dominant role of ATF4 in regulation of BiP, which is consistent with experimental evidence **?**. Regulation of BiP was the critical regulator of spliced XBP1 (XBP1s), which in turn acts as a key marker of progression through different stages of UPR (supplementary materials Fig. S**??**E). ATF4, cleaved ATF6, and XBP1s act as integrators of the signals coming from all the three branches of UPR and furthermore leads to regulation of BiP, thereby leading to a negative feedback or control of UPR signal. Another interesting note was the regulation of pro-apoptosis phenotype via regulation of Bcl2. PERK and ATF4 KO led to delay in the onset of apoptosis (marked by slower and lower reduction of Bcl2 levels, supplementary materials Fig. S**??**F). This effect could be attributed to the lack of CHOP mediated branch of Bcl2 regulation. On the other hand, IRE1 and CHOP KO leads to drastic reduction in apoptosis (marked by little or no change of Bcl2 levels, supplementary materials Fig. S**??**F). CHOP KO implicated the importance of CHOP in the down-regulation of Bcl2. IRE1 KO implicated the critical role of IRE1-TRAF2 mediated route of apoptosis.

A few parameter sets for the sensitivity analysis were diversely selected based on the scatter in the CV values (supplementary materials Fig. S**??**). Infrastructure parameters e.g. nuclear transport, RNA polymerase or ribosome binding were globally critical, independent of stress (black points, Fig. **??**). Additionally, apoptotic species and parameters were also important, both in the presence and absence of UPR (yellow points, Fig. **??**). Thus, as expected, components such as RNA polymerase, or caspase activation were globally important irrespective of the folding state of the ER. More interesting, however, were coefficients that shifted above or below the 45^o-line in the presence of UPR. These points denote differentially important network components. While the majority of parameters and species became more important in the presence of stress, we found a band of parameters (Fig. **??** Inset) that were differentially important under stressed. For example, the rank-ordering of the sensor and stress-transducer modules clearly increased in the presence of UPR; approximately 172 or 15% of the parameters were significantly more important. These parameters were largely associated with adaptation and processing of

unfolded or misfolded proteins, e.g., unfolded protein degradation, cleaved ATF6-induced gene expression, IRE1-TRAF2 mediated apoptosis regulation, and RCAN1 regulation. Likewise, 75 or 12% of the species were significantly more important in UPR compared with normal protein loads (data not shown).

Interestingly, upon knockout of any individual feedback branch like that of ATF4, ATF6 and XBP1s, the system overall remains equally robust. However the sensitivity of the alternate feedback components increases. Overall $\sim$ 54 % of the parameters were differentially less sensitive upon removal of BiP feedback as compared to WT. This brings to light how the presence of BiP feedback makes the system more susceptible/sensitive to perturbations. The specific relevance of ATF4 in trageting BIP feedback was most evident upon KO of ATF4 feedback. We distinctly saw increase in sensitivity of feedback components associated with XBP1s and ATF6 (supplementary materials Fig. S**??**). Upon ATF6 and XBP1s feedback KO, there wasn't much change in terms of sensitivity of the system. This further attests the key regulatory effect of ATF4 in mediating the positive BiP feedback which is an essential component of the adaptation phase of UPR. Another interesting observation was that when we completely knockout all the feedback branches of BiP in the adaptation phase, the system overall becomes relatively more robust (supplementary materials Fig. S**??**). We distinctly saw a major shift of sensitivity of BiP upon removal of positive feedback. KO of ATF6 and XBP1s mediated feedback of BiP was seen to have little effect (as marked by robustness coefficients for BiP, supplementary materials Fig. S**??**). However, ATF4 mediated feedback KO led to significant amount of reduction in BiP levels (supplementary materials Fig. S**??**) thereby highlighting the significance of ATF4 in BiP feedback. Upon KO of all branches of BiP feedback, we found overall reductions of BiP levels. However, there were two distinct populations. One with a $\sim$ 10 fold reduction in BiP levels while the other had $\sim$ 1000 fold reduction in BiP levels. These two populations could resemble two distinct operational paradigms within UPR. In the first mode of operation feedback, BiP regulation is really strong resulting in drastic reductions in BiP levels and ultimately a stronger and faster UPR response upon knockout of BiP feedback.

Structural and parametric uncertainty associated with current version of UPR model

First, the cytosolic kinase domain of PERK can be inhibited by the action of the DNAJ family member P58^{IPK}. P58^{IPK} was initially discovered as an inhibitor of the eIF2α protein kinase PKR **?**. P58^{IPK}, whose expression is induced following ATF6 activation, binds to the cytosolic kinase domain of PERK, inhibiting its activity **??**. Inhibition of PERK kinase activity relieves eIF2α phosphorylation, thereby removing the translational block. Interestingly, P58^{IPK} expression occurs several hours after PERK activation and eIF2α phosphorylation. Thus, P58^{IPK} induction may mark the end of UPR adaptation, and the beginning of the alarm/apoptosis phase of the response **?**. Second, PERK induces a negative feedback loop, through its downstream effector CHOP, involving the direct de-phosphorylation of eIF2α. CHOP induces the expression of GADD34 which, in conjunction with protein phosphatase 1 (PP1), assembles into a phosphatase which dephosphorylates the S51 residue of eIF2α **?**. GADD34 is a member of the GADD family of genes which are induced by DNA damage and a variety of other cellular stresses **?**. The GADD34 binding partner in this complex appears to be responsible for PP1α recognition and targeting of the phosphatase complex to the ER. Association between GADD34 and PP1 is encoded by a C-terminal canonical PP1 binding motif, KVRF, while approximately 180 residues, near the N-terminus of GADD34, appear to be responsible for ER localization **?**. Currently, little is known about deactivation of ATF6. Recently, XBP1u, the unspliced form of XBP1, has been implicated as a negative regulator for ATF6 **?**. Following, the induction of ER stress, two versions of XBP1 exist: XBP1u and sXBP1 **?**. In the recovery phase following ER stress, high levels of XBP1u may play a dual role. First, XBP1u binds sXBP1, promoting complex degradation **??**. Second, XBP1u can bind ATF6α rendering it more prone to proteasomal degradation **?**. Taken together, these two steps may slow the transcription of ER chaperones and ERAD components during the recovery phase following ER stress. IRE1α activity is regulated by several proteins, including tyrosine phosphatase 1B (PTP-1B), ASK1-interactive protein 1 (AIP1) and members of the Bcl2 protein family. PTP-1B has been implicated in a number of IRE1α signaling events. The absence of PTP-1B reduced IRE1α dependent JNK activation, XBP1 splicing and EDEM transcription in im-

mortalized and primary mouse embryonic fibroblasts **?**. However, no physical interaction between IRE1α and PTP-1B was established. On the other hand, AIP1 physically interacts with both TRAF2 and IRE1α, suggesting a model in which AIP1 facilitates IRE1α dimerization and activation **?**. The C-terminal period-like domain (PER) of AIF1 binds the N-terminal RING finger domain of TRAF2, followed by ASK1-JNK signaling **?**. Thus, based on these findings, Luo *et al.* postulated that AIF1 may be directly involved in the IRE1α-TRAF2 complex and its activation of the ASK1-JNK signaling axis **?**. This hypothesis was validated in AIP1-KO mouse studies; AIP1-knockout mouse embryonic fibroblasts and vascular endothelial cells showed significant reductions in ER-stress induced ASK1-JNK activation that was rescued in AIP1 knock-in cells **?**. IRE1α has also been shown to directly interact with Bcl-2 family members Bax and Bak. Hetz *et al.* showed that Bax and Bak complex with the cytosolic domain of IRE1α and modulate IRE1α signaling **?**. Bax and Bak double knockout mice failed to signal through the IRE1α UPR branch following tunicamycin-induced ER stress; however, PERK signaling markers, e.g., eIF2α phosphorylation, responded normally **?**. This pro-activation role of Bak and Bax may be modulated by one of the few negative regulators of IRE1α activity, Bax inhibitor 1 (BI-1). BI-1 is an anti-apoptotic protein that enhances cell survival following several intrinsic death stimuli **?**. Bailly-Maitre *et al.* were the first to suggest that BI-1 may downregulate IRE1α and possibly ATF6 activity **?**. BI-1 deficient mice displayed increased XBP1s and enhanced JNK activity in the liver and kidney, while eIF2α phosphorylation remained normal under ER-stress conditions **?**. Lisbona *et al.* later showed that BI-1 directly interacts with the cytosolic domain of IRE1α, inhibiting its endoribonuclease activity **?**. Interestingly, BI-1 interacts with several members of the Bcl2 protein family e.g., Bcl2 and Bcl-X_L, even though it has no homology **?**. Members of the HSP family of proteins have also been shown to regulate IRE1α. For example, HSP90 interacts with the cytosolic domain of IRE1α, potentially protecting it from degradation by the proteasome **?**. HSP72 interaction with the cytosolic IRE1α domain has also recently been shown to enhance IRE1α endoribonuclease activity **?**. Taken together, these modes of IRE1α regulation with the exception of B1-1, largely promote or enhance IRE1α signaling. Given the importance of CHOP

in regulation of Bcl2, it is vital to establish the exact connectivity. However, while CHOP expression is negatively correlated with Bcl2 levels, there is no CHOP binding site in the *bcl2* promoter **?**. McCullough *et al.* have suggested that the bZIP domain of CHOP could act with other bZIP transcription factors to regulate *bcl2* expression **?**. Thus, it's likely that the connection between CHOP expression and apoptosis is more complex than simple down-regulation of Bcl2 expression. These missing structural connections shall allow us to establish a detailed model and extract more relevant insights into manipulating UPR.