Structure and Intermolecular Interactions of the Luminal Dimerization Domain of Human IRE1α*

Chuan Yin Liu, Zhaohui Xu, and Randal J. Kaufman‡

From the Howard Hughes Medical Institute, Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109-0650

The endoplasmic reticulum (ER) monitors the folding status of newly synthesized secretory and transmembrane proteins and ensures that only properly folded proteins transit to the Golgi compartment. In response to accumulation of unfolded proteins in the ER, cells activate an intracellular signal transduction cascade that culminates in the transcriptional induction of genes encoding adaptive functions. One proximal sensor for this unfolded protein response is the protein kinase/endoribonuclease IRE1α. IRE1α is a type-I transmembrane glycoprotein for which the N-terminal luminal domain (NLD) senses the accumulation of unfolded proteins. Previously we demonstrated that the NLD forms a stable ligand-independent dimer linked by disulfide bridges. In this report we have identified the cysteine residues responsible for intermolecular disulfide bonding. However, this covalent interaction was not required for dimerization and/or signaling, suggesting that a cryptic dimer interface exists in the NLD that is involved in the control of dimerization. Limited proteolysis of the NLD revealed characteristic fragments, all containing the same N-terminal sequences as full-length NLD. Biochemical and functional studies using NLD truncation mutants indicated that the dimerization domain of the NLD is confined to the conserved motifs at the N-terminal regions where putative hydrophobic interactions exist. In addition, the peptide binding domain of the endoplasmic reticulum protein chaperone BiP interacted with the N-terminal region within the NLD. Our findings suggest that the NLD has at least two distinct types of interactions mediating dimerization and function in signaling, i.e. covalent interactions involving disulfide bond formation and hydrophobic interactions, with the hydrophobic interaction being the driving force for dimerization.

The yeast UPR is a linear pathway in which the protein kinase/endoribonuclease Ire1p signaling mediates transcriptional activation of UPR target genes. When Ire1p in Saccharomyces cerevisiae is activated, it functions as a site-specific endoribonuclease (RNase) that splices HAC1 mRNA encoding Hac1p, a basic-leucine zipper-containing transcription factor. Hac1p binds to the UPR element in the promoter region and induces transcription of target genes, including KAR2, encoding the polypeptide-binding protein chaperone BiP/GRP78 that is a classical hallmark of UPR activation (1).

IRE1, PERK, and ATF6 are the three proximal ER stress transducers that regulate UPR signaling in metazoan species. IRE1 (yeast scire1p homolog) and PERK are two type-I ER transmembrane serine/threonine protein kinase receptors, and ATF6 is a type-II ER transmembrane-activating transcription factor. The mammalian UPR includes three adaptive cellular responses that are activated to cope with the accumulation of unfolded proteins in the ER; 1) transcriptional induction of ER chaperones and folding catalysts; 2) transcriptional activation of genes encoding components of ER-associated protein degradation; and 3) general translational attenuation (1–3). Recent studies at the organismal level showed that IRE1 plays critical roles in normal embryogenesis in early development, and PERK function is required for glucose homeostasis in vivo (4–7).

IRE1 and PERK are structurally similar to serine/threonine protein kinase receptors. Dimerization and trans-autophosphorylation is a universal mechanism for activation of this class of cell surface receptors (8–9). However, the biochemical and structural basis for this transmembrane signaling in response to conditions of ER stress is not understood. A biochemical and structural analysis of the NLD should provide insights into this novel transmembrane signaling event, which will in turn establish the foundation to understand the physiological functions of IRE1 and PERK.

IRE1 and PERK contain a remarkably large N-terminal domain that resides in the ER lumen. The N-terminal luminal domains (NLDs) of IRE1 and PERK sense the accumulation of unfolded proteins by a common mechanism and transmit the signal across the ER membrane to induce receptor activation (10). To provide an experimental system amenable to study the biochemical and structural basis for transmembrane signaling mediated by the NLD, the entire IRE1α luminal region was produced in a soluble form by transient DNA transfection in COS-1 cells, termed the NLD (11). The soluble NLD formed homodimers in a ligand-independent manner. In addition, the NLD interacted with the membrane-bound full-length IRE1α receptor and the ER chaperone BiP. Interestingly, the NLD homodimer was stabilized by disulfide bridges.

In this report we analyzed the biochemical and structural properties of the purified NLD homodimer. The cysteine residues responsible for intermolecular disulfide bond formation
were identified, and their requirement for UPR signaling was examined. The core domain required for dimerization was defined by limited proteolysis and analysis of truncation mutants. Our studies demonstrated that sequences required for dimerization and signaling are confined to conserved motifs at the N terminus. The existence of a cryptic dimer interface and the significance of multiple types of intermolecular interactions within the NLD dimer are discussed.

EXPERIMENTAL PROCEDURES

Materials

Proteases Glu-C, Lys-C, and trypsin, N-α-tosyl-l-lysine chloromethyl ketone, phenylmethylsulfonyl fluoride, and other protease inhibitors were purchased from Roche Applied Science. Mouse α-His antibody and Ni-NTA agarose were from Qiagen. Mouse α-NLBD antibody was previously described (12). S-protein-agarose and S-protein horseradish peroxide conjugate were purchased from Novagen. Dithiothreitol was from Calbiochem, and β-mercaptoethanol was from Sigma. All other reagents were from Sigma, Fisher, or Calbiochem.

Construction of Expression Vectors

Mammalian Expression Constructs—The mammalian expression vector of pED-NLD-His, and pED-NLD-S were described previously (11). The fragments corresponding to different C-terminal truncations were amplified and cloned into the XhoI and EcoRI sites of the pED vector. The 5′ primer used in PCR reactions contained an XhoI site and encodes MPARRLL of the N terminus. All the 3′ antisense primers included an EcoRI site at the 3′ end and encode seven specific residues corresponding to the C terminus of the truncations followed by a His6 tag and KEKDEL sequence for ER retention. The mature NLD protein included 418 amino acid residues (Ser24–Leu441). The mature C-terminal truncations are D5 (Ser24–Val390), D4 (Ser24–GLY365), D3 (Ser24–ILE331), D2 (Ser24–Val307), D1 (Ser24–ALA346), 4M (4 motifs, Ser24–Leu441), and 3M (3 motifs, Ser24–Leu441). All these NLD proteins have a sequence motif (HHHHHHHEKDEL) at the C terminus. To generate N-terminal truncation mutants, XhoI was introduced to pED-NLD-C148S/C332S by mutating CCTGAA to CCGCGG (P299P309G). PCR fragments encoding Δ2M (Ser117–Leu441), Δ3M (Leu165–Leu441), ΔD1 (ALA346–Leu441), or ΔD3 (ILE334–Leu441) were cloned into this vector to XhoI and EcoRI sites.

Yeast Expression Constructs— pRS316-IRE1, pRS316-IRE1-A (AccIII and PacI sites introduced), C193 and C35 were described previously (10). To make C6, a pair of primers encoding ISCSNS was designed. The two primers were annealed and inserted into AccIII and PacI sites in pRS316-AD. In pRS316-AD-hsIRE1-NLD, the entire yeast scIre1p-NLD was replaced with the entire NLD of hsIRE1 vector pEmcBiP was described previously (13). A sequence encoding the PacI terminus. The existence of a cryptic dimer interface and the significance of multiple types of intermolecular interactions within the NLD dimer are discussed.

Construction of Expression Vectors

Mammalian Expression Constructs—The mammalian expression vector of pED-NLD-His, and pED-NLD-S were described previously (11). The fragments corresponding to different C-terminal truncations were amplified and cloned into the XhoI and EcoRI sites of the pED vector. The 5′ primer used in PCR reactions contained an XhoI site and encodes MPARRLL of the N terminus. All the 3′ antisense primers included an EcoRI site at the 3′ end and encode seven specific residues corresponding to the C terminus of the truncations followed by a His6 tag and KEKDEL sequence for ER retention. The mature NLD protein included 418 amino acid residues (Ser24–Leu441). The mature C-terminal truncations are D5 (Ser24–Val390), D4 (Ser24–GLY365), D3 (Ser24–ILE331), D2 (Ser24–Val307), D1 (Ser24–ALA346), 4M (4 motifs, Ser24–Leu441), and 3M (3 motifs, Ser24–Leu441). All these NLD proteins have a sequence motif (HHHHHHHEKDEL) at the C terminus. To generate N-terminal truncation mutants, XhoI was introduced to pED-NLD-C148S/C332S by mutating CCTGAA to CCGCGG (P299P309G). PCR fragments encoding Δ2M (Ser117–Leu441), Δ3M (Leu165–Leu441), ΔD1 (ALA346–Leu441), or ΔD3 (ILE334–Leu441) were cloned into this vector to XhoI and EcoRI sites.

Yeast Expression Constructs— pRS316-IRE1, pRS316-IRE1-A (AccIII and PacI sites introduced), C193 and C35 were described previously (10). To make C6, a pair of primers encoding ISCSNS was designed. The two primers were annealed and inserted into AccIII and PacI sites in pRS316-AD. In pRS316-AD-hsIRE1-NLD, the entire yeast scIre1p-NLD was replaced with the entire NLD of hsIRE1 vector pEmcBiP was described previously (13). A sequence encoding the PacI terminus. The existence of a cryptic dimer interface and the significance of multiple types of intermolecular interactions within the NLD dimer are discussed.
The difference in migration between the different dimeric forms likely represents different conformations resolved by electrophoresis. Lanes 3 and 4 in Fig. 1B likely represent homodimers with Cys-332/Cys-332 and Cys-148/Cys-148 bonds, respectively. Lane 2 in Fig. 1B likely represents a mixture of these two species. Note that, unlike wild-type D5, the higher multimeric forms were not observed in the Cys mutants. Taken together, Cys-148 and Cys-332 are essential for formation of intermolecular disulfide linkages. These results support that two disulfide bonds bridge each homodimer. One occurs through Cys-148/Cys-148, and the other occurs through Cys-332/Cys-332. At this point we cannot rule out that a disulfide bridge occurs between Cys-148 and Cys-332 in some homodimers.

NLD Dimer Formation Is Independent of Covalent Disulfide Interactions—Next we directly examined the ability for these Cys mutant NLD to form dimers. Wild-type or mutant D5-H (His₆-tagged) were cotransfected with wild-type or mutant NLD-S (S-tagged). NLD-S bound to S-protein agarose (Fig. 2A, top and middle panel, pellet, lane 1), whereas wild-type and mutant D5-H did not (Fig. 2A, bottom panel, pellet, lanes 1–7). All the D5 constructs were pulled down by S-protein-agarose through their associations with NLD-S or with NLD-S-C109S/C148S/C332S (Fig. 2A, top and middle panel). In all the transfection experiments, D5 and NLD were expressed to a similar level as assessed by Coomassie Blue staining of total lysate (data not shown). The two asterisks (Fig. 2A, top, lane 2, and middle, lane 7) represent NLD and D5 heterodimer formation in the presence or absence of intermolecular disulfide bridges, respectively. This result demonstrates that NLD dimer formation is independent of disulfide bonding. However, elimination of the two disulfide bridges within the NLD did weaken the subunit association in the dimer (Fig. 2A, top, lanes 2–4, middle, lanes 3–7), suggesting that disulfide interaction within the NLD dimer contributes to affinity and/or stability. It should be noted that the intermolecular interactions between NLD and D5-C148S/C332S, between NLD and D5-C109S/C148S/C332S, and between NLD-C109S/C148S/C332S and D5 were weak, with only a fraction of D5 proteins pulled down in the assay (Fig. 2A, top, lanes 5–7, middle, lane 2). This observation indicates that dimer formation is favored when the two subunits in the dimeric complex contain either C148/C332 or S148/S332. When one subunit of the homodimer contains C148/C332 and the other contains S148/S332, the subunit association within the dimer is weaker. The basis and significance for this observation is not known and awaits structural determination.

Intermolecular Disulfide Bonding in the NLD Is Not Required for the UPR—Previously we developed an assay in S. cerevisiae to examine the function of the NLD (10). In this assay system, Δire1 yeast cells harbor a single copy of the lacZ gene under the control of the UPR element from Kar2. Activation of this β-galactosidase reporter upon tunicamycin treatment requires the introduction of a single copy of wild-type Ire1. When the NLD of yeast Ire1p (sc Ire1p) was replaced by the NLD of human IRE1α while retaining the signal peptide sequence, transmembrane sequence, and the cytoplasmic domain of the sc Ire1p, the resultant chimeric Ire1p receptor restored UPR-dependent β-galactosidase induction in Δire1. Introduction of the luminal domain of PERK also restored UPR signaling (10). Therefore, mutations in the NLD region can be introduced into chimeric Ire1p receptors, and their function can be tested by monitoring UPR-dependent β-galactosidase expression. Receptors carrying single cysteine point mutations in hs IRE1α-NLD, C109S, C148S, and C332S, were able to sense ER stress in an identical manner to wild-type NLD (Fig. 2B). The double cysteine mutant C148S/C332S and the triple mu-
The asterisks indicate two types of heterodimer-forming partners, wild-type NLD or NLD
receptors were grown to mid-log phase in media lacking uracil and treated with tunicamycin (2
minutes, cells were harvested, and extracts were prepared. Specific
expression upon ER stress. K1058A is an RNase mutant that is functional in replacing the scIRE1
major accessible cleavage sites to Glu388, Glu 377, Lys 374, 34.5, and 30.3 kDa. Based on their masses, we mapped five
Lanes 5 and 6 represent two individually isolated mutant clones of C148S/C332S. The asterisks in lane 2 and lane 7 indicate two types of heterodimer-forming partners, wild-type NLD-D5 or NLD-D5, with each monomer carrying triple Cys mutations. B and C, different chimeric Ire1p receptors carrying wild-type or mutant hsIRE1α-NLD in B and cePERK-NLD in C were tested for their ability to induce lacZ expression upon ER stress. K1058A is an RNase mutant of scIRE1α as a negative control. Cells expressing the indicated forms of receptors were grown to mid-log phase in media lacking uracil and treated with tunicamycin (2 μg/ml) for 0, 1, or 3 h as indicated. At each time point, cells were harvested, and extracts were prepared. Specific β-galactosidase activity (milliunits/mg of protein crude extract) represents an average of three independent assays with at least two independent clones. Error bars indicate S.D.

These results demonstrated that intermolecular disulfide bonds are not required for either receptor dimerization or UPR signaling upon ER stress. Our data suggest that the driving force for NLD dimerization is from interactions other than intermolecular disulfide bonds.

Proteolytic Cleavage Analysis of the NLD—To identify the molecular interactions responsible for dimerization and to understand the structural organization of the NLD, we performed limited proteolysis of the non-glycosylated NLD-N176Q. Limited proteolysis by Lys-C gave rise to one major and two minor bands (Fig. 3A, 3, 6, and 7). Proteinase Glu-C digestion also generated one major and two minor, albeit different bands (Fig. 3A, 1, 2, and 4). Trypsin digestion resulted in yet another different major band (Fig. 3A, 5) and five other minor bands. Western blotting confirmed that all these trypsinized bands originated from the NLD (Fig. 3B). Each of the three major bands and two relatively strong minor bands were excised from a polyvinylidene difluoride blot and sequenced at the N-terminal end by Edman degradation. The N-terminal amino acid sequences of all the five proteolytic fragments were STVTL-PETLL, identical to that of the full-length NLD (Fig. 3C) (11). It should be noted that there are 32 Glu and 28 Lys residues that are distributed throughout the entire NLD sequences. These results indicate that the NLD possesses relatively few sites accessible to proteases despite many potential cleavage sites.

All five bands were subjected to analysis by mass spectrometry. The determined molecular masses were 41.2, 39.9, 39.0, 34.5, and 30.3 kDa. Based on their masses, we mapped five major accessible cleavage sites to Glu377, Glu334, Lys374, Glu331, and Lys286, respectively (Fig. 3C). The fact that all the major protease accessible sites are located at the C-terminal part of the NLD indicated that the C-terminal part of the molecule is much more flexible than the N-terminal part. In addition, mass spectrometry analysis showed that the three major proteolytic fragments by Glu-C were dimers (data not
shown). Therefore, the N-terminal region is likely to contain the core structure for dimerization.

**C-terminal Truncation Mutants Can Form Dimers**—To test the hypothesis that the core structure for NLD dimerization is located at the N terminus, we constructed C-terminal truncation mutants of the NLD: D5, D4, D3, D2, D1, 4M, and 3M (Fig. 4A). Truncations were expressed in COS-1 cells by transient DNA transfection. The D5 was expressed to a level comparable to the stability of the soluble NLD. However, 2M was defective in UPR induction, suggesting that deletion of motif 3 and motif 4 abolished NLD function and Ire1p signaling (Fig. 5B) (10). It is noted that none of the N-terminal deletion mutants contains the four conserved motifs. To define the sequence requirement for NLD function, Ire1p receptors containing deletions from the C terminus of hsIRE1α-NLD were generated, and their function in UPR activation was tested using the β-galactosidase reporter assay. Like full-length hsIRE1α-NLD, truncations of D4, D2, D1, 4M, and 3M were all able to induce lacZ expression upon ER stress. However, 2M was defective in UPR induction, suggesting that deletion of motif 3 and motif 4 abolished NLD function and Ire1p signaling (Fig. 5B).

**Fig. 4.** The C-terminal truncations of the NLD can form dimers. **A**, schematic diagram of the C-terminal truncation constructs of the NLD. All the mature NLD proteins contain a Ser24 at the N terminus and an 11-residue sequence (HHHHHHHEKDEL) at the C terminus as indicated by a filled rectangle. Their relative expression in transfected cells is shown on the right. **B**, analysis of transfected cell lysates by Coomassie Blue staining showing the relative expression of D1, D2, and D3. **C**, the NLDs deleted of the C-terminal regions can form heterodimers with full-length NLD. The full-length NLD-S (S-tagged) was co-transfected into COS-1 cells with the truncation mutants, and heterodimer formation was detected by an S-protein binding assay. All the five truncations were pulled down by S-protein-agarose through their interaction with NLD-S (Fig. 4C, lanes 7–11). In the absence of NLD-S, no interacting protein was detected (Fig. 4C, lanes 2–6). Because 3M and 4M did not express well in transfected cells, their dimerization with the NLD was not examined. These results demonstrated that the N-terminal region of the NLD can form dimers.

**Fig. 5.** The three conserved motifs at the N terminus of the NLD are necessary and sufficient for UPR signaling. **A**, deletion of the conserved N-terminal motifs of the NLD from yeast scIRE1p receptor abolished the UPR. Deletion constructs are shown on the left. B, the three conserved motifs at the N terminus of hsIRE1α-NLD are sufficient and required for the UPR. Deletion constructs are shown on the left. The scIRE1p signal peptide (SP), transmembrane domain (TM), and cytoplasmic domain (CD) are indicated. The scIRE1p-NLD is shown as an open rectangle in A, and the hsIRE1α-NLD is shown as a dot-filled rectangle in B. K1058A carries a mutation in the RNase domain and is an inactive form of scIRE1p. **C**, the highly conserved residues within the four motifs of scIRE1p-NLD are not critical for NLD function. All the experiments were carried out essentially the same as in Fig. 2, B and C.
There are only about 16 identical/similar residues (~6%) among the NLDs of IRE1 and PERK including five glycines, and yet both NLDs were able to substitute for yeast Ire1p-NLD to signal the UPR. All of the highly conserved residues are localized to the four motifs at the N terminus of the NLD (10). To test the requirement of the conserved residues, they were each singly or doubly mutated to alanine in sCre1p-NLD. All of the single or double mutant receptors constructed were able to restore tunicamycin-dependent β-galactosidase induction in Δire1 cells (Fig. 5C) (10). Similarly, single mutations of the conserved residues of human IRE1a-NLD including D39N, W64A, D79N, P97A, R158E, and N176Q did not affect human NLD function in inducing the UPR when compared with wild-type human IRE1a-NLD (data not shown).

Taken together, these results demonstrated that the three conserved motifs at the N terminus are sufficient and required for dimerization and UPR induction. It is interesting to note, however, that neither of the conserved residues within this region of the NLD is critical for NLD structure and/or function.

The Peptide Binding Domain of BiP Interacts with the NLD—We asked whether BiP interacts with the NLD. BiP has two major domains, the N-terminal ATPase domain with no peptide affinity and the C-terminal PBD. The crystal structures of ATPase domain from bovine HSC70 and of PBD-peptide complex of Escherichia coli DnaK aided our understanding of the mechanism of BiP function (17–18). The ATPase domain has a four-domain structure with the nucleotide bound in a deep cleft. The PBD consists of a compact β sandwich followed by α helical elements, and the peptide substrate is bound in an extended conformation in the β sandwich. The α helical domain acts as a hinged lid to regulate the peptide binding and release in communication with the ATPase domain. We first constructed wild-type hamster BiP expression constructs, pED-BiP-S (S-tagged). IAD-S is a 27-amino acid deletion mutant of the 1A domain of the ATP binding cleft. Although IAD is capable of binding ATP and immunoglobulin heavy chain, this mutant BiP has no ATPase activity (19). 32K-S is the full-length PDB without the ATPase domain containing all the eight β-strands (β1-β8) and five α-helices (αA-αE). The 20K-S contains β1-β8 and αA, and 15K-S contains only β1-β8.

Wild-type BiP interacted with D5 and D3 by a protein binding assay using S-protein-agarose (Fig. 6, A, lane 6, and B, lane 2). Mutant IAD also interacts with D3 (Fig. 6B, lane 3), D4, and D5 (data not shown). Second, all the three PBD constructs, 32K, 20K, and 15K, bound to D5 or D3 (Fig. 6, A, lanes 7–9, and B, lanes 4–6). Third, although the interaction between BiP and D2 was detectable (Fig. 6C, lanes 1–3), the interaction was much weaker compared with that of D3 (Fig. 6C, lanes 4–6). The amounts of D2 and D3 proteins bound to Ni-NTA beads represent their expression levels in total cell lysates (Fig. 6C, bottom panel). The interaction between D1 and BiP was weak also (data not shown). We were unable to obtain 4M protein in COS-1 cells, and therefore, its interaction with BiP was not examined. Under our experimental washing conditions, all the NLD truncations from D1 to D5 did not bind to S-protein-agarose to a detectable level by either silver-staining or Western blot analysis (Fig. 6, A, lane 1, B, lane 1, and D and C, lanes 2–6). Our results showed that the PBD of BiP interacts with the N-terminal regions of the NLD. However, the Val597–Ile604 region, which is absent in D2 but present in D3, contributes to high affinity BiP binding.

Potential BiP-binding Sites in the NLDs of Human IRE1α and PERK Are Distributed Similarly—To further analyze potential BiP-binding sites within the NLD, we used an algorithm based on peptide binding motifs for DnaK identified by an extensive peptide scan (15). This algorithm allows for prediction of DnaK-binding sites with high accuracy in natural proteins by sequence alignment. The DnaK binding motif consists of 13 residues, a central hydrophobic core of 5 residues, and two 4-residue flanking regions. The nature of this consensus DnaK recognition motif is similar to the extended hydrophobic heptapeptide consensus BiP binding motif previously described (16). This algorithm is applicable to BiP also because the general features of the peptide-binding sites of HSP70 family proteins are conserved. In particular, all of the amino acids that contact the bound peptides in DnaK are conserved in BiP (18).

This algorithm calculates the statistical energy distribution (ΔΔGk) of each amino acid in the motif. ΔΔGk is a measure of the potential for a particular peptide to bind BiP. The lower the ΔΔGk value obtained for a specific segment, the higher the predicted affinity for BiP. A positive score by this scoring system indicates the peptide motif has a low probability to bind to BiP, whereas scores below −5 indicate a significant probability of binding to BiP (15). In the NLD of human IRE1α and human PERK, four motifs (M1–4) at the N terminus and three other regions (R5–7) at the C terminal were found to contain peptides with scores below or close to −5 (dotted line) and, therefore, represent potential BiP-binding sites (Fig. 7A). Consistent with this finding, these sites are also predicted to be hydrophobic in nature (data not shown). However, it remains to be determined if any of these predicted sites are in a binding-competent conformation in the folded native NLD. Sequence
The NLD Dimerizes through Hydrophobic Interactions

Fig. 7. Potential BiP-binding sites in the NLDs of human IRE1α and PERK are distributed similarly. A, multiple potential BiP-binding sites were identified by averaged energy distribution (ΔΔG) analysis. Four motifs (M1–4) and three other regions (R5–7) were found to contain peptides with scores below −5 (dotted line). Energy scores of −5 or lower possess significant probability of binding to BiP (15). The x axis represents relative residue numbers where the first residue is Ser24 for human IRE1α-NLD and Glu96 for human PERK-NLD. B, sequence alignment of conserved BiP binding regions between human IRE1α and PERK. The five-residue hydrophobic core for each potential BiP-binding site is underlined, and the ΔΔG values are shown above the core sequence for IRE1 and below the core sequence for PERK. Dots denote identical/similar residues. Dashes represent omitted sequences, and spaces represent gaps to obtain maximum sequence alignment. Arrows show the three truncations of IRE1α-NLD (D1, D2, and D3) and a deletion mutant of PERK-NLD (Δ4).

The four conserved motifs at the N terminus each contain one or more potential BiP-binding sites, and the residues of these sites are also well conserved. A number of potential BiP-binding sites are distributed throughout the non-conserved C-terminal region of the NLD. These sites can be grouped into regions 5, 6, and 7, although their positions are not well conserved (Fig. 7B). In summary, although the primary sequences of the NLDs of IRE1α and PERK have diverged, they share a common feature of hydrophobicity and multiple potential BiP-binding sites, which may provide the biochemical basis for a common mechanism of dimerization and function.

**DISCUSSION**

The NLD Dimerizes through Hydrophobic Interactions—In this study we analyzed the molecular nature by which the luminal domain of IRE1α forms dimers. Both full-length IRE1α receptor and the NLD form homodimers through intermolecular disulfide bonds (11). However, mutants carrying Cys → Ser mutations did not disrupt dimer formation and UPR induction, suggesting that a cryptic dimer interface exists that is independent of the covalent interaction. To probe the subdomain structure of the NLD, we performed limited proteolysis. The results indicated that all the major sites accessible to cleavage are located at the C-terminal region of the NLD. C-terminal-truncated forms of the NLD were functional in their ability to associate with full-length NLD and to induce the UPR in yeast. Furthermore, we showed that deletion of the conserved motifs from the N terminus abolished UPR signaling. In search for the interactions responsible for dimerization, we demonstrated that the cysteine-less NLD dimer is sensitive to SDS detergent, indicative of noncovalent interactions such as hydrophobic and electrostatic interactions. We showed that the N-terminal fragment (D1, Ser24–Ala246) formed dimers and was functional. Although the low level expression and/or instability precluded analysis of dimer status of shorter truncations, 3M (Ser20–Leu147), which constitutes the first three conserved motifs at the N terminus, was functional in sensing ER stress to signal the UPR. We conclude that the NLD dimerizes mainly through these three conserved motifs at the N terminus. Previously we demonstrated that a basic-leucine zipper dimerization motif conferred ER-stress inducible Ire1p receptor activation and UPR induction (10). This implies that dimerization and the subsequent receptor activation can occur simply through hydrophobic interactions within the leucine zipper. It is interesting to note that the conserved motifs in the NLDs of IRE1 and PERK are characterized by an abundance of hydrophobic amino acids and hydrophobic motifs. Taken together, we propose that hydrophobic interactions within the N terminus of the NLD, rather than intermolecular disulfide bonding, are the driving force for dimer formation.

The NLD Evolved as a Robust ER Stress Sensor—The ER is a unique protein-folding compartment and differs from cytosol in its high oxidizing potential, high Ca2+ concentration, and the presence of glycosylation machinery. Perturbation of this unique environment elicits the up-regulation of chaperones and folding catalysts that are present at basal levels under normal conditions. Our results support that the NLD structure is not sensitive to conditions that activate the UPR. First, because NLD dimerization does not depend upon covalent disulfide interactions, its function will not be disturbed by reductive stresses such as dithiothreitol or β-mercaptoethanol treatment. Second, although IRE1α is a glycoprotein, NLD glycosylation is not essential for dimer formation or UPR signaling. This allows the NLD to function even under conditions that disrupt oligosaccharide addition and/or removal. Third, the maintenance of the purified dimeric receptor does not require nucleotides such as Mg2+-ATP or metal ions such as Ca2+, for which the ER is the major storage site (11). Thus, the function of the NLD is not sensitive to fluctuations in ATP or calcium concentrations in the ER. Fourth, the NLD can tolerate amino acid mutations of conserved residues and even extensive deletions from the C terminus. The only region within IRE1α and PERK that is conserved, although very weakly, is at the N terminus of the NLD. A common feature between these regions is the hydrophobic character of amino acids that may provide a hydrophobic interface for dimerization. This may account for the ability for these two divergent domains to respond to the same ER stress despite their low sequence homology. Finally, the NLD displays high affinity self-association, also suggesting that a large hydrophobic dimer interface exists. Indeed, prediction of potential BiP-binding sites identified extensive hydrophobic regions in the conserved N-terminal motifs (see below). Our data demonstrate that the NLD can tolerate a local disturbance of its three-dimensional framework without disrupting the dimer interface. This ensures that the NLD maintains structural integrity and function even in the presence of mutations.
tions under conditions that reduce the fidelity of protein translation or under conditions that disrupt protein folding in the ER. We, therefore, conclude that the NLD has evolved as a robust ER stress sensor, suitable for responding to conditions that induce protein misfolding in the ER.

**Molecular Nature of UPR Regulation by BiP**—BiP is a negative regulator of the UPR and interacts with all three ER stress sensors, IRE1, PERK, and ATF6, under non-stress conditions (10–11, 20–23). Upon accumulation of unfolded proteins in the ER, these stress sensors are released from BiP to initiate downstream signaling. It is proposed that unfolded proteins bind and sequester BiP so that it is no longer available for interaction with IRE1, PERK, and ATF6. Upon release from BiP, ATF6 transits to the Golgi compartment, where it is processed to its transcriptionally active form (21). The release of IRE1 and PERK from BiP promotes their respective homodimerization and trans-autophosphorylation for activation. In this manner BiP senses unfolded proteins in the ER lumen and activates the UPR.

Although BiP is the master regulator of UPR activation, the molecular nature of its interaction with the luminal domains of the stress transducers remains to be defined. Three distinct regions within the ATF6 luminal domain were identified that have different affinities for BiP (23). A region (Lys411–Leu481) in human PERK-NLD was identified as a strong BiP-binding site (20). In the current study, we analyzed interaction of IRE1α-NLD truncations with BiP to conclude that BiP interacts with the N-terminal region of the NLD, with apparent higher affinity to fragments that contain residues Val307–Ile314. Interestingly, both the Val307–Ile314 region of IRE1α-NLD and the Δ4 region (Lys411–Leu481) of PERK-NLD are in region 6. This region in human IRE1α-NLD is not required for signaling the UPR in yeast (10). It remains to be determined whether this region in human PERK is required to confer ER stress-dependent UPR activation.

Our studies also demonstrated that the NLD binds to the eight-stranded compact β sandwich within the peptide binding domain of BiP (Asp110–Thr527). Therefore, we assume that the NLD binds to the peptide binding pocket in BiP. Our results support that there are multiple regulatory BiP-binding sites in the NLD, and these sites may be functionally redundant. BiP binding to the N-terminal region of the IRE1-NLD may be functionally significant. Our studies suggest that the Val307–Ile314 motif binds BiP with high affinity. Interestingly, proteolytic and deletion analysis showed that the Val307–Ile314 motif is not present in the dimer interface, allowing its association with BiP. It is, therefore, likely that hydrophobic motifs within the N terminus up to residue Val307 form a dimer interface, and the Val307–Ile314 motif is an exposed site available for BiP binding. To test this hypothesis, we constructed N-terminal deletion mutants of the NLD, Δ2M (Ser112–Leu441), Δ4M (Leu185–Leu441), ΔD1 (Ala246–Leu441), and ΔD3 (Ile334–Leu441). We showed that all these truncations also interacted with BiP (data not shown), although sequences in Δ4M were not required for the UPR. It remains to be examined whether these regions including the Val307–Ile314 motif are the bona fide BiP-binding sites in vivo for regulating NLD dimerization. We speculate that the weak BiP interaction with the N-terminal region, compared with the high affinity homodimer association, may be sufficient for BiP to regulate IRE1 dimerization. The great excess of endogenous BiP over the level of endogenous IRE1 provides a kinetic advantage for BiP to interact with the NLD. The high affinity self-association of the NLD would provide the driving force for receptor dimerization and activation under conditions of ER stress.

In summary, we have identified cysteine residues responsible for intermolecular disulfide bond formation. We showed both biochemically and functionally that removal of the intermolecular disulfide linkages does not affect receptor dimerization or its function in UPR signaling. The structural organization of the NLD was defined by limited proteolysis and functional analysis. We showed that the functional dimerization domain is located at the N terminus of the NLD. The characterization of the NLD permitted the production of functional truncations that are neither glycosylated nor covalently disulfide-linked. These proteins should prove useful in generating crystals for structure determination of IRE1 to elucidate how the accumulation of unfolded proteins in the ER leads to receptor activation.

**REFERENCES**

1. Patil, C., and Walter, P. (2001) *Curr. Opin. Cell Biol.* 13, 349–355
2. Mori, K. (2000) *Cell* 101, 451–454
3. Kaufman, R. J., Scheuner, D., Schroder, M., Shen, X., Lee, K., Liu, C. Y., and Arnold, S. M. (2002) *Nat. Rev. Mol. Cell Biol.* 3, 411–421
4. Uramoto, Y., Wang, X., Bertolatti, A., Zhang, Y., Chung, P., Harding, H. P., and Ron, D. (2000) *Science* 287, 664–666
5. Harding, H. P., Zeng, H., Zhang, Y., Jungries, R., Chung, P., Plesken, H., Sabatini, D. D., and Ron, D. (2001) *Mol. Cell* 7, 1153–1163
6. Scheuner, D., Song, B., McEwen, E., Liu, C., Laybutt, R., Gillespie, P., Saunders, T., Bonner-Weir, S., and Kaufman, R. J. (2001) *Mol. Cell* 7, 1165–1176
7. Lee, K., Tirastrophon, W., Shen, X., Michalak, M., Prywes, R., Okada, T., Yoshida, H., Mori, K., and Kaufman, R. J. (2002) *Genes Dev.* 16, 452–466
8. Lemmon, M. A., and Schlessinger, J. (1994) *Trends Biochem. Sci.* 19, 459–463
9. Weiss, A., and Schlessinger, J. (1996) *Cell* 84, 277–280
10. Liu, C. Y., Schroder, M., and Kaufman, R. J. (2000) *J. Biol. Chem.* 275, 24881–24885
11. Liu, C. Y., Wong, H. N., Schauerte, J. A., and Kaufman, R. J. (2002) *J. Biol. Chem.* 277, 18346–18356
12. Tirastrophon, W., Weihinda, A. A., and Kaufman, R. J. (1998) *Genes Dev.* 12, 1812–1824
13. Dorner, A. J., Wasley, L. C., and Kaufman, R. J. (1992) *EMBO J.* 11, 1563–1571
14. Morris, J. A., Dorner, A. J., Edwards, C. A., Hendershot, L. M., and Kaufman, R. J. (1997) *J. Biol. Chem.* 272, 4327–4334
15. Rudiger, S., Germeroth, L., and Schneider-Mergener, J., and Bukau, B. (2002) *Curr. Opin. Cell Biol.* 14, 431–437
16. Blond-Elguindi, S., Cwirla, S. E., Dorner, A. J., Wang, X., Bertolatti, A., Zhang, Y., Chung, P., Harding, H. P., and Ron, D. (2002) *Cell* 101, 451–454
17. Flink, C. M., DeLuca Flink, C., and McKay, D. B. (1990) *Nature* 346, 623–628
18. Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E., and Hendrickson, W. A. (1996) *Science* 273, 1606–1614
19. Gauthier, J., and Hendershot, L. M. (1993) *J. Biol. Chem.* 268, 7248–7255
20. Ma, K., Vattem, K. M., and Wek, R. C. (2002) *J. Biol. Chem.* 277, 18728–18735
21. Bertolotti, A., Zhang, Y., Hendershot, L. M., Harding, H. P., and Ron, D. (2000) *Nat. Cell Biol.* 2, 326–332
22. Okamura, K., Kimata, Y., Higashio, H., Tsuru, A., and Kohno, K. (2000) *Biochem. Biophys. Res. Commun.* 279, 445–450
23. Shen, J., Chen, X., Hendershot, L., and Prywes, R. (2002) *Dev. Cell* 3, 99–111