

# On the mechanism of sensing unfolded protein in the endoplasmic reticulum

Joel J. Credle<sup>\*†‡</sup>, Janet S. Finer-Moore<sup>†‡</sup>, Feroz R. Papa<sup>\*†§</sup>, Robert M. Stroud<sup>†¶</sup>, and Peter Walter<sup>\*†||\*\*</sup>

<sup>\*</sup>Howard Hughes Medical Institute, Departments of <sup>†</sup>Biochemistry and Biophysics and <sup>§</sup>Medicine, University of California, San Francisco, CA 94143-2200

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 29, 2003, and April 20, 2004.

Contributed by Peter Walter, November 1, 2005

**Unfolded proteins in the endoplasmic reticulum (ER) activate the ER transmembrane sensor Ire1 to trigger the unfolded protein response (UPR), a homeostatic signaling pathway that adjusts ER protein folding capacity according to need. Ire1 is a bifunctional enzyme, containing cytoplasmic kinase and RNase domains whose roles in signal transduction downstream of Ire1 are understood in some detail. By contrast, the question of how its ER-luminal domain (LD) senses unfolded proteins has remained an enigma. The 3.0-Å crystal structure and consequent structure-guided functional analyses of the conserved core region of the LD (cLD) leads us to a proposal for the mechanism of response. cLD exhibits a unique protein fold and is sufficient to control Ire1 activation by unfolded proteins. Dimerization of cLD monomers across a large interface creates a shared central groove formed by  $\alpha$ -helices that are situated on a  $\beta$ -sheet floor. This groove is reminiscent of the peptide binding domains of major histocompatibility complexes (MHCs) in its gross architecture. Conserved amino acid side chains in Ire1 that face into the groove are shown to be important for UPR activation in that their mutation reduces the response. Mutational analyses suggest that further interaction between cLD dimers is required to form higher-order oligomers necessary for UPR activation. We propose that cLD directly binds unfolded proteins, which changes the quaternary association of the monomers in the membrane plane. The changes in the ER lumen in turn position Ire1 kinase domains in the cytoplasm optimally for autophosphorylation to initiate the UPR.**

Ire1 | unfolded protein response | MHC | protein folding | secretory pathway

**M**ost secretory proteins and noncytosolic domains of transmembrane proteins are cotranslationally transported in an unfolded state into the lumen of the endoplasmic reticulum (ER), where resident enzymatic activities prevent these nascent proteins from aggregating as they fold into their native conformations (1, 2). If demand exceeds the protein folding capacity, unfolded proteins accumulate in the ER (referred to as ER stress), which activates an ER-to-nucleus signaling pathway called the unfolded protein response (UPR). In yeast, the UPR is a massive transcriptional program triggered in response to ER stress by the transmembrane sensor Ire1 (3). The UPR's transcriptional output includes a large subset of genes ( $\approx 5\%$  of the genome), including chaperones, oxido-reductases, phospholipid biosynthetic enzymes, ER-associated protein degradation components, and many proteins that function downstream in the secretory pathway (4). The aggregate effect of UPR activation, therefore, is containment and reversal of ER stress. The UPR thereby constitutes a classic homeostatic feedback loop that adjusts the ER protein folding capacity according to need.

In addition to direct transcriptional control, the UPR in higher eukaryotes is accompanied by a global dampening of translation through the phosphorylation of eIF2 $\alpha$ , while biasing any residual translation toward "translationally privileged" mRNAs vital for adaptation to ER stress (5). This translational dampening im-

poses a second level of control in parallel to the Ire1-mediated transcriptional program and is mediated by another ER transmembrane signaling protein called PERK. The concerted effect of these responses affords proteins passing through the ER an extended opportunity to fold to their native state, reduces the load on the ER, disposes of unsalvageable unfolded polypeptides through ER-associated protein degradation, and increases the capacity for ER export and downstream transport. Additionally, mammalian cells display another unique commitment step in this process: in the event that ER stress is not contained during a finite time window, the UPR directs the cell to an apoptotic pathway (6).

Ire1 is a single-spanning ER transmembrane protein with three functional domains. The most C-terminal domain of Ire1 is a regulated, site-specific endoribonuclease (RNase) that is responsible for transmitting the unfolded protein signal to the nucleus. The RNase of yeast Ire1 has a single known substrate, *HAC1<sup>u</sup>* mRNA ("u" for uninduced) (7). This mRNA encodes the Hac1 transcriptional activator necessary for activation of UPR target genes (8). *HAC1<sup>u</sup>* mRNA is constitutively transcribed but not translated. This is attributable to the presence of a nonconventional intron located toward the 3' end of the ORF, which base-pairs to the 5' untranslated region to prevent translation (9). Upon Ire1 activation through ER stress, Ire1's RNase cleaves the *HAC1<sup>u</sup>* mRNA at two specific sites, excising the intron (10). The liberated 5' and 3' exons are rejoined by tRNA ligase (11), resulting in spliced *HAC1<sup>i</sup>* mRNA, ("i" for induced). *HAC1<sup>i</sup>* mRNA lacks the translation inhibitory intron and thus is actively translated to produce the transcriptional activator Hac1, which in turn up-regulates UPR target genes (8).

A kinase domain precedes the RNase domain on the cytosolic side of the ER membrane. Activation of Ire1 changes its quaternary association in the plane of the ER membrane, resulting in transautophosphorylation by its kinase domain akin to activation of growth factor receptor tyrosine kinases in mammalian cells (12, 13). We recently showed that the Ire1 kinase additionally must bind a ligand, most probably an adenosine nucleotide, in its ATP binding site after the phosphorylation event, which evokes a conformational change that activates the Ire1 RNase (14).

Conflict of interest statement: No conflicts declared.

Freely available online through the PNAS open access option.

Abbreviations: ER, endoplasmic reticulum; UPR, unfolded protein response; LD, luminal domain; cLD, core LD; HA, hemagglutinin.

Data deposition: The coordinates of Ire1 cLD reported in this paper have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 2BE1).

<sup>†</sup>J.J.C. and J.S.F.-M. contributed equally to this work

<sup>¶</sup>R.M.S. was elected to the National Academy of Sciences in 2003.

<sup>||</sup>P.W. was elected to the National Academy of Sciences in 2004.

<sup>\*\*</sup>To whom correspondence should be addressed. E-mail: walter@cgl.ucsf.edu.

© 2005 by The National Academy of Sciences of the USA

**Table 1. Statistics of crystallographic analyses**

Data set	Ire1-LD	Hg <sup>2+</sup> -Ire1-LD	Semet-Ire1-LD	Ire1-cLD
Space group	P6 <sub>5</sub> 22	P6 <sub>5</sub> 22	P6 <sub>5</sub> 22	P6 <sub>5</sub> 22
Unit cell, a(b), c, Å	102.2, 405.2	102.5, 402.4	102.8, 402.8	102.6, 403.7
Resolution range, Å	81.2–3.10	87.42–3.17	48.22–3.49	48.39–3
High-resolution bin, Å	3.26–3.10	3.37–3.17	3.66–3.49	3.19–3
$\lambda$ , Å	1.006	1.006	0.980	1.116
Completeness, %	92.4 (96.8) <sup>†</sup>	85.7 (82.1)	95.9 (9739)	90.8 (76.3)
Average $I/\sigma(I)$	6.1 (1.1)	7.4 (1.1)	12.7 (3.8)	11.8 (1.8)
Observations	87,326 (13,143)	149,174 (18,009)	189,876 (22,951)	93,953 (7,751)
Unique reflections	21,699 (3,236)	19,179 (2,617)	16,357 (2,010)	24,150 (3,260)
$R_{\text{merge}}$ ,* %	11.7 (71.8)	9.3 (33.6)	15.6 (44)	10.6 (24.5)
Wilson B-factor, Å <sup>2</sup>	75.0	86.7	68.9	65.7
Refinement cutoff, $\sigma$	0			0
Protein atoms	4,973			4,858
Solvent atoms	26			26
$R_{\text{cryst}}$	25.6 (45.1)			24.0 (41.6)
$R_{\text{free}}$	28.4 (45.3)			27.9 (46.8)
Mean B-factor, Å <sup>2</sup>	72.6			41.4
rmsd <sub>bond</sub> , Å	0.008			0.026
rmsd <sub>angle</sub> , °	1.34			2.3
Ramachandran outliers	1			0
Most favored, %	76.6			78.8
Allowed, %	21.8			19.1
Generously allowed, %	1.5			2.1

rmsd, rms deviation; Semet, selenomethionyl protein.

\* $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$ .

<sup>†</sup>Statistics for the high-resolution bin are in parentheses.

On tracing the unfolded protein signal back to its source in the ER, an outstanding mystery remained: How does Ire1's most N-terminal domain, which resides in the ER lumen, sense unfolded proteins? The dissociation of ER chaperones from Ire1's luminal domain (LD) as they become engaged with unfolded proteins is widely held to be the mechanistic step that triggers Ire1 activation. Indeed, Ire1 activation is temporally linked to reversible dissociation from ER-luminal chaperones, most notably BiP (15, 16). However, genetic and structural evidence in direct support of the notion that Ire1-BiP dissociation is mechanistically important for Ire1 activation, and not merely correlative, has not been readily forthcoming. Here we report a structural approach to understanding the mechanism of sensing unfolded proteins by Ire1 LD.

## Results

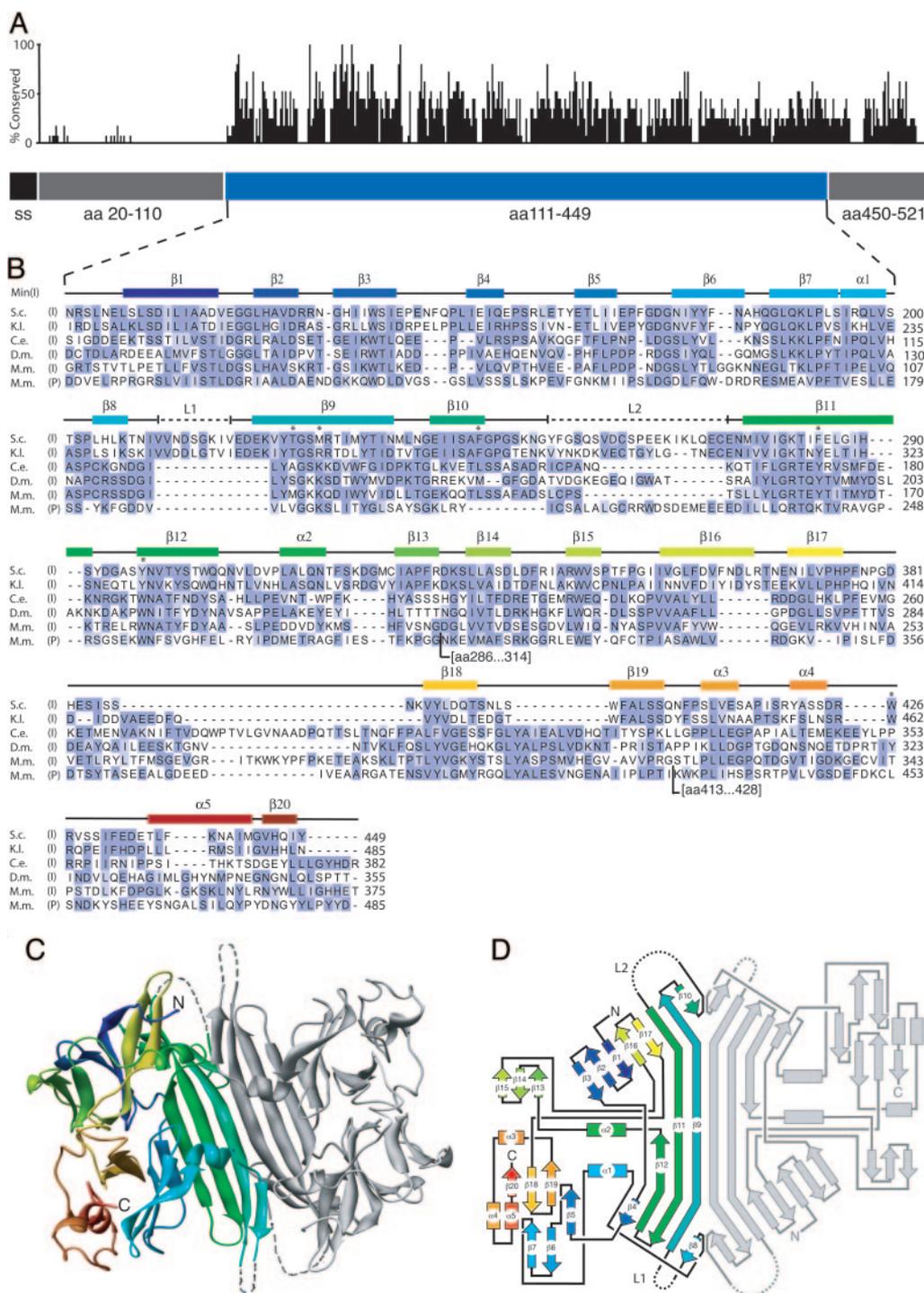
**Crystallization of the Yeast Ire1 LD.** To gain insight into the mechanism by which Ire1 recognizes unfolded proteins in the ER, we determined the crystal structure of its ER-LD. To this end, we expressed the LD of yeast Ire1 without its signal sequence ("LD"; amino acids 20–521 of Ire1) fused to an N-terminal GST tag in *Escherichia coli*. The fusion protein was soluble in *E. coli* extracts and was purified by affinity chromatography on glutathione-Sepharose. The GST moiety was removed by protease digestion (see *Materials and Methods*), and the resulting protein was further purified by ion-exchange chromatography on Mono-Q Sepharose and subjected to crystallization trials. After refining the crystallization and cryoprotectant conditions, we obtained crystals that diffract to 2.7 Å and fall in space group P6<sub>5</sub>22 with two independent LD molecules per asymmetric unit. Phases were calculated by using both isomorphous replacement and anomalous dispersion, from an Hg<sup>2+</sup> derivative and from selenomethione-substituted protein. The structure of LD was determined at 3.1-Å resolution and refined to  $R = 25.6\%$  and  $R_{\text{free}} = 28.4\%$  (Table 1).

With the exception of two segments comprising the most

N-terminal 91 aa and the most C-terminal 72 aa, and two short internal stretches (residues 210–219 and 255–274), we traced the protein sequence in well defined electron density (see Fig. 6, which is published as supporting information on the PNAS web site). The structure thus suggested that the LD folds into a compact core domain (cLD; amino acids 111–449), which is flanked by sequences that are equally disordered in both independent copies of the monomers in the crystal. The two internal disordered stretches map to surface loops in both copies of cLD. This finding suggests that all four regions of disorder are disordered in the molecules themselves and are not simply so as a consequence of crystal packing. The two independent core domains in the asymmetric unit associate in an almost perfectly twofold symmetric head-to-head arrangement (Fig. 1C).

Using the structural definition of the well structured core as a guide, we devised an expression system for the cLD (residues 111–449) that lacks the terminal regions (residues 20–110 and 450–521) that were disordered in the LD. We expressed the cLD, also as a GST-fusion protein in *E. coli*, proteolytically amputated the GST tag, and purified the protein chromatographically by methods similar to those described above for the intact LD. Purified cLD crystallized in the same unit cell (within 0.4%), and crystals diffracted to the same resolution. We determined the phases by molecular replacement using the LD structure. The resulting electron density map for cLD was virtually identical to that of full-length LD. Thus, the two separately expressed and purified proteins, one 502-aa and the other 339-aa-long, make identical crystal lattice contacts, and removal of the extra amino acid sequences in LD does not markedly improve crystal quality. However, crystals of cLD grew faster and larger. Most importantly, cLD crystals more consistently diffracted to 2.7 Å, whereas the diffraction limit of individual LD crystals was more varied, ranging from 4.0 to 2.7 Å.

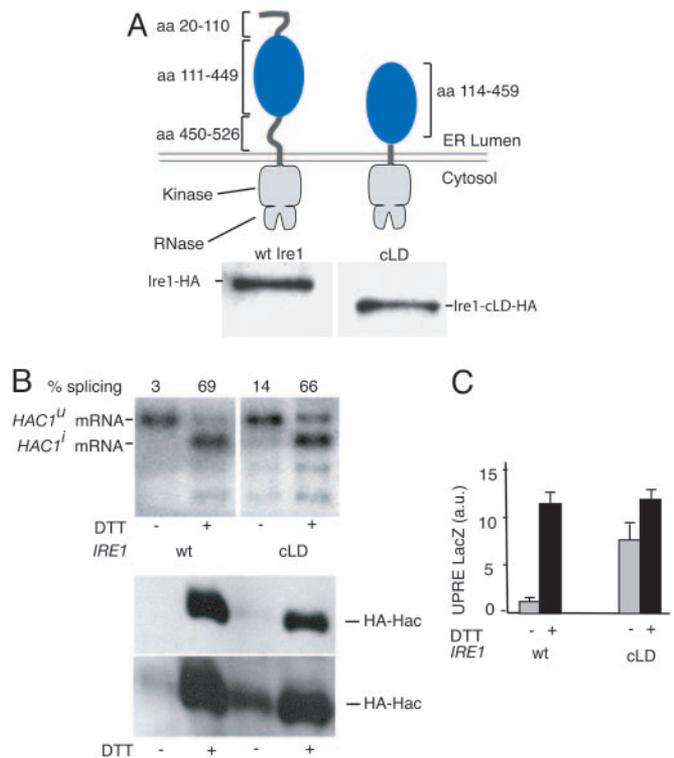
One possible artifactual explanation for the unexpected isomorphism of crystals could be that LD might have been inadvertently proteolyzed, *de facto* producing cLD during the crystallization



**Fig. 1.** The Ire1 cLD. (A) The relative conservation of amino acids is plotted along the sequence of Ire1 LD. The blue bar represents the cLD, the structure of which is shown below. The gray bars represent regions that were disordered in LD crystals and absent in cLD crystals. The black bar represents the signal sequence (ss). (B) Amino acid alignment of IRE1 and PERK LDs. (S.c., *Saccharomyces cerevisiae*; K.L., *Kluyveromyces lactis*; C.e., *Caenorhabditis elegans*; D.m., *Drosophila melanogaster*; M.m., *Mus musculus*; I, Ire1 cLD; P, PERK cLD. Conservation of residues among species was scored by using *BLOSSUM62* (46). Blue represents residues of high conservation. Secondary structural elements are indicated above the alignment and correspond in color to those of the ribbon diagram of the Ire1 cLD in C. Dashed lines (L1 and L2) represent regions found disordered in the structure. The asterisks mark residues that have been mutated in this study. For each sequence, amino acid number 1 is the initiating Met. The D.m. sequence is incorrect in the databases; an in-house resequenced sequence is used in the alignment (Julie Hollien and Jonathan Weissman, personal communication). The PERK sequence has two additional insertions (amino acids 286–314 and 413–428) where indicated. (C) Ribbon diagram of the cLD dimer as seen in the asymmetric unit corresponding to residues 111–449 have been colored with a rainbow gradient with from N terminus (blue) to C terminus (red). (D) Schematic connectivity diagram (road map) of the cLD using the same coloring scheme as in B and C.

procedure for LD, so yielding the same crystallographic unit cell. To assess this possibility, we tested the integrity of both proteins by dissolving crystals and subjecting them to SDS/PAGE followed by

silver staining, or to MALDI MS analyses. Although the mass of cLD was close to the predicted value, we found that LD was smaller than predicted. A loss of  $\approx 50$  aa from either the N or C terminus



**Fig. 2.** Functional analysis of cLD-Ire1. (A) (Upper) Topography of Ire1 and cLD-Ire1. The Ire1 cLD construct contains an ER-LD starting with amino acid 114 and ending in amino acid 449. (Lower) Immunoblot of HA-tagged Ire1 and cLD-Ire1. (B) Northern blot analysis of *HAC1* mRNA in control and DTT-treated cells expressing wild-type Ire1 or cLD-Ire1. Unspliced *HAC1*<sup>U</sup> and spliced *HAC1*<sup>I</sup> mRNAs are indicated; the lower bands are splicing intermediates. Immunoblot against Hac1-HA in control or DTT-treated cells expressing wild-type (wt) Ire1 or cLD-Ire1. A short exposure (Upper) and a long exposure (Lower) are shown. (C) LacZ activity assay in control cells (gray bars) and DTT-treated cells (black bars) expressing wild-type Ire1 or cLD-Ire1.

or both was possibly caused by cleavage at a cryptic site by thrombin used to remove the GST tag, as we observed a closely spaced doublet on SDS/PAGE under limiting digestion conditions. Nevertheless, the crystallized version of LD is larger by 13 kDa than cLD, and the mass of the additional N- and C-terminal disordered regions in the LD crystals must therefore occupy spaces in the crystal lattice that are filled with solvent in the cLD crystals.

#### The Core of the Ire1 LD Is Sufficient for Unfolded Protein Recognition.

We next tested cLD in a functional context. To this end, we engineered a mutant gene of *IRE1*, in which we replaced the LD with cLD, thus deleting both the ER-luminal sequences leading and trailing the core that were unstructured in the crystal (Fig. 2A). We expressed Ire1 cLD from its own promoter in *ire1*Δ cells. Both Ire1 cLD and wild-type Ire1 expressed equally well, as monitored by Western blotting using an antibody directed toward a hemagglutinin (HA)-epitope present at the C terminus. Cells then were treated with DTT to induce the UPR, and UPR activity was monitored by following the splicing of *HAC1* mRNA by Northern blot analysis. Cells expressing Ire1 cLD induce *HAC1* mRNA splicing to an extent (Fig. 2B, lane 4; 66% splicing) that is indistinguishable from that of control cells expressing wild-type Ire1 (Fig. 2B, lane 2; 69% splicing). Thus, cLD can efficiently activate Ire1 in response to unfolded protein accumulation in the ER.

Although fully inducible, Ire1 cLD did not appear as tightly regulated as the wild-type control, as indicated by an elevated level of *HAC1* mRNA splicing in absence of UPR-inducers (Fig.

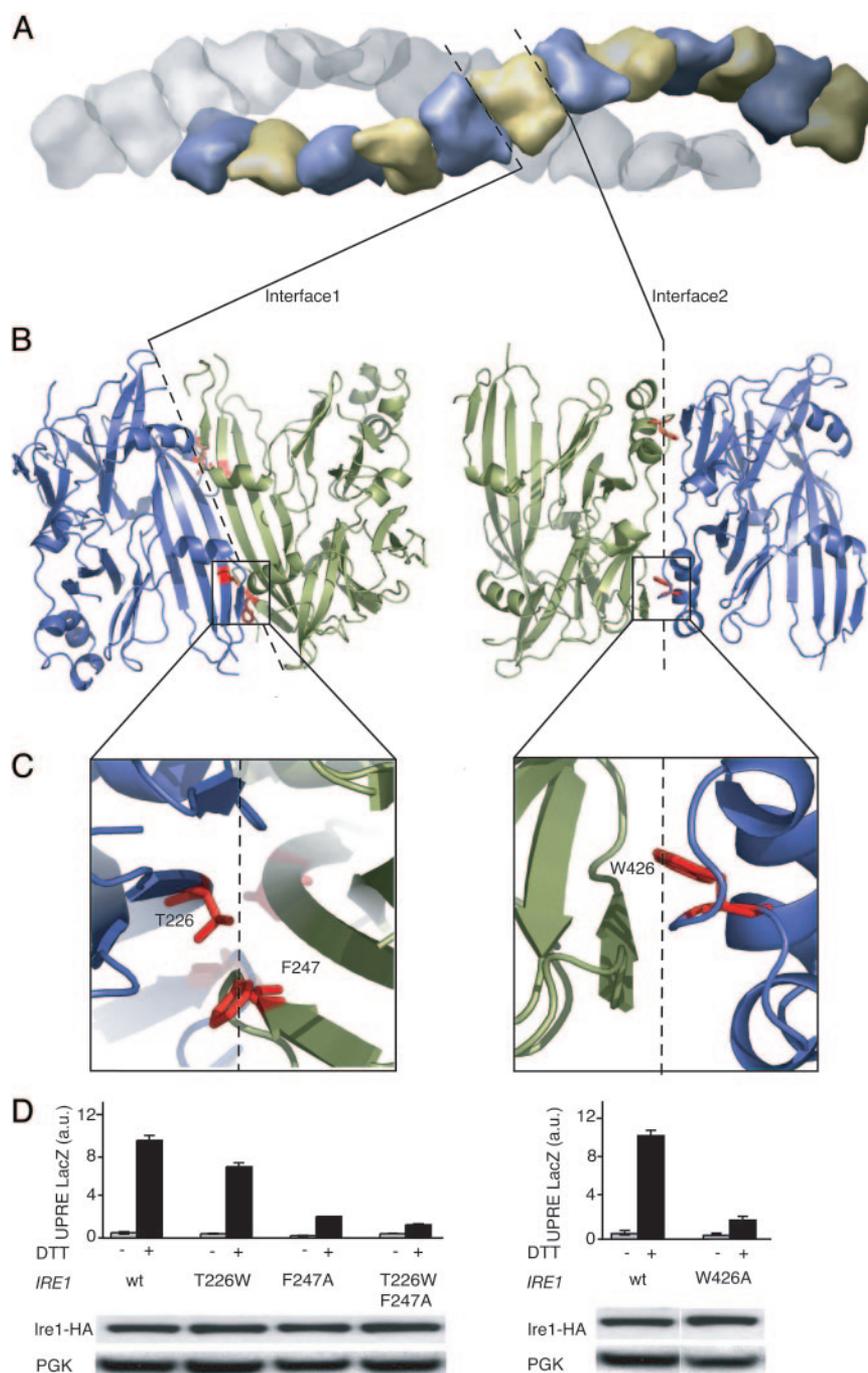
2B, compare lanes 1 and 3; 3% vs. 14% splicing). This result was confirmed by Western blotting that monitored the accumulation of Hac1<sup>I</sup>, which is exclusively produced from the spliced mRNA (Fig. 2B, Lower). To our surprise, this small amount of leakiness at the level of mRNA splicing translates into an apparently much larger “constitutive induction” when the UPR is monitored by a UPR element-driven β-galactosidase reporter (Fig. 2C, compare lanes 1 and 3). This result indicates that even a small amount of splicing can yield sufficient quantities of Hac1 to approach saturation of the transcriptional response measured with these commonly used reporter constructs. Thus, for measurements of Ire1 mutants that exhibit such “leaky” behavior, the degree of *HAC1* mRNA splicing is a more robust indicator, as it monitors the more ER-proximal reaction of the UPR signaling pathway.

**The Structure of the Ire1 cLD.** An alignment of the amino acid sequences of phylogenetically distant LD homologs shows significant conservation in the cLD region, with another block of conservation found in the disordered C-terminal region that links the cLD to Ire1’s transmembrane segment (Fig. 1A). By contrast, there is no significant sequence conservation in the N-terminal region. The cLD sequences align well with each other and with the corresponding region of the mammalian paralog PERK, with all secondary structure elements from the yeast Ire1 cLD represented (Fig. 1B). Gaps in the sequences as aligned are mostly found in loop regions, including the two unstructured loops L1 and L2. This analysis suggests that the structural features and functional insights gleaned for the yeast cLD are evolutionarily conserved among the Ire1 protein family and can be extended to include PERK, whose LD likewise responds to the accumulation of unfolded proteins in the ER.

The protein fold of the Ire1 cLD is unique and is dominated by an abundance of β-strands. A prominent antiparallel β-sheet, whose centrally located residues are exposed to solvent on either side of the sheet, links two Ire1 cLD monomers through their zippered-up central strands (β9). In each monomer, two α-helices (α1 and α2) are positioned on the β-sheet platform to form the walls of a deep groove that becomes connected to the groove in the neighboring monomer in this mode of dimeric association across the interface of the dimer. This central sheet/helix arrangement is flanked by two lobes made of short β-strands and α-helices, one lobe forming a distorted barrel (β5–7, β18–19, and α5) and the other a partial β-propeller (β1–3 and β13–15). The chain weaves back and forth between these structural features as shown schematically in the “road map” (Fig. 1D); these subdomains therefore do not represent independent folding units that are continuous in the linear protein sequence.

#### Functional Importance of Residues at both Crystallographically Defined Interfaces.

The arrangement of cLD in the crystal lattice instantly suggests biological roles (Fig. 3). Six pairs of cLD monomers line up to form one turn of a continuous helical arrangement. Two such continuous helices intertwine without any connection across the helix axis along the 6<sub>5</sub> axis of the unit cell. Monomers in each helix are arranged head-to-head forming the noncrystallographic dimeric contacts (Interface 1) shown in Fig. 1 through hydrogen-bonding of two central antiparallel strands β9, one from each monomer, and tail-to-tail around the crystallographic twofold axis (Interface 2). The ends of the central β9 strands each contribute to a large area of hydrophobic packing, formed by interaction of central strands β9 and two short-strand β8 and β10 contributed by each monomer. We refer to the contact zone resulting from the noncrystallographic head-to-head arrangement of the monomers as Interface 1. Interface 1 is essentially twofold symmetric and buries 2,380 Å<sup>2</sup> of solvent-accessible surface, contributed by 32 aa from each monomer. We refer to the crystallographic contact zone resulting from the tail-to-tail arrangement of the dimers as Interface



**Fig. 3.** Analysis of Interfaces 1 and 2. (A) Surface representation the unit cell of cLD monomers in space group  $P6_522$ . There are 24 cLD monomers per unit cell, arranged in two strands that twist around the  $6_5$  axis. Dashed lines represent the position of interfaces 1 and 2 within the strand between monomers. (B) Ribbon diagram of cLD dimers connected through Interface 1 (Left) or Interface 2 (Right). Dashed lines represent the interfaces between the twofold symmetrical dimers as seen in the asymmetric unit. The red residues shown in stick representation have been mutated. (C) Enlarged view of residues that were mutated (T226W and F247A in Interface 1 and W426A in Interface 2). (D) LacZ activity assay in control cells (gray bars) and DTT-treated cells (black bars) expressing wild-type Ire1 and Ire1 with the indicated mutations. Lower shows an immunoblot of Ire1-HA and its mutant forms.

2. Interface 2 buries  $2,117 \text{ \AA}^2$ , contributed by 26 aa from each monomer.

There is one further crystallographic contact in the lattice (Interface 3). Interface 3 is small by comparison to Interfaces 1 and 2. It links together pairs of the intertwined helices shown in Fig. 3A through contacts between the outsides of the helices and buries  $947 \text{ \AA}^2$  of solvent-accessible surface area from three molecules of Ire1 cLD where they meet. Part of Interface 3

buries a short stretch of electron density that clearly identifies a small, 8-aa-long peptide in an extended chain configuration. We observed this density in both crystals of LD and cLD. The identity of the peptide is unknown. It is not part of the LD polypeptide chain, and it is likely that this density represents a peptide (or mixture of peptides) that was carried through the purification. Although the density for the backbone and side chains is quite clear, the resolution is not good enough to deduce

**Table 2. Size determination of cLD in solution**

Protein, 10 $\mu$ M	Protein mass, kDa				
	AUC			Gel	
	Predicted	equilibrium sedimentation	AUC velocity sedimentation		DLS
cLD	39	43	34	28	34
cLD [T226W, F247A]	39	33	ND	30	44
cLD [W426A]	38	36	ND	30	35

Gel filtration was performed on a Superdex 200 10/30 gel filtration column. AUC, analytical ultracentrifugation; DLS, dynamic light scattering; ND, not detected.

its amino acid sequence unambiguously. We have not been able to identify it by MS analysis, although some low-molecular weight species are present in the spectrum.

In light of the surprisingly large areas of Interfaces 1 and 2, we assessed the oligomerization state of cLD in solution by analytical ultracentrifugation (employing both velocity and equilibrium sedimentation), gel filtration, and dynamic light scattering. These techniques yielded consistent results (Tables 2 and 3), which showed that cLD is a monomer at all concentrations up to 10  $\mu$ M and is monodisperse in solution. The expansive buried surfaces at Interfaces 1 and 2 are therefore formed during crystal growth.

At both Interfaces 1 and 2, two Ire1 LD monomers are paired with quasi-twofold, and twofold symmetry respectively (Fig. 3A). Both interfaces are extensive and continuous, suggesting that protein-protein contacts in these regions may contribute to biological function. To distinguish whether these interfaces are biologically meaningful interaction surfaces, or are simply crystal packing interactions of no biological significance, we analyzed the characteristics of the buried surfaces (17). Interfaces in true homodimers are generally more hydrophobic and contain more fully buried atoms than are crystal contacts in which the buried surface more closely resembles the protein solvent-accessible surface in character. A scatter plot of fraction of buried residues versus nonpolar surface area at 310 interfaces seen in crystal structures showed that these two parameters alone could distinguish most biological homodimers (17). For Interface 1, these two parameters cluster in a region corresponding to protein interfaces with established biological significance. By contrast, the parameters for Interface 2 fall into a gray zone corresponding to interfaces resulting from both true biological dimerization and crystal contacts. Both interfaces contain a higher than average number of residues with a propensity to lie in dimer interfaces (17).

We next tested the functional significance of residues at either interface by mutational analysis (Fig. 3B and C). To this end, we changed F247 to alanine, thereby removing a large hydrophobic side chain that becomes buried in Interface 1. The mutation impaired Ire1 function significantly as monitored by a  $\beta$ -galactosidase reporter assay, without affecting the expression level of the Ire1 mutant allele. When combined with a second mutation (T226W), designed to introduce extra bulk into Interface 1, Ire1

**Table 3. AUC equilibrium sedimentation data statistics**

Protein species	Variance of fit	Sum of residual squared	Square root of variance
Unconstrained monomer	$8.8 \times 10^{-5}$	$2.4 \times 10^{-2}$	$9.4 \times 10^{-3}$
Forced monomer	$2.2 \times 10^{-4}$	$5.9 \times 10^{-1}$	$1.5 \times 10^{-2}$
Forced dimer	$4.9 \times 10^{-4}$	$1.3 \times 10^{-2}$	$2.2 \times 10^{-2}$
Monomer/dimer	$7.6 \times 10^{-5}$	$2.1 \times 10^{-2}$	$8.7 \times 10^{-3}$

AUC, analytical ultracentrifugation.

activity was further impaired. As an additional means beyond monitoring expression levels to exclude that the mutations in Interface 1 caused deleterious folding defects, we expressed cLD harboring both T226W and F247A mutations in *E. coli*. When assayed by gel filtration, the cLD[T226W, F247A] behaved indistinguishably from wild-type cLD, arguing that the mutations did not cause the protein to be grossly misfolded (data not shown). Taken together, our data suggest that homodimerization across Interface 1 is an important step in Ire1 function.

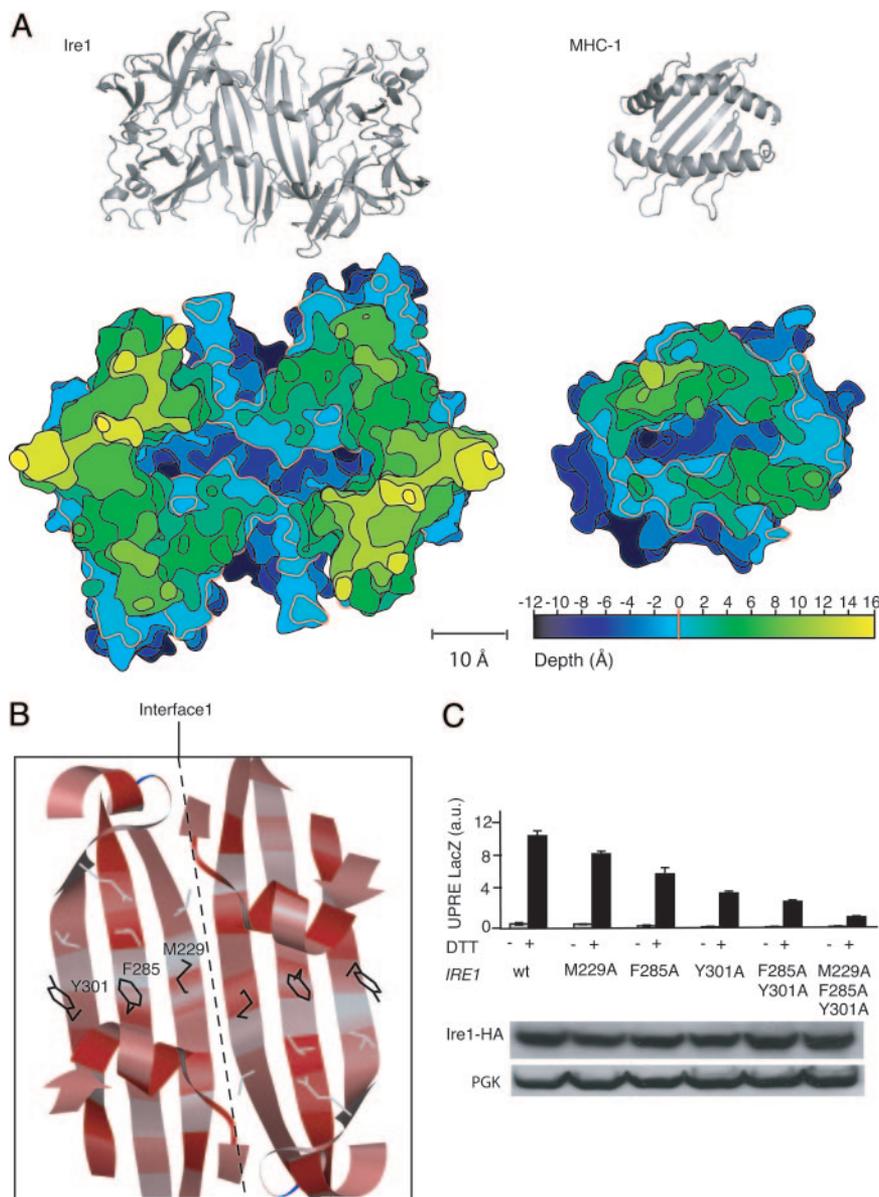
To our surprise, mutational analysis of Interface 2 yielded similar results. Removal of a single tryptophan side chain, W426 buried in Interface 2, resulted in significant loss of Ire1 activity without affecting its expression level in the cell. Again, purified recombinant cLD[W426A] behaved properly by gel filtration analysis, and the expression level of Ire1[W426A] in cells was not diminished. By contrast, mutation of many other surface residues (F174A, D176A, K190A, R196A, L204A, K223A, and F377A) that do not map to either interface had no effect on Ire1 activity, indicating that the LD is not unusually sensitive to mutational inactivation. The significant phylogenetic diversity in most surface positions also supports this notion. Thus, taken together, the mutational analyses show that residues at both Interface 1 and 2 are important for Ire1 activity (Fig. 3D). Because the two interfaces map to opposite ends of the domain, these results imply an essential role for oligomerization, rather than only dimerization, during Ire1 activation, and that the contacts in the crystal lattice provide valuable clues about physiologically important interactions.

#### Phylogenetic Conservation and Functional Importance of Residues in the Major Histocompatibility Complex (MHC)-Like Groove.

The theoretical and mutational analyses of Interface 1 suggest that LD monomers form such dimers *in vivo*. One of the most remarkable aspects of the LD homodimer formed through interactions across Interface 1 is the resemblance of its gross architectural features to the peptide-binding domain of the MHCs. Like Ire1, MHCs contain a  $\beta$ -sheet that forms a platform on which two parallel  $\alpha$ -helices are placed such that they form the walls of a deep central groove. The dimensions of the groove in LD and MHC are similar. This similarity is illustrated in the topographic map shown in Fig. 4A, which depicts a stack of sections cut parallel to the bottom of the groove of LD and 2-Å apart, alongside a similar map for a representative MHC. The map defines a rim (Fig. 4A, red contour line, set arbitrarily at 0 Å) as the height at which the rim becomes discontinuous, i.e., the lowest level where bulk solvent could access the groove. Comparison of these topographic maps shows that the width and depth of the grooves are similar. In addition to the overall geometric similarity of the two grooves, two 11-Å-deep pockets that resemble the single A or F pockets in MHC-I are found on either end of the groove in Ire1. In MHC, A and F pockets provide anchor sites for the N and C termini of bound peptides (18).

Extensive mutational analysis of LD has already been carried on throughout evolution. To explore this wealth of information, we plotted the degree of sequence conservation onto the cLD structure. In Fig. 4B, the most highly conserved residues are indicated in red, whereas less conserved residues are in light gray. A striking picture emerges: amino acid side chains that line the groove are highly conserved, whereas those facing away from the groove are much less conserved. This result is particularly apparent for residues that are part of the central  $\beta$ -sheet. In a  $\beta$ -sheet, side chains of alternating residues in each strand are exposed to the same side of the sheet. The strong conservation of the amino acid side chains that point into the groove are therefore manifest in the candy-cane striping seen in Fig. 4B.

We explored the consequences of three mutations of residues that line the groove. As shown in Fig. 4C, changing M229, F285, or Y301 individually to alanine in each case impairs but does not abolish Ire1 activity. When the mutations were combined either



**Fig. 4.** Analysis of the central groove in cLD dimers. (A) (Upper) Ribbon diagrams of the cLD dimer (Left) and MHC-1 (Right) shown in the same scale for comparison. Note that the slant of the  $\beta$ -strands is opposite between cLD and MHC. (Lower) A topographic map of cLD and MHC-1 seen from the top. The map displays the grooves as deep canyons of roughly equivalent depths and widths in the two structures. The vertical spacing of the contour lines connecting points of equal depths is 2 Å, and different elevations are colored according to the scale provided. The red index line at depth = 0 is set in both structures at the point where the rim becomes discontinuous. Relative to this contour, the grooves in both structures are 11-Å deep at their lowest point. (B) Ribbon representation looking into the cLD groove, displaying the residues mutated. The ribbon drawing is colored by amino acid conservation. Red corresponds to phylogenetically conserved amino acids. Note the “candy cane” pattern of conserved residues pointing into the groove. (C) (Upper) LacZ activity assay in control cells (gray bars) and DTT-treated cells (black bars) expressing wild-type Ire1 or Ire1 with the indicated mutations. (Lower) Immunoblot of Ire1-HA and its mutant forms.

in pairs or in triplet, Ire1 activity was significantly reduced versus the single mutations. The mutant alleles of Ire1 were expressed at wild-type levels, indicating that the mutated residues are unlikely to contribute to the ability of LD to fold into its native conformation. Taken together, these observations support that residues whose side chains point into the groove are important for Ire1 activation.

**Discussion**

The structure of yeast Ire1 cLD and the functional studies presented here provide a number of unexpected insights into the molecular mechanism by which Ire1 recognizes and responds to the accumulation of unfolded proteins in the ER lumen. First,

our studies define the compactly folded and phylogenetically conserved cLD as the functional center of the LD. Regions preceding and trailing it in the sequence are largely dispensable for regulation of Ire1, suggesting that cLD contains all necessary elements to sense unfolded proteins and transmit this information across the membrane. This result is in agreement with previous mutagenesis studies, which showed that segments can be deleted from either end of the LD without affecting its function (19, 20).

Second, the crystal structure of cLD defines two regions of extensive contacts at opposing ends of the monomeric cLD. Surprisingly, mutations at both of these interfaces impair Ire1 activation, suggesting that residues at both interfaces are bio-

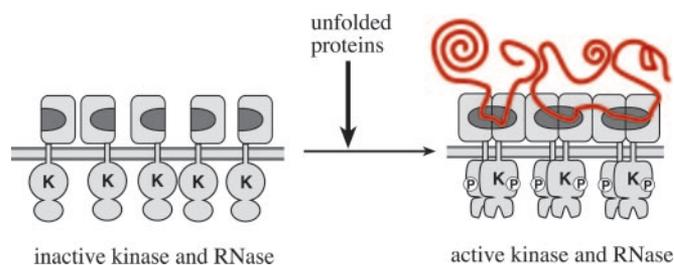
logically important. A direct implication of this result is that dimerization at either interface is insufficient for activation. Thus, it is likely that the formation of higher-order linear oligomers is important for Ire1 activation.

Third, Ire1 dimers contain a deep groove resembling in its geometry that seen in the MHCs. Conserved amino acid side chains line the bottom and walls of the groove, suggesting that the binding of unfolded polypeptide chains may be encoded there.

How unfolded proteins in the ER lumen are recognized has been a subject of debate. Before this work, the prevailing model posed that the chaperone BiP binds to the LD of Ire1 where it acts as a negative regulator, thus preventing Ire1 activation (16). According to this view, the equilibrium  $\text{Ire1} \cdot \text{BiP} \leftrightarrow \text{Ire1} + \text{BiP}$  regulates Ire1 activity: if free BiP is in sufficient supply, the equilibrium lies to the left and Ire1 is off, whereas when free BiP levels fall because BiP becomes engaged with unfolded proteins, the equilibrium shifts to the right, and Ire1 is turned on as BiP is titrated away from it. Although there is convincing correlative evidence for BiP binding to inactive but not active Ire1 (and PERK) (16, 20), no causality between release of BiP and Ire1 activation has yet been established. One of the main drawbacks of this model is that BiP is present in the ER lumen in millimolar concentrations. Activation of Ire1 thus would require large concentrations of unfolded proteins to provide a sufficiently large sink to make a significant difference in the pool of free BiP. This is clearly not the case because the UPR responds to small fluctuations in the ER protein folding state, as would seem appropriate for a sensor that adjusts the ER protein folding capacity of cells according to need homeostatically. Moreover, in a recent study, Oikawa *et al.* (21) identified the BiP binding site in the Ire1 LD to lie within amino acids 448–520 in yeast Ire1 and showed that deletion of this region did not impair the Ire1 regulation by presence and absence of unfolded protein. Direct recognition of unfolded proteins by the cLD suggested by our results provides an attractive alternative model.

Although BiP binding and release is not a requirement for control of Ire1 activity, it might nevertheless provide a regulatory role under extreme activation conditions when the pool of free BiP becomes severely depleted. Such situations might arise under nonphysiological experimental conditions using high concentrations of tunicamycin or DTT, or upon prolonged UPR induction. BiP release under such conditions could serve to enter a different activation state, perhaps signaling that the UPR is not able to reestablish homeostasis in the ER and leading the cell down an apoptotic pathway. Conversely, BiP binding may dampen activation of Ire1 under conditions of mild unfolded protein accumulation (i.e., during conditions that may be dealt with through existing concentrations of ER chaperones), consistent with our observation that Ire1 cLD lacking the BiP binding site is mildly constitutively induced. In this view, BiP binding would buffer Ire1 against normal fluctuations of ER unfolded proteins, thereby reducing “noise” in UPR signaling. Regardless of the precise contribution of Ire1-BiP binding to UPR activation, our results suggest that it is of secondary importance for UPR signaling by Ire1.

Ire1 cLD's resemblance to MHC is likely to be an example of convergent evolution: it reflects a common architectural design principle rather than a common evolutionary origin. Both classes of molecules use a flat  $\beta$ -sheet as a platform on which  $\alpha$ -helices form the walls of a deep groove. The construction differs greatly in detail, however, which is most obvious from the opposite chirality with which the  $\beta$ -sheet and the  $\alpha$ -helices intersect (Fig. 4A, gray ribbon models). If, as postulated here, the groove indeed serves to bind portions of unfolded polypeptide chain, then evolution converged on this common feature by different routes. The groove of cLD is lined by an about even mixture of hydrophobic and polar amino acid side chains. We have shown



**Fig. 5.** Model for unfolded protein recognition by Ire1. The model depicts Ire1 activation through oligomerization brought about by binding of unfolded proteins (indicated in red). Direct or indirect interactions between unfolded protein chains may contribute to activation. On the ER-luminal side of the membrane, the postulated unfolded protein-binding groove formed by Ire1 cLD dimerization through Interface 1 is indicated in dark gray. On the cytoplasmic side of the ER membrane, oligomerization juxtaposes the Ire1 kinase domains, which undergo a conformational change after autophosphorylation that activates the RNase function of Ire1. Inactive Ire1 could either be monomeric as shown or exist already in oligomeric yet inactive states whose quaternary associations change upon unfolded protein binding.

here that amino acid side chains exposed at the floor of the groove matter for Ire1 activation. There are additional features that MHC-type molecules and Ire1 cLD potentially have in common, such as conserved arginine side chains (R196 in Ire1 LD) that extend from the walls of the groove and may gate access (22, 23). Both phylogenetic conservation of the groove-lining residues and the functional importance of the ones tested here suggest that a ligand binds there, which, if it is a protein, must be conformationally predisposed to reach deep into the groove. By modeling, a linear 10-polyvaline peptide can be accommodated within the groove without steric clashes, while allowing ample room at either end for the peptide to loop out of the groove. Thus, we hypothesize that unfolded polypeptide chains and possibly partially folded proteins with exposed loops on their surface bind to Ire1 directly, providing the primary signal mediating its activation (Fig. 5).

In principle, there are two extreme but not mutually exclusive ways by which unfolded proteins could be recognized. First, the groove could provide a binding environment that is specific for certain amino acid side chains. This principle is realized in hsp70-type chaperones, such as BiP, where a signature binding motif on the unfolded substrate that consists of hydrophobic amino acids in every other position has been characterized (24). Such a sequence resembles a  $\beta$ -strand, one side of which is destined to pack onto the hydrophobic core of a folded protein but has not yet been properly accommodated in the protein fold. Hence, although low in information content, the sequence properties of the unfolded protein provide an intuitive means of recognizing unfolded proteins that need chaperone assistance. By contrast, the peptide binding grooves of MHCs bind peptides with high sequence specificity (25). Genetic variation allows different MHC subclasses and alleles to construct binding pockets with different specificities, which is an important feature of immune surveillance. Structurally, the groove in Ire1 resembles those found in MHCs in that it is similarly lined by a patchwork of conserved hydrophobic and hydrophilic residues. Thus, recognition of specific side chains or classes of side chains in preferred positions could play an important part in unfolded protein recognition.

A second way of achieving recognition of unfolded proteins is steric discrimination. Given the depth of the groove, it is inaccessible to interact with surface residues on compactly folded proteins. Some viruses use a similar strategy of hiding residues that are important for infectivity in deep canyons on their surface where the antigen binding sites of globular immunoglobulins cannot reach (26). In principle, steric discrimination

alone could be sufficient to distinguish folded from unfolded proteins. Interactions in the groove might therefore be limited to backbone contacts only, paying little or no attention to amino acid sequence of the polypeptide. Both steric discrimination and sequence specificity might be important parameters in recognition of the unfolded protein. Studies designed to assess the binding properties of cLD and peptides or unfolded polypeptide chains are required to distinguish between the relative importance of these possibilities.

Previous work showed that Ire1 activation results in formation of large oligomers. In particular, we demonstrated that upon activation, full length Ire1 coimmunoprecipitated with significantly more than equimolar amounts of a C-terminally truncated version, if both Ire1 constructs were expressed in the same cell (27). Here, we show that residues at Interface 1 and Interface 2 are important for activation, suggesting that both regions participate in oligomerization. Because both regions are found on opposing ends of the monomer, the arrangement would be linear, and it is quite plausible that the filamentous assembly formed in the crystal lattice may represent a view of biologically relevant interfaces, albeit in a possibly distorted way. We confirmed with a variety of techniques that the solution state of cLD is monomeric. Thus, inactive Ire1 may also be a monomer as depicted in Fig. 5. An unfolded polypeptide chain might favor dimerization by binding to the groove as it is formed across the dimer, mediated by Interface 1. It is not clear whether dimer formation alone is enough to lead to even partial Ire1 activation. Indeed, the Interface 1 and Interface 2 mutants analyzed here retained some activity (Fig. 3), and Liu *et al.* (28) previously showed that Ire1 dimerization in constructs in which the Ire1 LD was replaced by an inherently dimerizing leucine zipper led to partial activation. We show here, however, that residues at both interfaces are important, strongly suggesting that a higher-order quaternary structure is required for full activation. Thus, we propose that the unfolded polypeptide chain helps form and then tethers Ire1 dimers, and that it is the further association of dimers through Interface 2 that properly juxtaposes Ire1 kinase domains on the other side of the membrane, resulting in efficient Ire1 activation.

An alternative view to this two-step activation process is that full-length, inactive Ire1 may already be dimerized or be in some other associated but inactive state in the plane of the membrane. The affinity of receptors for each other contributed by interactions between the cytosolic and/or transmembrane domains that by themselves are insufficient for activation might locally concentrate LDs in the ER lumen and engender interactions through either Interface 1 or Interface 2 despite an immeasurably low affinity between the cLDs in solution. As above, it would be the association of dimers in a particular orientation, triggered by the added stability provided by binding unfolded polypeptide chains that leads to activation. In the cytokine receptor systems it is not simply dimerization but also orientation of the receptors within the dimer that is important. In Ire1, the importance of orientation of Ire1 versus simple dimerization for signaling remains to be determined.

Irrespective of which of the proposed models proves to be correct, the mechanistic insights suggest intriguing ways to design Ire1 inhibitors that could prove to be powerful pharmaceuticals. A steadily increasing number of publications point to roles for the UPR in a variety of diseases. Enveloped viruses, for example, exploit the UPR to make more ER so that the cell can handle the load of viral membrane protein production (29). Similarly, rapidly growing cancer cells rely on the UPR for their survival (30). If a compound could be developed that binds to and occupies the half-groove in a cLD monomer, or otherwise prevents further oligomerization of Ire1 receptors, it might serve as an antagonist of Ire1 activation. Such a reagent might constitute proof of principle that inhibitors could potentially lead

to the development of broad-spectrum antiviral or chemotherapeutic cancer agents.

## Materials and Methods

**Strains and Plasmids.** The N terminus of yeast Ire1 from residues 20–521 was cloned (Primer 1, 5'-GGGGGATCCTCCATCATTTCATGCTC-3'; Primer 2, 5'-GCTCTCTTAATCTACTTATTGAGCTCGGGG-3') into a BamHI–EcoRI linearized pGEX4T-2 expression vector (Amersham Pharmacia Biosciences) containing an N-terminal GST tag (PW420). The *ire1Δ* strain PWY260 (*ire1Δ::TRP1*; *his3-11,-15::HIS<sup>+</sup>UPRE-lacZ*; *leu2-3,-112::LEU<sup>2+</sup>UPRE-lacZ;ura3-1*), derived from W303 (R. Rothstein, Columbia University, New York), contained two integrated copies of lacZ, encoding β-galactosidase, under control of the UPR element (UPRE1) as described previously (8). IRE1 and IRE1 mutants used in this study were expressed from a CEN-ARS low-copy yeast-shuttle vector (derived from Yc-plac33) transformed into PWY260 by using LiOAC. All mutations were confirmed by DNA sequencing on a 3100-Avant DNA Sequencer (Applied Biosystems Appliedera).

The *IRE1* gene used in all shuttle vectors was contained on a 5 kb XhoI–HindIII genomic fragment maintained in the CEN-ARS low-copy yeast-shuttle vectors YCplac33. In this plasmid, a single HA epitope was incorporated at the 3' end of *IRE1* as previously described (14). The HA-tagged variant exhibited wild-type activity and served as the parent plasmid for all subsequent mutagenesis. Introduction of missense mutations used the QuickChange XL kit (Stratagene). The gene encoding Ire1 cLD was constructed as follows. A cLD-encoding PCR fragment was amplified from a wild-type IRE1 template with forward and reverse primers that had deletions encoded into their sequence. The forward primer had the sequence 5'-TCCATCATTTTCATGCTCAATCCCATTGTCTGCTCTCGCACCTCATTGAACGAAGTGTATCAG-3' (the coding sequence resulted in a deletion of residues 34 to 113, i.e., the N-terminal fusion occurred after the signal peptide sequence and appended the first 33 residues to Leu-114). Similarly the reverse primer had the sequence 5'-TAGACTTCCAAACTTCAGTAGCAAAGAATTTTGGTTCTTGTTTTTCATAAAGGTGATCATATTC-3' (the coding sequence—complementary strand—appends N459 to K521; therefore, residues 460–520 are deleted, and the transmembrane region is completely preserved). This PCR fragment was used to transform strain PWY260 along with the parent plasmid, which had been linearized with PflM1 and Afe1. Uracil prototrophs were selected on synthetic defined (SD) *ura<sup>-</sup>* media, and gap-repaired plasmid was recovered in *E. coli* DH5α. The plasmid was sequenced at both ends to confirm deletion of the intended sequences. Anti-HA Western blot confirmed that the size of the protein matched the expected size and that levels of cLD protein approximated those of wild type.

**Protein Expression and Purification.** Plasmids encoding GST-tagged fusion proteins (GST-LD or GST-cLD) were transformed into *E. coli* BL21-DE3\* and grown in LB-ampicillin until mid-log phase ( $OD_{600} = 0.7$ ). Culture were induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3.5 h and then incubated at 4°C overnight. The next day, cells were pelleted by centrifugation and resuspended in cold lysis buffer (50 mM Tris, pH 7.5/0.5 mM EDTA/5 mM DTT) containing Complete Protease Inhibitors (Roche). After resuspension, cells were lysed by three passages through a MicroFluidizer (Microfluidics Corp.). The cell lysate was centrifuged for 1 h at 4°C and 35,000 rpm in a Ti45 rotor (Beckman). The cleared supernatant was bound in batch to glutathione-Sepharose (Amersham Pharmacia Biosciences) at a concentration of 1 ml of resin per liter of lysate. Binding was performed in batch at 4°C for 2 h with gentle agitation. The beads were collected by centrifugation and

washed three times with cold PBS, pH 7.4. The recombinant LD was removed from the bound GST tag by incubating the 5.0 ml of beads with 7.0 ml of thrombin (Amersham Pharmacia Biosciences) at 32.5 units/ml of beads at 4°C for 15 h. The eluate was dialyzed into buffer A (20 mM bis-Tris, pH 6.0/2 mM  $\beta$ -octyl glucoside/5 mM DTT) and loaded onto a MonoQ 10/10 ion-exchange column (Amersham Pharmacia Biosciences) preequilibrated in buffer A. The column was washed twice with 2 column volumes of buffer A and eluted with a linear gradient of 10 column volumes of buffer A  $\rightarrow$  buffer B (buffer A containing 2 M NaCl). Peak fractions were collected, checked for purity and correct size by SDS/PAGE (12.5% NuPage gel, Invitrogen), and pooled. The purified recombinant LD was dialyzed overnight at 4°C into buffer C (20 mM Tris, pH 7.5/0.5 mM DTT/0.01 mM PMSF/2 mM  $\beta$ -octyl glucoside) and concentrated to 5.0 mg/ml by using a 30-kDa cutoff Centricon Concentrator (Amicon). cLD was purified in an identical manner, except that after chromatography on the MonoQ column, peak fractions were pooled, concentrated, and further purified on a Superdex 200 gel filtration column (Amersham Pharmacia Biosciences). Peak fractions were pooled and dialyzed overnight at 4°C into the minimal crystallization buffer and concentrated to a final protein concentration of 5 mg/ml. For selenomethionyl protein production, the Met auxotroph *E. coli* strain B834 was used and grown as previously described (31). Purification of selenomethionyl LD was performed as described above except for the addition of 5 mM methionine to the purification buffers as a supplemental reducing agent.

**Crystallization.** Crystals of the LD (residues 20–521), and the cLD (residues 111–449) were grown by hanging-drop vapor diffusion at 4.0°C by mixing 2  $\mu$ l of recombinant protein (5 mg/ml) with 2  $\mu$ l of the reservoir buffer (1.6 M ammonium sulfate/50 mM bis-Tris, pH 6.5/5% dioxane/0.6% MeOH). Hexagonal crystals appeared within 2 days and grew to their maximum size in 2 weeks. Crystals were cryopreserved by soaking them in the original reservoir buffer with added glycerol (17%) at room temperature for 30 s and flash-frozen in liquid N<sub>2</sub>. The Hg<sup>2+</sup> derivative was prepared by soaking crystals in a 10:1 mixture of mother liquor with a 100 mM HgCl<sub>2</sub> solution for 5 min. Crystals were then back-soaked in the cryoprotectant solution and flash-frozen. Crystals of selenomethionyl LD were grown as described above except for the addition of 5 mM methionine to the crystallization solution as a supplemental reducing agent.

**X-Ray Diffraction.** X-ray diffraction data were collected at  $-170^{\circ}\text{C}$  on a QUANTUM image plate at beamline 8.3.1 at the Advanced Light Source at Lawrence Berkeley National Laboratories (Berkeley, CA). Native cLD and selenomethionyl Ire1 LD data sets were processed and scaled with the HKL processing software (32) (Table 1). Structure factor amplitudes were derived from intensities by using TRUNCATE (CCP4 program suite). Native LD data and the Hg<sup>2+</sup> derivative data set were reduced by using MOSFLM as implemented in ELVES (33) (Table 1).

**Structure Solution and Refinement.** X-ray data from a single crystal of a mercury derivative revealed two mercury sites that were used to determine phases to 3.2 Å for the LD structure by the SIRAS method, as implemented in MLPHARE (CCP4 program suite). No minor heavy atom sites could be identified in double-difference Fourier maps calculated with the SIRAS phases. The electron density map calculated with these phases showed a clear protein-solvent boundary consistent with one dimer of Ire1 LD in the asymmetric unit and several sets of secondary structure elements that were related to each other by a noncrystallographic twofold axis. Phases were improved by density modification, histogram matching, and noncrystallographic symmetry averaging by using DM (34). Approximately 60% of the residues

of one quasi-twofold symmetric dimer of the LD were initially fitted as serines to the experimental density by using CHAIN (35).

The x-ray data from selenomethionine-substituted Ire1 LD were not initially useful for phasing, because the selenium sites could not be located by inspection of difference Patterson maps or by the program SOLVE (36). However, once a partial model of the protein had been built, a difference Fourier calculated with data from native and selenomethionyl protein with partial model phases [(F<sub>SeMet</sub> - F<sub>Nat</sub>)  $\alpha$  calc.] showed the locations of 6 of the 20 selenium atoms (10 selenomethionines per monomer) in the asymmetric unit. MIRAS phases calculated from the Hg<sup>2+</sup> and selenomethionyl Ire1 LD sets, improved and extended by density modification (solvent flipping) using SHARP (37) (Table 1) gave the experimental map with the greatest connectivity (Fig. 6), which was used for determining connectivity and assigning sequence. The sequence assignment was later confirmed by calculating an updated [(F<sub>SeMet</sub> - F<sub>Nat</sub>)  $\alpha$  calc.] map with phases from the refined structure. This map had density peaks greater than 3.5 $\sigma$  at all but one of the 12 methionine side chains in the ordered region.

The N-terminal residues 20–110, C-terminal residues 450–521, and loops 210–219 and 255–274 were not visible in electron density maps. Residues 111–209, 219–254, and 275–449 in each protomer were refined at 2.8 Å to  $R = 23.7\%$  and  $R_{\text{free}} = 27.6\%$  with combinations of simulated annealing and positional and restrained isotropic B-factor refinements, with noncrystallographic symmetry restraints, a bulk solvent correction, and an anisotropic B-factor correction by using CNS and REFMAC5 (38–40). The data were anisotropic and diminished faster versus resolution in the  $a$ ,  $b$  plane. Therefore, we report statistics only to 3.1 Å (Table 1). An 8-residue stretch of polypeptide chain of unknown origin was visible in strong electron density at a crystallographic interface of one protomer with another. There was no noncrystallographic symmetry-related density for these residues. Although the residues were well ordered, the resolution of the maps precluded sequence assignment, thus it was not known whether the residues comprised a cocrystallized peptide or were part of the otherwise disordered LD chain. The peptide was built and carried through refinement by using valine for all residues and is represented in the Protein Data Bank as chain D.

The crystal structure of the Ire1 cLD, residues 111–449, was isomorphous to that of the full-length LD. The structure factor amplitudes for the cLD scaled to those of the LD with an overall  $R$  factor of 17.2%, higher than could be accounted for by random errors in amplitudes ( $R_{\text{m}} \approx 10\%$ ) (Table 1). However, the weighted  $R$  factor (12.6% overall) was fairly constant versus resolution, suggesting that the disordered regions are partially ordered structural domains that are flexibly tethered to cLD. A difference map calculated with coefficient (F<sub>o1</sub> - F<sub>o2</sub>)  $\alpha$  calc., where F<sub>o1</sub> was an amplitude from Ire1 LD data and F<sub>o2</sub> was the corresponding amplitude from the data from Ire1 cLD, indicated no significant differences between the two structures, even at their N and C termini. In particular, the crystal structure of the truncated domain contained the same isolated 8-residue peptide at the crystal interface as was seen in the crystal structure of the full-length domain. Because both disordered loops (210–219 and 255–274) are too far from the position of the peptide, and all other residues are accounted for in the structure, the peptide is likely to have copurified with Ire1 LD and Ire1 cLD from the cell lysate. This is particularly unusual because the protein was purified as a monomeric species, and the peptide is seen only in a crystal interface between three molecules. The model for Ire1 LD, omitting highly mobile residues 379–386, refined against the Ire1 cLD 3Å data, yielded  $R = 24.0\%$  and  $R_{\text{free}} = 27.9\%$  (Table 1). A composite simulated annealing (2F<sub>o</sub> - F<sub>c</sub>) omit map for Ire1 cLD showed density for the majority of the residues in the model, although density for a few of the more mobile loops was weak or absent in the omit map.

**Structure Analysis.** Surface area buried by protein–protein interfaces was calculated in CNS by the method of Lee and Richards (41). Structure quality was assessed with PROCHECK (42). Figures were produced by using the University of California, San Francisco, CHIMERA package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (43), and PYMOL (Delano Scientific).

**Enzyme Assays.**  $\beta$ -Galactosidase assays were carried out as previously described (8). In brief, yeast cultures were grown at 30°C to  $OD_{600} = 0.7$  in synthetic defined (SD) media lacking uracil supplemented with 100  $\mu$ g/ml of inositol. Freshly prepared DTT buffered with NaOAc, pH 5.2, was added to the culture to a final concentration of 2 mM to induce the UPR. Cells were harvested after 45 min by centrifugation and assayed as previously described (14) using 0.8 mg/ml *o*-nitrophenyl  $\beta$ -D-galactoside (ONPG) (Sigma) according to the manufacturer's instructions. The reaction was incubated at 32°C for 10 min, at which point 400  $\mu$ l of 1 M  $Na_2CO_3$  was added to quench the reaction. LacZ arbitrary units (a.u.) are defined as  $(OD_{420} \times 1,000)/(OD_{600} \times t \times v)$ , where  $v$  is the volume of the sample used in the assay and  $t$  is the time of the incubation of the reaction at 32°C. Values for LacZ a.u. were expressed as the mean  $\pm$  SD measured for three independent transformants for each condition.

**Northern and Western Blot Analyses.** Northern blot analysis to detect unspliced *HAC1* mRNA (*HAC1<sup>u</sup>* mRNA) and spliced *HAC1* mRNA (*HAC1<sup>i</sup>* mRNA) was performed as described (9). For detection, a  $^{32}P$ -labeled 5' exon probe of *HAC1* mRNA was made by using the Ready-To-Go DNA Label kit (Amersham Pharmacia Biosciences). Yeast protein sample preparation and Western blot analysis to detect Hac1 and Ire1 has been described (14). Anti-HA horseradish peroxidase conjugate (at a 1:2,000 dilution) and SuperSignal ECL (Pierce) were used to detect wild-type and mutant forms of Ire1-HA.

**Dynamic Light Scattering (DLS).** Molecular weights and dispersity of the cLD and Interface 1 and Interface 2 mutant cLDs were determined by DLS at room temperature using a DynaPro MS/X and accompanying analysis software (Proterion, Piscataway, NJ). This analysis yielded the  $R_h$  and dispersity of the sample. Solvent viscosity and refractive index were calculated accordingly. Lysozyme was used to calibrate the DLS instrument before and after all runs. All protein solutions were tested at 10  $\mu$ M.

**Analytical Ultracentrifugation Velocity Equilibrium and Sedimentation.** All experiments were performed on a Beckman Optima XL-A centrifuge using the An60Ti rotor with quartz windows. All protein samples were at 10  $\mu$ M and had been buffer-exchanged into a buffer containing 20 mM Hepes, pH 7.5/2 mM DTT/200 mM KCl. For velocity sedimentation, the cLD was centrifuged at 20,000 rpm for 3 h at room temperature and scanned continuously with an absorbance reading at 280 nm. Sedimentation scans were analyzed with SEDFIT to calculate an  $S$  value (44). Equilibrium sedimentation was monitored by absorption at 280 nm at each of the following speeds: 10,000,

14,000, and 20,000 rpm at 25°C. Scans were analyzed by using MATCHV7, REEDIT9 (Jeff Lary, National Analytical Ultracentrifuge Facility, Storrs, CT), and WINNONLIN (Scientific Consulting, Inc., Apex, NC) programs assuming a single globular species to determine  $\sigma$  values, which were then used to calculate molecular weights and  $K_{ds}$ .

**MS.** Protein masses were determined by MALDI-TOF MS using a Voyager-DE STR Biospectrometry Workstation (Applied Biosystems). All recombinant proteins were buffer-exchanged into water and prepared for analysis by the addition of sinapinic acid matrix in a 1:1 ratio.

**Sequence Alignment.** Global alignments were initially performed on the full length LDs of the six species of Ire1 and PERK by using T-COFFEE (45). Scoring of analogous residues was performed with BLOSSUM62 (46). The sequence alignment and the conservation plot were made by using JALVIEW (47).

**Interface Specificity Evaluation.** The following metrics were calculated for Interface 1 and Interface 2 according to Bahadur *et al.* (17): B, total area of the interface;  $F_{np}B$ , fraction of interface created by nonpolar atoms;  $F_{bu}$ , fraction of interface made by fully buried atoms; and RP, residue propensity score. The values calculated for Interfaces 1 and 2 were compared with a reference data set for which the same four values had been calculated for each interface in that set. This reference set is composed of 122 homodimeric proteins, 70 protein–protein complexes, and 188 crystal packing interfaces. According to published work (17), B,  $F_{np}$ ,  $F_{bu}$ , and RP provide a predictive power of 95% and 93% success for correctly identifying a protein as monomeric or homodimeric, respectively. The values obtained for LD Interface 1 were B, 2,380  $\text{\AA}^2$ ;  $F_{np}B$ , 72.9%;  $F_{bu}$ , 41.2%; and RP, 2.0; and for Interface 2 were: B, 2,117  $\text{\AA}^2$ ;  $F_{np}B$ , 56.3%;  $F_{bu}$ , 26.5%; and RP, 5.9.

**Topography Map.** Surface representations of both yIre1-cLD and MHC-1 (1A1N;  $\alpha 1\alpha 2$  domain without bound peptide) were oriented with the grooves being parallel to the plane of the paper and perpendicular to the quasi-twofold axis ( $z$  axis), and with the long axis horizontal. The intersections of planes perpendicular to  $Z$  with the molecular surfaces were recorded at 2- $\text{\AA}$  intervals from the bottoms of the grooves to the outer boundaries of the protein. Contours of the sections are colored according to their position along the  $z$  axis and stacked. Zero depth (red line) was set for the position where the rim becomes discontinuous.

We thank Martha Stark for the construction of the expression plasmid for LD; James Holton, Chris Wadding, and Pascal Egea for their invaluable help with the data collection and analysis; and Damien Devos for his help with sequence alignment. We also thank Dyché Mullins for his expert help with the analytical centrifugation experiments, and the members of the Walter and Stroud laboratories for valuable discussion and comments on the manuscript. This work was supported by National Institutes of Health Grants DK 065671 (to F.R.P.), GM 60641 (to R.M.S.), and GM 32384 (to P.W.). P.W. is an Investigator of the Howard Hughes Medical Institute. The Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, is supported by National Institutes of Health Grant P41 RR-01081.

- Gething, M.-J. & Sambrook, J. (1992) *Nature* **355**, 33–45.
- Stevens, F. J. & Argon, Y. (1999) *Semin. Cell Dev. Biol.* **10**, 443–454.
- Cox, J. S., Shamu, C. E. & Walter, P. (1993) *Cell* **73**, 1197–1206.
- Patil, C. & Walter, P. (2001) *Curr. Opin. Cell Biol.* **13**, 349–355.
- Ron, D. (2002) *J. Clin. Invest.* **110**, 1383–1388.
- Zinszner, H., Kuroda, M., Wang, X., Batchvarova, N., Lightfoot, R. T., Remotti, H., Stevens, J. L. & Ron, D. (1998) *Genes Dev.* **12**, 982–995.
- Niwa, M., Patil, C. K., DeRisi, J. & Walter, P. (2005) *Genome Biol.* **6**, R3.
- Cox, J. S. & Walter, P. (1996) *Cell* **87**, 391–404.
- Rueggsegger, U., Leber, J. H. & Walter, P. (2001) *Cell* **107**, 103–114.
- Sidrauski, C. & Walter, P. (1997) *Cell* **90**, 1031–1039.

- Sidrauski, C., Cox, J. S. & Walter, P. (1996) *Cell* **87**, 405–413.
- Weiss, A. & Schlessinger, J. (1998) *Cell* **94**, 277–280.
- Stroud, R. M. & Wells, J. A. (2004) *Sci. STKE* **2004**, re7.
- Papa, F. R., Zhang, C., Shokat, K. & Walter, P. (2003) *Science* **302**, 1533–1537.
- Kimata, Y., Kimata, Y. I., Shimizu, Y., Abe, H., Farcasanu, I. C., Takeuchi, M., Rose, M. D. & Kohno, K. (2003) *Mol. Biol. Cell* **14**, 2559–2569.
- Bertolotti, A., Zhang, Y., Hendershot, L. M., Harding, H. P. & Ron, D. (2000) *Nat. Cell Biol.* **2**, 326–332.
- Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. (2004) *J. Mol. Biol.* **336**, 943–955.

18. Jardetzky, T. S., Lane, W. S., Robinson, R. A., Madden, D. R. & Wiley, D. C. (1991) *Nature* **353**, 326–329.
19. Liu, C. Y., Wong, H. N., Schauerte, J. A. & Kaufman, R. J. (2002) *J. Biol. Chem.* **277**, 18346–18356.
20. Kimata, Y., Oikawa, D., Shimizu, Y., Ishiwata-Kimata, Y. & Kohno, K. (2004) *J. Cell Biol.* **167**, 445–456.
21. Oikawa, D., Kimata, Y., Takeuchi, M. & Kohno, K. (2005) *Biochem. J.* **391**, 135–142.
22. Achour, A., Persson, K., Harris, R. A., Sundback, J., Sentman, C. L., Lindqvist, Y., Schneider, G. & Karre, K. (1998) *Immunity* **9**, 199–208.
23. Olson, R., Huey-Tubman, K. E., Dulac, C. & Bjorkman, P. J. (2005) *PLoS Biol.* **3**, e257.
24. Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F. & Gething, M. J. (1993) *Cell* **75**, 717–728.
25. Bjorkman, P. J. & Parham, P. (1990) *Annu. Rev. Biochem.* **59**, 253–288.
26. Rossmann, M. G. (1989) *J. Biol. Chem.* **264**, 14587–14590.
27. Shamu, C. E. & Walter, P. (1996) *EMBO J.* **15**, 3028–3039.
28. Liu, C. Y., Schroder, M. & Kaufman, R. J. (2000) *J. Biol. Chem.* **275**, 24881–24885.
29. Tardif, K. D., Waris, G. & Siddiqui, A. (2005) *Trends Microbiol.* **13**, 159–163.
30. Reimold, A. M., Ponath, P. D., Li, Y. S., Hardy, R. R., David, C. S., Strominger, J. L. & Glimcher, L. H. (1996) *J. Exp. Med.* **183**, 393–401.
31. Leahy, D. J., Erickson, H. P., Aukhil, I., Joshi, P. & Hendrickson, W. A. (1994) *Proteins* **19**, 48–54.
32. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276**, 307–326.
33. Holton, J. & Alber, T. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 1537–1542.
34. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter Protein Crystallography* **31**, 34–38.
35. Sack, J. S. (1988) *J. Mol. Graphics* **6**, 244–245.
36. Terwilliger, T. C. & Berendzen, J. (1999) *Acta Crystallogr. D* **55**, 849–861.
37. de La Fortelle, E. & Bricogne, G., eds. (1997) *Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multi-wavelength Anomalous Diffraction Methods* (Academic, New York), Vol. 276, pp. 472–494.
38. Brunger, A. T. (1988) *J. Mol. Biol.* **203**, 803–816.
39. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998) *Acta Crystallogr. D* **54**, 905–921.
40. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) *Acta Crystallogr. D* **53**, 240–255.
41. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400.
42. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1983) *J. Appl. Crystallogr.* **26**, 283–291.
43. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004) *J. Comput. Chem.* **25**, 1605–1612.
44. Schuck, P. (2000) *Biophys. J.* **78**, 1606–1619.
45. Notredame, C., Higgins, D. G. & Heringa, J. (2000) *J. Mol. Biol.* **302**, 205–217.
46. Henikoff, S. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919.
47. Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. (2004) *Bioinformatics* **20**, 426–427.