Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction

Chris P. PONTING¹

MRC Functional Genetics Unit, Department of Human Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, U.K.

Sequence database searches, using iterative-profile and Hidden-Markov-model approaches, were used to detect hitherto-undetected homologues of proteins that regulate the endoplasmic reticulum (ER)-associated degradation pathway. The translocon-associated subunit Sec63p (Sec = secretory) was shown to contain a domain of unknown function found twice in several Brr2p-like RNA helicases (Brr2 = bad response to refrigeration 2). Additionally, Cue1p (Cue = coupling of ubiquitin conjugation to ER degradation), a yeast protein that recruits the ubiquitin-conjugating (UBC) enzyme Ubc7p to an ER-associated complex, was found to be one of a large family of putative scaffolding-domain-containing proteins that include the autocrine motility factor receptor and fungal Vps9p (Vps = vacuolar protein sorting). Two other yeast translocon-associated molecules, Sec72p

and Hrd3p (Hrd = 3-hydroxy-3-methylglutaryl-CoA reductase degradation), were shown to contain multiple tetratricopeptide-repeat-like sequences. From this observation it is suggested that Sec72p associates with a heat-shock protein, Hsp70, in a manner analogous to that known for Hop (Hsp70/Hsp90 organizing protein). Finally, the luminal portion of Ire1p (Ire = high inositol-requiring), thought to convey the sensing function of this transmembrane kinase and endoribonuclease, was shown to contain repeats similar to those in β -propeller proteins. This finding hints at the mechanism by which Ire1p may sense extended unfolded proteins at the expense of compact folded molecules.

Key words: CUE domains, SEL-1, sequence analysis, translocon subunits, unfolded protein response.

INTRODUCTION

Recent advances in sequence-database-searching algorithms have provided considerable insights into the evolution, structure and function of proteins. Sequence analysis, when undertaken with due regard to statistical significance of similarities, has elicited an increasing realization that a great number of eukaryotic protein sequences contain multiple domains. Homologues of these domains are often seen in non-orthologous gene products and in differing domain arrangements. This enables the prediction of function for one domain when experimental information is available for at least one other homologue.

Many analyses have targeted the identification of domain homologues in intracellular signalling, extracellular matrix and DNA- or RNA-associated proteins (reviewed in [1]). One evolutionarily conserved set of proteins, however, that has hitherto escaped detailed analysis are molecules, sited in the endoplasmic reticulum (ER), that represent a quality-control system for misfolded or unassembled secretory proteins. The goal of the present study was to address this imbalance by undertaking thorough sequence analyses of proteins known to contribute to ER-associated degradation.

Studies (reviewed in [2,3]) have shown that an evolutionarily conserved process of folding quality control exists in eukaryotic cells' ERs that specifically targets misfolded molecules for proteolytic degradation. This process centres on the translocon. This is the aqueous channel formed in the ER that translocates proteins from the cytoplasm into the ER [4]. Secretory proteins

are conducted through the membrane via interactions with translocon protein components, which may vary according to the type of secretory protein involved [5,6]. In yeast, Sec61p (Sec = secretory) and its two paralogues, Sbh1p (Sbh = Sec61 beta homologue) and Sss1p (Sss = Sec61 suppressor), as well as Sec62p, Sec63p, Sec71p and binding protein (BiP), are required to reconstitute post-translational translocation in protoliposomes [7]. In mammals, a heterotrimer of Sec61p homologues with the translocation-associated membrane protein ('TRAM') are similarly required for translocation of most proteins.

In recent years it has become apparent that, in addition to its role in importing secretory proteins from the cytoplasm into the ER, the translocon and its associated subunits are responsible for conducting misfolded or unnecessary proteins from the ER to the cytoplasm [3]. This retrograde transport mechanism appears to deliver proteins for ubiquitination and subsequent digestion by the proteasome [2]. Thus the ER membrane is the scene of great industry, with protein synthesis and degradation occurring in the cytoplasm, and protein-folding quality control occurring within the ER lumen.

Aberrant protein folding is associated with the pathogenesis of amyloidoses, such as Alzheimer's and prion diseases [8]. The surfeit of proteins in amyloidoses is in contrast with those diseases associated with protein deficits, such as cystic fibrosis. Mutations in the cystic-fibrosis transmembrane conductance regulator result in misfolded molecules that are targeted for translocation into the cytoplasm, where they are degraded by the proteasome [9]. In yeast, ubiquitination is facilitated by ubiquitin-

Abbreviations used: ER, endoplasmic reticulum; Hsp, heat-shock protein; Hop, Hsp70/Hsp90 organizing protein; PSI-BLAST, position-specific iterative BLAST; nrdb, non-redundant protein sequence database; TPR, tetratricopeptide repeat; AUP1, ancient ubiquitous protein 1; AMF(R), autocrine motility factor (receptor); PQQ, pyrroloquinoline quinone; snRNPs, small ribonucleoprotein particles; Hsc70, heat-shock cognate protein 70; UBC, ubiquitin-conjugating; TM, transmembrane; Sec, secretory; Hrd, 3-hydroxy-3-methylglutaryl-CoA reductase degradation; Vps, vacuolar protein sorting; Brr2, bad response to refrigeration 2; BiP, binding protein; Cue, coupling of ubiquitin conjugation to ER degradation; Der, degradation in the ER; Ire, high inositol-requiring; LAMA, local alignment of multiple alignments; Lin-12, lineage-12; SEL-1, suppressor and/or enhancer of LIN-12; Sbh, Sec61 beta homologue; Sss, Sec61 suppressor.

¹ e-mail Chris.Ponting@anat.ox.ac.uk

conjugating (UBC) enzymes Ubc6p and Ubc7p [10] and the scaffolding protein Cue1p (Cue = coupling of ubiquitin conjugation to ER degradation) [11].

Detailed investigation of the translocon subunits in yeast has facilitated a greater understanding of the corresponding molecules in mammals. However, the identities of the mammalian counterparts (orthologues) of yeast Cue1p, Sec72p, Der3/Hrd1p (Der = degradation in the ER; Hrd = 3-hydroxy-3-methylglutaryl-CoA reductase degradation), Der1p, Hrd3p and Sec71p remain unclear. Similarly, many of these proteins lack well-defined domains or motifs, thus hindering the understanding of the roles of these proteins in ER-associated translocation and degradation. Here I describe the use of iterative database searching [12,13] and other methods to identify key domain structures in translocon subunits and thereby provide predictions of their functions.

EXPERIMENTAL

Position-specific iterative BLAST (PSI-BLAST; [12]) sequence searches employed the non-redundant protein sequence database (nrdb) available from the National Center for Biotechnology Information (<ftp://ncbi.nlm.nih.gov/blast/db>). MoST and HMMer2 searches used a database (nrdb90) containing no pairs of sequences that are greater than 90% identical [14]. PSI-BLAST was the method of choice in analysing protein sequences. The threshold of $E(x)$ (the number of random sequences expected purely by chance to be aligned with scores greater than, or equal to, a value x) was chosen typically at 10^{-2} or 10^{-3} . Since database searches can never be guaranteed to generate no false-positive sequences, alignments were inspected visually in order to ensure conservation of signature motifs that are characteristic of domain families.

Sequences were also subjected to the automatic domain prediction protocol embodied in SMART ([15]; <http://smart.embl-heidelberg.de/>). Sequences established as being significantly similar to warrant claims of homology (evolutionary descent from a common ancestor) were multiple aligned and re-compared with databases using the HMMer suite of algorithms (v.1 and v.2; S. Eddy, personal communication; <http://HMMer.wustl.edu/>). This approach was useful in detecting the presence of additional tandem repeats that scored lower than the optimal alignment score.

Significantly similar sequence motifs were detected using MoST [13] with an E -value inclusion threshold of 0.05 and a sequence redundancy identity threshold (I) of 80%. Pairs of ungapped alignment blocks were compared for establishing the significance of their similarities using LAMA [16]. This method allows estimation of the number of blocks (E_{5000}) in a database of 5000 blocks expected, by chance, to score at least score Z .

In constructing multiple alignments of tandem repeat sequences, the known tertiary structures of homologues were used as guides. Multiple alignments were presented using CHROMA (L. Goodstadt and C. P. Ponting, unpublished work; <http://www.lg.ndirect.co.uk/chroma/index.htm>) and default options. Secondary-structure predictions are those of PHD [17].

RESULTS

A combination of iterative-profile, Hidden-Markov-model and motif (ungapped-alignment-block) analyses was applied to the sequences of yeast proteins known to be involved in ER-associated degradation. No additional insights into evolution, structure or function were obtained for proteins such as Der1p, Sec61p, Sbh1p, Sss1p and Sec71p. Novel observations of homology, however, were obtained for the following five proteins.

Sec63p

A gapped BLAST [12] search of protein sequence databases with the sequence of human Sec63p [18] (amino acid residues 210–760) revealed significant similarity ($E < 10^{-8}$) with yeast Brr2p-like helicases (Brr2 = bad response to refrigeration 2). Further investigation using PSI-BLAST and HMMer2 revealed that many of these helicases contain pairs of tandem repeats of a domain found singly in Sec63p (Figures 1 and 2).

Sec72p

Querying the SMART database [15] with the yeast Sec72p sequence reveals a single tetratricopeptide repeat (TPR) ($E = 4.8 \times 10^{-3}$). A subsequent PSI-BLAST search with Sec72p was consistent with this proposal, as it revealed significant similarity ($E = 9 \times 10^{-8}$) with three TPRs within rat heat-shock protein (Hsp) 70/Hsp90 organizing protein (Hop). Detailed comparison of the Sec72p sequence with a Hidden Markov model constructed from a TPR multiple alignment resulted in delineation of three TPRs in yeast Sec72p (Figure 3). A BLASTP search of databases with the N-terminal 536 amino acids revealed similarity ($E = 5 \times 10^{-3}$) to a region of mouse p58 [19] that contains five TPRs.

Cue1p

A search of sequence databases with yeast Cue1p (amino acids 21–203) using PSI-BLAST [12] and an E -value inclusion threshold of 10^{-2} revealed significant similarity to yeast YML101c, *Schizosaccharomyces pombe* Cue1p (SPCC4G3.13c), metazoan ancient ubiquitous protein 1 (AUP1) orthologues and metazoan autocrine-motility-factor receptor (AMFR) orthologues within four iterations; these sequences represent group I CUE domains (Figure 4).

Group II CUE domains were identified similarly using mouse Tollip (amino acids 152–274) as a PSI-BLAST query; this identified a *Caenorhabditis elegans* orthologue and *Fugu rubripes* (pufferfish) BAW (where BAW = between AKAP84 and WSB1 genes) as homologues within 1 iteration. Group II CUE domains were supplemented with three further members [*C. elegans* F25H2.1, yeast Vps9p (Vps = vacuolar protein sorting mutants) and yeast YKL090w] by querying databases using a 26-amino-acid alignment block and MoST [13] (parameters: $E < 0.05$ and $I < 80\%$).

Groups I and II CUE domains were found to be significantly similar when compared using local alignment of multiple alignments (LAMA) [16] (Z -score, $Z = 7.2$; $E_{5000} = 1.3 \times 10^{-2}$). Group III CUE domains were found in fungal, plant and metazoan orthologues of yeast YGL110c, whereas Group IV CUE domains were found in yeast YDR273w and YOR042w (and fungal orthologues). Comparison of Group III and IV CUE domain alignment blocks with a single multiple alignment of Groups I and II CUE domain sequences with LAMA produced significant E_{5000} values of 3.8×10^{-2} ($Z = 6.7$) and 2.5×10^{-3} ($Z = 7.9$) respectively.

Hrd3p

PSI-BLAST analysis using yeast Hrd3p as query identified with significance ($E < 10^{-3}$) TPR-containing sequences, such as those in *Arabidopsis thaliana* SPINDLY [20] within two rounds. The second round of searching also revealed three significant ($E < 10^{-3}$) sub-optimal alignments between yeast Hrd3p and mouse suppressor and/or enhancer of lineage-12 (LIN-12) (SEL-1), thereby indicating the presence of multiple repeats. These repeats were analysed using the TPR periodicity as a starting basis.

200k_HUMAN/1	NFQVTELGRITASHY---YITNDVTQYTNQLLK-PTLSEIELEFRVFLSSSEFK---NITVREEEKLZQKLLERVPVVK---ESIEEPSAKINVLLOAFISOLK---LEGFALMAMVYVTQSA
RNAh_HUMAN/1	YFSSDLDGRITASHY---YIKYNTIETFNELFD-AHKTEGDIFAIVSKAEEFD---QIKVREEEIEELDTLLSNFCBLST---PGGVENSYGKINILLOTYISRGE---MDSFSLISDSAYVAQNA
BRR2_YEAST/1	VIEATDLGNIASSE---YINHASMDVYNRELD-EHTTQIDLEFRIFSMSEEFK---YVSVRYEEKRELKQLEKAPIPIR---EDIDDLPLAKVNVLLQSYFSQLK---FEGFALNSIVFIHQNA
YG50_YEAST/1	HFIADKDLGRVSSDF---YLLNESVEIFNQMC-D-PRATEADVLSMISMSSEFD---GIKPREEESKELKRLSDESVEQC-I---GSQOLDTPQGANVILLOAYISQTR---IFDSALSSDSNYVAQNS
200k_HUMAN/2	DVAPLNLGMIAAAY---YINYTTIELFMSMLN-AKTKVRGLIEIISNAABYE---NIPTRHHEDNLLRQLAQKVPHKLN---NPKFNDPHVKTNLLOAHLSRM---QLSABLQSDTEEILSKA
RNAh_HUMAN/2	SIEPLTYGRIASYY---YLKHQTVKMFKDR-LK-PECSTEELLSILSDAEYNT---DLPVRHNEHDHMSSELAKCLPIESN---PHSFDSPHTKAHLILLOAHLSRAM---LPCPDYDTTKTVLDQA
BRR2_YEAST/2	IISTLSNGLIASHY---GVSFFTIOSEVSSLS-NTSTLKNMLYVLSTAVFEFE---SVPLRKGDRALVVKLSKRLPLRFPE---HTSSGSVSRVFLLOAYFSRL---ELPVDVFQNLKDILEKV
YG50_YEAST/2	DIVATPFLSISSYY---YISHLTIQQLLQI-H-DHATFQEVLRWLSLAVEYN---ELPVRGGEIIMNEEMSQQSRSYSVE(7)ELPMWDPHVKTNLLOAHLSRVDD---LPIADYIQDTVSVLDQS
HFM1_YEAST/1	EYKSTAYGNAMTRH---YISFESMKQFINAKK---FLSLQGLINLLATSEEFSS---VMRVRHNEKKLFKEINLSPLLKYP(7)SQIIDRVSOVSLILQYELGGLE(8)KLHQTIVQDKFLVFRHC
SEC63_HUMAN/1	GSWWYRSIRYSGDQ---ILIRTTQIYTYFYVYKT-RNMDMKRLIMVLGASGFD(9)RPTDNTLIPQLITREIGSINLKNE---PPLTCPYSTIARVILLSHLARM---KIPETLEEDQQFMLKCC
NPL1_YEAST/1	SRWWARTQSYTTKGIHNVASNFVSNLVNYPSEIVTTDLILHLWLSFAHEFK---QFPDLOPTD-----FEKTIQDHINRR---DSGKLNNAKFRIVAKC
Consensus/80%	.h.shphhphss.a...hlp..slp.a.p.bp..phs...h.h.l.s.s.Eap...h.hRb.c..bbppl.pp..b.....h.ps.hKhpILLOsalsph...b.s.sh.pd...lhpps
2-structure	eee e hhhHHHHHHHHHH hHHHHHHHHHHHH hHHHHHHHHHHHH hHHHHHHHHHHHH hHHHHHHHHHHHH hHHHHHHHHHHHH
200k_HUMAN/1	GRMLRAIFEIVLNRGWAQL---TDKTLNLCKMIDKRMQSM---CPDRQFRKLPEEVVKKIEK---KNFPFERLYDIN---HNEIGELIRMP---KMGTIKHYVHLFKLELSVHLQPI--
RNAh_HUMAN/1	ARIVRALFEIALRKRWPTM---TYRLNLNLSKVIDKRLAGWA---SPDRQFSEILPPHILTRLEE---KKLTVDKLKDMR---KDEIGHILNHV---NIGLKVKQCVHQIPSSVMMEASIQPI--
BRR2_YEAST/1	GRLLRAMFEICLRGCGHP---TRMLNLNCKSATTKMPTN---CPDRQFKTCPEVEIKRLEA---STVPWGGYLOPET---PAEVGRAIRSE---KYGKQVYDLLKRFKMSVTCNAQPI--
YG50_YEAST/1	VRICRALFLIGVNRWGF---SNVMLNICKSIEKRLWAFD---HPLCQFD---LPENIIRIRRD---TKPSMEHLLE---ADELCGLVHNK---KAGSRLYKILSRFEKINTEAEIIFI--
200k_HUMAN/2	IRLIQACVDVLSNGWLSP---ALAAAMELAQMVTQAMSED---SYLRRLLPPFPSPGLFKRCTD---KGVESVFDIME---DEERNALLQTD---SQIADVARFCNRYENTIELSYEVVDK--
RNAh_HUMAN/2	LRVQCAMLDVAANQGWLV---VLNITNLIQMVIQGRMLKD---SSLTLPLNIENHHLFLFK(10)RGRTSIECLPETIACGGKDHVFSMV(5)AAKTKQAWNFLSHLPVINVGISVKGSGW
BRR2_YEAST/2	VPLINVVVILSANGYLN---ATTAMDLAQMLIQGVVDVD---NPLRQIPHFNKILEKCKE---INVEYVYDIMA---DEERDEILTLD---SOLAQVAAFVNNYINVELTYSLNNS--
YG50_YEAST/2	LRILQAYIDVASELGYFHT---VLTMIKMMQCKQGYWYED---DPVSVLPLGLQLRRIKDYTF(7)MTPQKKKKLLTTEE---IGRFGYKKLLNV(12)DTKKRFVSVQRLPVLGEMKFEEQEN-
HFM1_YEAST/1	FRLLKCMVDTFIEKSDGTS---LKNLTFLRLSLNGHQENTP---MVRQLKKTIGLVSVRLIR---HGITNLEEMGHSS---DTQIEYLLNLK---IGNGIKIKNDISLLPCLNIRTKLENC-
SEC63_HUMAN/1	PALLQEMVNVICQLIVMAR(14)LENCMKLSQMAVQGLQQFK---SPILQLPHIEEDNLRVRSNHK---KYKIKTIQDLVSIK---ESDRHTLLHFLE---DEKYEEVMAVLGSFVYVMDIKSQVL(5)
NPL1_YEAST/1	HSLHGLLDIACFRNLDI---ALGAINTFKCTIQVAVPLTP---NQILQLPNVDKEHPITKTG---DIHTLGLKLTTE---DAKIGEVLGIDK---QAKLNETLRVASHIENLKIKADFLV(5)
Consensus/80%	hplhphshclh.pb.hhp.....s..hhplhp.l.p.hW..s..ssLbph..h..p.lcchp.....sh.clhplp...scbsbhlp.....h.phhphpphP.lph.hp.bs...
2-structure	HHHHHHHHHHHHHH hH HHHHHHHHHHHHHHHHH hHHHHHHH hHHHHhh eEEEE hHHHHHHH eEEEEe
200k_HUMAN/1	-TRSTLKVLTITPDPQWDEKVGHS---SEAFWILVEDVDSE---VILHHEYFLKAKYAQ---DEHLITFEVVFEP---LPPQYFIRVVSRLW---SCTQLPVSF 3255965 (543- 852)
RNAh_HUMAN/1	-TRTVLRVILSIYADFTWNDQVHGTV---GEPWWIIVEDPTND---HIYHSEYFLALKKQVISKEAQLLTIPIFEP---LPSQYYIRAVSRLW---GAEAVCIINFQ 2842424 (39- 352)
BRR2_YEAST/1	-TRSVMRFNIEIADWIWDMNVHGS---LEPFLMLEDTDGD---SILYYDVLFITPDIVG---HEFTLSBTYELKQH(4)LPPNFFLTILISENWW-HSEFEIPVSFN 731380 (995-1309)
YG50_YEAST/1	-TTNVMRIHIALGPDVWDSRIHGD---AQFFWVFVEESDKS---QILHFEKFLINRRQLN---NQHEMDMIFLSDP---LPPQVVVKVSDTWI---GCSTHAISFQ 1723776 (792-1101)
200k_HUMAN/2	-DSIRSGGPVVVLQLEREEVETGP(10)EEGWVVIGDAKSN---SLISIKRLTLQQA---KVKLDEVAPATGG---RHN-TLYFMSDAYM-GCQOEYKFSVD 3255965 (1374-1690)
RNAh_HUMAN/2	-DDLVEGHNELSVSTLTADKRDDNK(41)DEGWFLILGEVDKR---ELTALKRVGYIRNH---HVASLSHYTPEIPG---RYIYTLFMSDCYL-GLDQQYDNLSD 2842424 (873-1242)
BRR2_YEAST/2	-DSLISGVKQKITIQLTRDVEPENL(11)LESWVLVLGEVSKK---ELYAIKVTILNKET---QOYELEEDTPTSG---KHNLTWCVDOSYL-DADKELSFEIN 731380 (1843-2161)
YG50_YEAST/2	-NEVLTFYSKHLSSKHNGFEVYCD-(5)QKELWFLIGHKGD---ELLMIKRKQPKQMN---KEVIHCLDFIPEIR---GEELQFSLINDALG-LRYDMVHKLIS 1723776 (1623-1967)
HFM1_YEAST/1	-KIENEELWLTTFKVEISATFKSSI(13)SSGELIDFRRLQVN---SKLQSPRGFRISAKISP---KLEKIEFSIHQEI---AGLGKTIYVSTDLA-SQFSAKTPNIR 1708195 (474- 814)
SEC63_HUMAN/1	)NNITVGLSVTLVLKLTROTMAEV(165)QEWVWLYIADRKEQ---TLISMPYHVCTLKD---TEEVELKVPAGKPG---NYQYTVFLRSDSYM-GLDQIKPLKLE 6005872 (209- 715)
NPL1_YEAST/1	)VTPSSTPYISLKVLSAKQPLIP(44)GSWCCLVSSQKDGKILQTPIIIEKLSYKLN(29)IKIPLGQPAETVG---DFFFRVIVKSTDYFTTDLITNMKVR 6324828 (240- 611)
Consensus/80%	.sp....h.b.l.sphs.p.p.....p.hhlhl.c.p.p...plh.hchh.hp.p.....bh.lpf.hp.....bhl.h.sD.ah.s.c..h..php
2-structure	eeeeeeEEEEEEEEe eEEEEEE eEEEEeeeeeE EEEEEe EEEEEEEe e

Figure 1 Multiple alignment of Sec63p-like domains

Consensus symbols: a, aromatic (FWHY, lightly shaded); b, big (EFHIKLMQRWY, lightly shaded and in italics); c, charged (DEHKR); h, hydrophobic (ACFGHILMTVWY, lightly shaded); l, aliphatic (ILV, lightly shaded); p, polar (CDEHKNQRTS); s, small (ACDGNPSTV). The black background indicates amino acid identity in > 80% of sequences. Numbers in brackets refer to the number of amino acids excised from the alignment. Predicted [17] secondary structures are indicated below the alignment (h/H indicates helix and e/E indicates beta strands); lower-case letters represent predictions that have expected accuracies of > 72%, and upper-case letters represent predictions that have expected accuracies of > 82%. Residue numbers and GenPept identifiers are shown following the alignment. Abbreviations: 200 k, U5 snRNP-specific 200kDa protein; BRR2, Bad Response to Refrigeration 2; HFM, 119 kDa DNA/RNA Helicase Family Member; NPL, Nuclear Protein Localization; RNAh, RNA helicase; YEAST, *Saccharomyces cerevisiae*; and YG50, YGR271W gene.

Figure 2 Schematic representation of the modular architecture of proteins discussed herein (approximately to scale)

Symbols and abbreviations of domains: ACYLT, tafazzin-like acyltransferase domain; C, CUE domain; C2, protein kinase C C2-like domain (SMART: C2); DEXD, DEXD-type helicases (ATPases, SMART: DEXDc); Hc, helicase C-terminal domain (SMART: HELICc); J, DnaJ molecular chaperone domain (SMART: DnaJ); P, PQQ dehydrogenase-like β -propeller repeat; Rn, endoribonuclease domain; SMR, small mutation S-related domain (SMART: SMR); S/T Kinase, Ser/Thr-type protein kinase (SMART: S/TKc); Vps9, Vps9-like Rab guanine nucleotide exchange factor domain (SMART: Vps9). Lightly and darkly shaded thin rectangles represent TPRs and Sel1-like repeats respectively. Thin open rectangles represent transmembrane sequences. Thick ring structures represent RING domains (SMART: RING). Multiple alignments for Sec63 and CUE domains, and for Sel-1-like repeats, are available from SMART. The scale corresponds to the number of amino acids.

Multiple repeats in Hrd3p and SEL-1 homologues were identified using an iterative approach of multiple alignment and HMM construction, and sequence comparison using an intentionally low bits-score threshold (> 10 bits). Predicted repeats were consistent with PSI-BLAST analysis output.

Ire1p (Ire = high inositol-requiring)

PSI-BLAST analysis of the ER luminal region of yeast Ire1p (amino acids 1–537) revealed significant similarity ($E < 0.001$) to prokaryotic kinases (for example, *Methanobacterium thermoautotrophicum* MTH1485) by round 2, and bacterial pyrroloquinoline quinone (PQQ) dehydrogenases (for example, *Paracoccus denitrificans* XoxF) by round 4. By inspection of the known tertiary structures of PQQ dehydrogenases (for example, that of *Pseudomonas aeruginosa* ethanol dehydrogenase [21]) it was apparent that four regions of significant sequence similarity between PQQ dehydrogenases and Ire1p represent the four-stranded β -sheet structures of β -propellers, termed 'blades'.

DISCUSSION

Sec63p and Brr2p-like RNA helicases

Sequence analysis clearly demonstrates that some components of the translocon, such as Sec61p and BiP/KAR2p, are ancient in origin, because their homologues are found in Bacteria and Archaea [22,23]. Mammalian BiP and its yeast orthologue Kar2p are members of the Hsp70 (DnaK) ATPase family that bind substrates in co-operation with DnaJ domain-containing proteins

[24]. Yeast BiP/Kar2p binds a DnaJ domain of Sec63p and mediates the import of polypeptides into the ER lumen [25] and the export of misfolded proteins into the cytoplasm [26].

The DnaJ domain of Sec63p is known to face the ER lumen [27]. The large cytoplasmic portion of Sec63p has hitherto not been found to be homologous with any other protein. However, sequence searches demonstrated significant similarity of this region of Sec63p to two regions of a family of helicases typified by yeast Brr2p and human U5-200kD [28,29]. Further analysis demonstrated that these RNA helicases contain two Sec63p-like domains (Figure 1) that each lie C-terminal to an ATPase domain (Figure 2). The yeast HFM1 putative helicase was found to contain only one Sec63p-like domain.

The Brr2p N-terminal ATPase domain is essential for *in vitro* RNA unwinding in U4/U6 small ribonucleoprotein particles (snRNPs) [30,31]. Although the functions of the helicase Sec63p-like domains remain unknown, the likely associations of the translocon and Brr2p with RNA-protein complexes (the ribosome and the spliceosome respectively) might indicate that Sec63p acts as a ribosome receptor. If true, this would be unexpected, since other translocon components have previously been thought to participate in anchoring the ribosome to the ER membrane, but this issue is far from being resolved [32].

Sec72p and TPRs

Yeast Sec71p/Kar7p and Sec72p represent additional components of a Sec63p-associated complex [33]. Only fungal homologues of Sec71p are evident from database searches with Sec71p. However, database searches (Figure 3) reveal the presence of

SC72_YEAST/1	IKGLFESGMKNLQKKLNALKNVSLAIEAQRK	1351043	(58- 91)
SC72_YEAST/2	HFMLRSKIDLCILGKHLALQDLDFLLGTGLIQ	1351043	(106-139)
SC72_YEAST/3	PDVFVRKADCLLKLQWEHARATCERGLALAPED	1351043	(140-173)
JEM1_YEAST/1	RLSIHYVISVLQTFVLNSDASFNIRKCLSIDMDY	6322388	(223-256)
JEM1_YEAST/2	DKCKKLSLTISKLNKVNPSKRQILDPAFYAFENK	6322388	(257-290)
JEM1_YEAST/3	FRSWDRIEFYLLKDKKPFITPMKILNKDTNFKN	6322388	(292-325)
HOP_HUMAN/1	VNELKEKGNKALSVGNIDDALQCYSEAIKLDPHN	7766912	(4- 37)
HOP_HUMAN/2	HVLYSNRSAAAYAKKGDYQKAYEDGCKTVDLKPDW	7766912	(38- 71)
HOP_HUMAN/3	GKGYSRKAALFLNRFEBAKRTYEEGLKHEANN	7766912	(72-105)
P58_MOUSE/1	VEKHLELGKKLLAAGQLADALSQFHAAVDGDPDN	1353272	(37- 70)
P58_MOUSE/2	YIAYYRRATVFLAMGSKRRVPDLTKVIALKMDF	1353272	(71-104)
P58_MOUSE/3	TAARLQRGHLLLKQGLDEAEDDFKKVLKCNPSE	1353272	(105-138)
P58_MOUSE/4	AELRELRAECFIKEGEPRKAISDLKAASKLKSDN	1353272	(188-221)
P58_MOUSE/5	TEAFYKISTLYYQLGDHLSLSEVRECLKLDQDH	1353272	(222-255)
P58_MOUSE/6	LNKLIESAEELIRDGRYTDATSKYESVMKTEPSV	1353272	(278-301)
P58_MOUSE/7	VRSKERICHCFSKDEKPVBAIRICSEVLQMEPDN	1353272	(306-339)
P58_MOUSE/8	VNALKDRAEAYLIEEMYDEAIQDYEAQEHEND	1353272	(340-373)
Consensus/80%	h..b.p.hp.hb...p..cs..php.sh.h..p.		
2-structure (HOP)	HHHHHHHHHHHHH HHHHHHHHHHHH		

Figure 3 Multiple alignment of TPRs in yeast Sec72p (SC72) and Jem1p, human Hop and mouse p58 (protein kinase inhibitor p58)

The known secondary structure of the TPRs in human Hop [38] is shown beneath the alignment. An additional C-terminal helix seen in the Hop crystal structure (helix C; [38]) is also apparent at the C-terminus of yeast Sec72p (not shown). Consensus and secondary-structure symbols are as in Figure 1. Residue numbers and GenPept identifiers are shown following the alignment. Abbreviation: YEAST, *Saccharomyces cerevisiae*.

I			
CUE1_YEAST	VTTQMVEVQNLAENLHPEQIRYSLE-NTGSVEETVERYLRGD	2506768	(65-106)
SPCC4G3.13C_SCHPO	VNSEHVHLVKTVFPHLESSAIAYDLQ-KTKNVDATIENALRGQ	7491914	(48- 89)
Yml101c_YEAST	VNSDMVEIVMTMAHPVQEKVQDILR-NTGSIETMENIFAGK	2497070	(74-115)
AUP1_HUMAN	QLATLAQRVKEVLPHVPLGVQIRDLA-KTGCVDLTITNLLEGA	5410292	(296-337)
AMFR_HUMAN	QLNAMAHOIQEMFPQVPYHLVQDLQ-LTRSVEITTDNILEGR	5931955	(456-497)
AMFR_CAEL	QLQTMLEQVREMPQMSVDIIMTDLR-QSGSAQSTIENILEGR	2829745	(432-473)
EG:95B7.9_DROME	DIDLEVTAVLEVLDPDLGSGFIRRLII-RYDNSEQVIAAILDDN	3255952	(453-494)
YL35_CAEL	QMDECAMRIKQSEFSPFHLSAIRRDLE-KTRSQTITVNNLKAGK	465791	(305-346)
ORF_HUMAN	ELDSLISQVKDLLPDLGEGFLLACLEYHYHDPEQVINNIIEER	6094659	(330-372)
T22C5.20_ARATH	VMQSKISQIKDIFPEYGNGLAACLEAYNQNPPEEVIQRIIEGT	6693023	(1493-1535)
II			
TOLLIP_HUMAN	CSEEDLKAIQDMFPNMDQEVIRSVLEAQRGNKDAAINSLQMG	6048243	(228-270)
BAW_FUGU	EFNQAMEDFKTMFPMSDYEVLECVLRSNNGAVDATIDQLQMS	4887239	(46- 88)
BAW_DROME	EFSQAMTDFKMFDPDIDREVIHAILRANLGAVDQTIDALLAMS	7292171	(1- 32)
YR75_CAEL	DFDRAMLDFOAMFPLSNSHLYVLRKYDGDVSATINELLYDN	1176809	(60-102)
F25H2.1_CAEL	ITEEDTKIEQEMFPIVDKEVIKCIIEERRVFFQLKYLPIKRIT	7499823	(195-237)
VPS9_YEAST	ERKDTLNTLQNMFPDMDPSLIEDVCI AKKSRI GPCVDALLSLS	665940	(408-450)
VPS9_SCHPO	EKAEAITALRAMFPADFSEVIEVVLNAQQRLSSSIDSLLEMS	SPBC4F6	
Ykl090w_YEAST/1	DHESKLSILDMFPAPISKVLQVHLENNNDLDTIGLLKEN	549704	(8- 50)
Ykl090w_YEAST/2	TVDNELHQLYDMFQDQDCSVIKDQFVINEKSVESTISDLNVE	549704	(55- 97)
L2602.6_LEIMA	ERHNALLTLISLFPQVDPVAVEAVLSQHHGAFADAYNVLLCAQ	7494013	(563-605)
III			
Ygl110cp_YEAST	VNEEQLSALMELFPQFSKYQLSQTLLAYDNNIELVTNKIFEDP	1723903	(316-358)
Ygl110cp_CANAL	DNIEGISILVDLWPNMTERKAKIILKQHNGDVEHVTNLLLENP	Con4-3058	
Ygl110cp_SCHPO	INPGDVKSLELFPQLSVEEAVEHLASLGNIDAAACESVITSS	7491869	(271-313)
IV			
Yor042wp_CANAL	PTQEIIDDMKQAFPNIEEKYIATLIASQGNPDPAFNALLYIS	Con4-2840	
Yor042wp_YEAST	KENPILQELKDAFPNLEEKYIKAVIIASQGVLSPAFNALLFLS	6324615	(97-139)
SPBC16E9.02c_SCHPO	MAEKARATLKEAFENTDDAIRAVLAASGYKLEPAFNALLGLS	7491403	(17- 59)
Ydr273w_YEAST	EKLCVLKELKIAFPEVDDTLKAILIASQGVLEPAFNALLYYS	6320479	(82-124)
Consensus/80%	..pp.hp.lbphaPphs...lb..L..pp.s.p.shp.lL..p		
2-structure	hHHHHHHHh HHHHHHHHHH HHHHHHh		

Figure 4 Multiple sequence alignment of CUE domains

Consensus and secondary-structure symbols are as in Figure 1. Residue numbers and GenPept or cosmid identifiers are shown following the alignment. Abbreviations: ARATH, *Arabidopsis thaliana*; CAEL, *Caenorhabditis elegans*; CANAL, *Candida albicans*; DROME, *Drosophila melanogaster*; FUGU, *Fugu rubripes*; LEIMA, *Leishmania major*; ORF, open reading frame; SCHPO, *Schizosaccharomyces pombe*; VPS9, vacuolar protein sorting-associated protein 9; YEAST, *Saccharomyces cerevisiae*. CUE domains are sorted into classes (I, II, III and IV), as discussed in the text.

three TPRs in Sec72p. TPRs are repeated structures that are present in most cellular organisms [34]. They adopt superhelical structures that offer an extended surface for protein–protein interactions [35,36]. The finding of three TPRs in Sec72p is of particular interest, given that Hop, a TPR-containing protein that is sequence-similar to Sec72p, is a co-chaperone that binds heat-shock cognate protein 70 (Hsc70), one of the Hsp70/DnaK family of chaperones that also include BiP and Kar2p. As the Hsc70-binding site in Hop is contained within its three N-terminal TPRs [37,38], it is expected that Sec72p binds a Hsp70 protein, possibly Kar2p, in the ER lumen via its three TPRs.

Mutant genes for Sec71p/Kar7p and Kar2p are associated with defects in nuclear fusion (karyogamy). Another such mutant is KAR8/JEM1, whose gene product contains a C-terminal DnaJ domain [39]. Detailed analysis of its sequence reveals that it too contains three TPRs (Figure 3). Consequently it is predicted that the TPRs in Sec72p and Kar8p/Jem1p serve to localize Hsp70/DnaK ATPase-dependent chaperone activity to Sec63p-associated complexes, thereby allowing the DnaJ domains of Sec63p (and Kar8p/Jem1p) to act as activators of Hsp70/DnaK ATPase activities.

Cue1p, a prototype of a novel domain family

Cue1p is a yeast protein that recruits Ubc7p to an ER membrane-associated complex that also includes Ubc6p [11]. In its absence, misfolded proteins remain un-ubiquitinated and are not translocated from the ER lumen. Thus Cue1p appears to be a scaffolding molecule that assembles Ubc7p and Ubc6p in the proximity of the translocon pore cytoplasmic exit. Although a Cue1p homologue is readily apparent in *S. pombe*, no metazoan homologue has hitherto been described. However, database searches demonstrate the presence of a novel domain structure (termed a 'CUE' domain) within the Cue1p sequence that is also present in several eukaryotic cytoplasmic proteins.

One of these CUE domain-containing proteins is AMFR, which is thought to be involved in tumour-cell invasion and metastasis [40]. The downstream function of AMFR upon binding AMF is unknown. However the finding of an AMFR CUE domain C-terminal to a previously recognized RING-finger domain [41] (Figure 2) suggests that AMFR recruits ubiquitination machinery via these two cytoplasmic domains. Really interesting new gene (RING)-finger domains have recently been recognized as binding UBC enzymes and acting as UBC-dependent ubiquitin ligases [42]. The molecular function of CUE domains, as assessed from investigations of Cue1p [11], remains unknown, but these structures may also involve recruitment of UBC enzymes.

AMFR is a marker for a subdomain of the smooth ER [43]. Although previously stated as being a seven-transmembrane (TM) protein [41], detailed analysis of the sequences of mammalian and *C. elegans* AMFR instead predicts five TM segments (results not shown). This number of TM segments and the presence of a RING finger in AMFR are identical with the architecture of the ER retrotranslocon-associated *Saccharomyces cerevisiae* protein Der3p/Hrd1p, whose RING finger is essential for ER-associated degradation [44]. Although AMFR is apparently not the mammalian orthologue of yeast Der3p/Hrd1p [a more closely similar sequence is apparent from a conceptual translation of an expressed sequence tag ('EST'), AA673624], it may nevertheless perform a cellular role similar to that of Der3p/Hrd1p. In addition, given that the domain architecture of metazoan AMFR is conceptually equivalent to a fusion of yeast Der3p/Hrd1p and Cue1p, it is predicted that the latter two proteins interact. This prediction is based on the proposition that

pairs of multidomain proteins physically interact if a fused counterpart is found in another genome [45].

Other CUE domain proteins are not expected to be associated with the ER. Tollip is an interleukin-1 receptor complex protein [46], fungal Vps9p proteins are guanine nucleotide exchange factors involved in vesicle-mediated vacuolar protein transport, and AUP1 is predicted, from significant similarity to tafazzin-like molecules ([47]; C. P. Ponting, unpublished work), to be an acyltransferase. These CUE domain-containing proteins, therefore, may recruit UBC enzymes to different membrane-associated complexes. The possibility that divergent CUE domains possess functions different from that of Cue1p, however, cannot be discounted.

Hrd3p and SEL-1 repeats

Recent studies have shown that yeast Hrd3p is a component of the ER translocon, by virtue of its genetic interactions with Der3p/Hrd1p and Sec61p [48]. The molecular function of Hrd3p remains elusive, although its sequence shows localized similarity to *C. elegans* SEL-1 [49], a negative regulator of LIN-12 [50], over approx. 40 amino acids. Techniques employed for this study reveal, however, that Hrd3p and SEL-1 contain significant similarities throughout their sequences and each contains multiple repeats, the 40-amino-acid sequence being but one of such repeats (Figure 5).

These approaches also determined that these so-called SEL-1 repeats are subtypes of TPRs, such as those in Sec72p and Kar8p/Jem1p (as above). SEL-1 repeats differ from standard TPRs in possessing a variable number of amino acids between the two TPR α -helices (Figure 5). Notwithstanding such differences, it remains possible that the SEL-1 repeats in Hrd3p possess a peptide-binding function similar to that of TPRs in Hop, and those predicted for Sec72p and Kar8p/Jem1p.

SEL-1 repeats have been found in several bacterial proteins [51]. These unusual occurrences are not uniformly distributed among the bacteria and are often most sequence-similar to eukaryotic versions. Thus it is proposed that these bacteria acquired SEL-1-like repeat genes from Eukarya, either directly or indirectly, via one or more horizontal gene-transfer events.

Ire1p and β -propellers

ER-associated degradation is known to be co-ordinated with the unfolded-protein-response pathway [52]. The transmembrane kinase and endoribonuclease Ire1p is a key member of the latter pathway, since deletion of its gene impairs protein degradation [53]. Ire1p spans the ER membrane with its N-terminal portion within the lumen, and with its catalytic kinase and endoribonuclease domains in the cytosol or nucleus. The function and structure of the N-terminal region of Ire1p have remained elusive, although it is thought to impart a sensing function [54].

PSI-BLAST analysis of the luminal sequence of Ire1p, and homologues such as interferon-inducible double-stranded RNA-dependent kinase-like ER kinase ('PERK'), shows that they are significantly similar to PQQ dehydrogenases (Figure 6). These enzymes contain a β -propeller structure containing eight four-stranded β -sheet structures termed 'blades'. Four such blade repeats were identified in yeast Ire1p and its mammalian and invertebrate homologues. Consequently, it is predicted that the Ire1p luminal region contains a β -propeller structure. It is likely that the number of blades in this structure is greater than four and that additional blade-like sequence repeats remain undetected.

HRD3_YEAST/1	SEATYTLSQLHLSQ--YNFPHNMTLHKYLEKFNDLTH	6323236	(103-139)
HRD3_YEAST/2	HSALFDLAVMYATGG(8)VIPQDSAKALLYYQRAAQLGN	6323236	(143-186)
HRD3_YEAST/3	LKAKQVLAYKYYSGF--NVPRNFHKSLLVLYRDIAEQLR	6323236	(187-222)
HRD3_YEAST/4	VRCLQLLGHMYFTGE--GSSKPNTHMAEILTTSLISR	6323236	(370-406)
HRD3_YEAST/5	GRACIDLGLINQY----ITNNISQALSYMMKAMKTQA	6323236	(413-445)
HRD3_YEAST/6	IPAIYEFVAVMIESGM--NSKSSVENTAYLFKTFVDKNE	6323236	(482-517)
HRD3_YEAST/7	AYLMYQLPYEFEDPP--RTTDQRKTLAISYYTRAFFKQGN	6323236	(559-595)
HRD3_YEAST/8	IDAGVVAGDIYFQ-----MQNYSKAMALYQGAALKYS	6323236	(596-627)
HRD3_YEAST/9	IQAIWNLGYMHEHGL---GVNRDFHLAKRYDQVSEHDH	6323236	(628-663)
SEL1_CAEEL/1	AEAEFQRMAYIERG-KGHGREGRVAHRVFERRAAQGH	7511639	(93-130)
SEL1_CAEEL/2	QEARKAVAFSQMFGD---YSRWSIQBAKTVFEDLEKNGS	7511639	(131-166)
SEL1_CAEEL/3	PDAQLALGFMHGAGI--GVEKSNQAKALVYMMFSAALGGN	7511639	(167-203)
SEL1_CAEEL/4	PLAQMAMGFRYSHGV---GVPQNCETALSYYQKVAKTVV	7511639	(204-239)
SEL1_CAEEL/5	TSAQLGLGQIYLAGG--RGLNQNFELAFRYLLAAAESGS	7511639	(289-325)
SEL1_CAEEL/6	ADALTYLGKMYLDGT--PFTPKDYQKSFYLMKSADKSS	7511639	(326-362)
SEL1_CAEEL/7	PSAQAVLGAMYMKGV---GVKKNYEKALKLLTSLADKKN	7511639	(363-398)
SEL1_CAEEL/8	ADGQMYLAELHYKGV(4)GVHRDFKKSVMKLYQLASQNGH	7511639	(399-438)
SEL1_CAEEL/9	ILAYYNLAQMHAAGT---GVPRSCSHAVDLFKSVARERKQ	7511639	(439-474)
SEL1_CAEEL/10	EVAQTNLAYILDRGE(7)PKDNNMERAFNLWQRSANQEQY	7511639	(507-549)
SEL1_CAEEL/11	AAARVKLGDDYYGYL---GTEVDHSLAFSNYKMAVDRHG	7511639	(550-585)
SEL1_CAEEL/12	AQAMFNLGYMHEVGE--GITRDLYLAKRFYDQAIHSQ	7511639	(587-622)
PPP5_HUMAN/1	AEELKTQANDYFK-----AKDYENAIKFYSQAIELNP	1709744	(28-59)
PPP5_HUMAN/2	AIYGNRSLAYLR-----TECYGALGDATRAIELDK	1709744	(62-93)
PPP5_HUMAN/3	IKGYRRAASNMA-----LGKFRALRDYETVVKVKP	1709744	(96-127)
HOP_HUMAN/4	ALKEKELGNDAYK-----KKDFDTALKHYDKAKELDP	7766910	(4-35)
HOP_HUMAN/5	MTYITNQAAVYFE-----KGDYKCRELCCKAEVGR	7766910	(38-69)
HOP_HUMAN/6	AKAYARIGNSYFK-----EEKYKDAIHFYKSLAEHR	7766910	(79-110)
Consensus/80%	..sbh.bs..bb.....psbp.Ab.hbp.shp...		
2-structure(HOP)	HHHHHHHHHHHHHHH	HHHHHHHHHHHHHHH	

Figure 5 Multiple sequence alignment of SEL-1-like repeats in SEL-1 and Hrd3p, and their comparison with six TPR sequences of known tertiary structure

Consensus and secondary-structure symbols are as in Figure 1. Residue numbers and GenPept identifiers are shown following the alignment. Abbreviations: CAEEL, *Caenorhabditis elegans*; HOP, Hop; PPP5, Ser/Thr protein phosphatase 5; YEAST, *Saccharomyces cerevisiae*.

IRE1_YEAST/1	ELSLSD---ILIAADVE---GGLHAVDR---RNGHIIWSTE	729857	(116-147)
IRE1_SCHPO/1	ESSLYN---AFIVATVD---GSLHSYDR---ITGOELWSLF	4164398	(78-109)
YQG4_CAEEL/1	EEKTSS---TILVSTID---GRLRALDS---ETGEIKWTLQ	1353135	(38-69)
IRE1_HUMAN/1	VTLPET---LLFVSTLD---GSLHAVSK---RTGSIKWTLK	3300094	(26-57)
IRE1_YEAST/2	GKIVED---EKVYTGSMR---TIMYTTINM---LNGEIIISAFG	729857	(216-248)
IRE1_SCHPO/2	PIRLLN---NNVFVSGKN---TTLFTTIDV---SNGDIVSQYP	4164398	(198-230)
YQG4_CAEEL/2	PCKGND---GILYAGSKK---DVWFGIDP---KTGLKVEYIL	1353135	(118-150)
IRE1_HUMAN/2	PCRSSD---GILYMGGKK---DIWYVIDL---LTGEKQQTLS	3300094	(108-140)
IRE1_YEAST/3	KLQECE---NMIVIGKTIFF---LGIHSYDGASYNVTYSTWQON	729857	(270-307)
IRE1_SCHPO/3	VTVSNK---SFVDIARTE---YFITIYSD(5)DLVYIDWTPT	4164398	(283-320)
YQG4_CAEEL/3	RICPANQ-KQTIFFLGRTE---YRVSMFDE---KNRGKTNAT	1353135	(172-206)
IRE1_HUMAN/3	SLCPST---SLLYLGRTE---YTITMYDT---KTRFLRWAT	3300094	(146-178)
IRE1_YEAST/4	NTFSKD---GMCIAFRPD---KSLLASDL---DFRIARWVSP	729857	(317-349)
IRE1_SCHPO/4	HSSSFD---SKLALSSYD---SSLHIVDT---HSKFIKQNIIP	4164398	(330-362)
YQG4_CAEEL/4	EVNTWP---FKHYASSSH---GYILTFDR---ETGEMRWEDQ	1353135	(217-249)
IRE1_HUMAN/4	DEGDYK---MSHFVSGND---GLVVTVDG---ESGDVLWION	3300094	(189-221)
QEDH_PSEAE/1	QAIVSD---GVLYVTASY---SRLFALDA---KTGKRLWTYN	7546210	(63-94)
QEDH_PSEAE/2	GAAIYG---DKVFFGTLD---ASVVALNK---NTGKVVWKKK	7546210	(112-144)
QEDH_PSEAE/3	APTIVK(6)VLLIHGSSG(5)VGRLFARDP---DTGEEIWMRP	7546210	(157-201)
QEDH_PSEAE/4	SASFDAET-NTLIIVGAGN(24)SGQVGVDP---SSGEVKWFYQ	7546210	(250-308)
QEDH_PSEAE/5	ELVLFD(6)KIVKATAHADRN-GFFYVVDG---SNGKLQNAFP	7546210	(322-363)
QEDH_PSEAE/6	PMAYSQDTGLFYVPANHW(32)GSLRAMDP---VSGKVVWEHK	7546210	(416-483)
QEDH_PSEAE/7	GVLATAG--NLVFTGTGD---GYFKAFDA---KSGKELWKFQ	7546210	(491-524)
Consensus/80%	p.....hbhsp.....bhshD....pssb..b.b.		
2-structure	EEE EEEEE EEEEE EEEEE		

Figure 6 Multiple alignment of β -propeller-like repeats in Ire1p and PQQ dehydrogenases

Consensus and secondary-structure symbols are as in Figure 1. Residue numbers and GenPept identifiers are shown following the alignment. Abbreviations: CAEEL, *Caenorhabditis elegans*; PSEAE, *Pseudomonas aeruginosa*; QEDH, quinoprotein ethanol dehydrogenase; SCHPO, *Schizosaccharomyces pombe*; YEAST, *Saccharomyces cerevisiae*; YQG4, *C. elegans* C41C4.4 gene product. The alignment represents the structural unit (β -propeller blade) observed in *P. aeruginosa* QEDH. The eighth blade in QEDH contains contributions from both N- and C-terminal regions; the structural repeats, therefore, are circular permutations of the sequence-based repeats.

The active-site residues of PQQ dehydrogenases are not conserved in Ire1p homologues, thereby indicating that Ire1p does not fulfil a similar enzymic role. Since Ire1p is thought to sense changes in the concentrations of unfolded proteins or unbound chaperones, it is likely to possess a protein-binding function. It has been noted that β -propellers tend to bind their substrates or ligands at or above the central chamber formed at the junction of the propeller blades [55,56]. It is likely, therefore, that the Ire1p β -propeller binds protein at a topologically equivalent site.

The seven-bladed β -propeller of prolyl oligopeptidase has been shown to contain a narrow central channel that is thought to exclude large folded molecules [57]. By analogy, the Ire1p β -propeller may also discriminate between folded and unfolded proteins by size exclusion from a central cavity positioned along its β -propeller axis.

Conclusions

The findings of the present study demonstrate that considerable insight into protein evolution, structure and function can be obtained if the full complement of sequence-analysis tools is employed to study functionally related protein sets. These methods have been used to detect two novel domain types: one present in Sec63p and RNA-dependent helicases, and a second in putative UBC-binding proteins. The detection of TPRs in Sec72p and the TPR-like SEL-1 repeats in Hrd3p, as well as the PQQ dehydrogenase-like β -propeller repeats in Ire1p, demonstrates the plurality of functions that homologues can possess.

Notwithstanding the multifunctionality of some homologues, this study generated sufficient evidence to predict, by analogy as well as homology, an Hsp70-binding function for Sec72p and a UBC-binding role for CUE domain-containing proteins. Moreover, it suggested, by analogy to other β -propeller proteins, a size-exclusion mechanism by which Ire1p may preferentially recognize unfolded over folded proteins. Results, however, were unable to suggest possible functions for the domain homologues that were unexpectedly identified in Sec63p and Brr2p. Whatever the successes of homologue detection in this and other sequence-based studies, it is clear that considerable experimental work will be required to investigate such sequence-based testable hypotheses of evolution, function and structure.

Sequence data for *Candida albicans* were generated at the Stanford DNA Sequencing and Technology Center with the support of the (U.S.) National Institute of Dental and Craniofacial Research and the Burroughs Wellcome Fund. C.P.P. is funded by the Medical Research Council (U.K.).

REFERENCES

- Ponting, C. P., Schultz, J., Copley, R. R., Andrade, M. A. and Bork, P. (2000) Evolution of domain families. *Adv. Protein Chem.* **54**, 185–244
- Kopito, R. R. (1997) ER quality control: the cytoplasmic connection. *Cell* **88**, 427–430
- Sommer, T. and Wolf, D. H. (1997) Endoplasmic reticulum degradation: reverse protein flow of no return. *FASEB J.* **11**, 1227–1233
- Johnson, A. E. and van Waes, M. A. (1999) The translocon: a dynamic gateway at the ER membrane. *Annu. Rev. Cell Dev. Biol.* **15**, 799–842
- Rapoport, T. A. (1992) Transport of proteins across the endoplasmic reticulum membrane. *Science* **258**, 931–936
- Johnson, A. E. (1997) Protein translocation at the ER membrane: a complex process becomes more so. *Trends Cell Biol.* **7**, 90–95
- Panzner, S., Dreier, L., Hartmann, E., Kostka, S. and Rapoport, T. A. (1995) Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. *Cell* **81**, 561–570
- Dobson, C. M. (1999) Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329–332
- Ward, C. L., Omura, S. and Kopito, R. R. (1995) Degradation of CFTR by the ubiquitin-proteasome pathway. *Cell* **83**, 121–127
- Chen, P., Johnson, P., Sommer, T., Jentsch, S. and Hochstrasser, M. (1993) Multiple ubiquitin-conjugating enzymes participate in the *in vivo* degradation of the yeast MAT α 2 repressor. *Cell* **74**, 357–369
- Biederer, T., Volkwein, C. and Sommer, T. (1997) Role of Cue1p in ubiquitination and degradation at the ER surface. *Science* **278**, 1806–1809
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
- Tatusov, R. L., Altschul, S. F. and Koonin, E. V. (1994) Detection of conserved segments in proteins: iterative scanning of sequence databases with alignment blocks. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12091–12095
- Holm, L. and Sander, C. (1998) Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics* **14**, 423–429
- Schultz, J., Milpetz, F., Bork, P. and Ponting, C. P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5857–5864
- Petrokovski, S. (1996) Searching databases of conserved sequence regions by aligning protein multiple alignments. *Nucleic Acids Res.* **24**, 3836–3845
- Rost, B. and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 584–599
- Kowronek, M. H., Rotter, M. and Haas, I. G. (1999) Molecular characterization of a novel mammalian DnaJ-like Sec63p homologue. *Biol. Chem.* **380**, 1133–1138
- Lee, T. G., Tang, N., Thompson, S., Miller, J. and Katze, M. G. (1994) The 58,000-dalton cellular inhibitor of the interferon-induced double-stranded RNA-activated protein kinase (PKR) is a member of the tetratricopeptide repeat family of proteins. *Mol. Cell. Biol.* **14**, 2331–2342
- Jacobsen, S. E., Binkowski, K. A. and Olszewski, N. E. (1996) SPINDLY, a tetratricopeptide repeat protein involved in gibberellin signal transduction in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9292–9296
- Keitel, T., Diehl, A., Knaute, T., Steszowski, J. J., Höhne, W. and Görsch, H. (2000) X-ray structure of the quinoprotein ethanol dehydrogenase from *Pseudomonas aeruginosa*: basis of substrate specificity. *J. Mol. Biol.* **297**, 961–974
- Schatz, P. J. and Beckwith, J. (1990) Genetic analysis of protein export in *Escherichia coli*. *Annu. Rev. Genet.* **24**, 215–248
- Gupta, R. S. and Singh, B. (1994) Phylogenetic analysis of 70kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. *Curr. Biol.* **4**, 1104–1114
- Cyr, D. M., Langer, T. and Douglas, M. G. (1994) DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. *Trends Biochem. Sci.* **19**, 176–181
- Lyman, S. K. and Schekman, R. (1997) Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BiP. *Cell* **88**, 85–96
- Plempner, R. K., Bohmler, S., Boddallo, J., Sommer, T. and Wolf, D. H. (1997) Mutant analysis links the translocon and BiP to retrograde protein transport for ER degradation. *Nature (London)* **388**, 891–895
- Feldheim, D., Rothblatt, J. and Schekman, R. (1992) Topology and functional domains of Sec63p, an endoplasmic reticulum membrane protein required for secretory protein translocation. *Mol. Cell. Biol.* **12**, 3288–3296
- Noble, S. M. and Guthrie, C. (1996) Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics* **143**, 67–80
- Lauber, J., Fabrizio, P., Teigelkamp, S., Lane, W. S., Hartmann, E. and Luhrmann, R. (1996) The HeLa 200 kDa U5 snRNP-specific protein and its homologue in *Saccharomyces cerevisiae* are members of the DEXH-box protein family of putative RNA helicases. *EMBO J.* **15**, 4001–4015
- Laggerbauer, B., Achsel, T. and Luhrmann, R. (1998) The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4188–4192
- Ragunathan, P. L. and Guthrie, C. (1998) RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8**, 847–855
- Rapoport, T. A., Jungnickel, B. and Kutay, U. (1996) Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* **65**, 271–303
- Green, N., Fang, H. and Walter, P. (1992) Mutants in three novel complementation groups inhibit membrane protein insertion into and soluble protein translocation across the endoplasmic reticulum membrane of *Saccharomyces cerevisiae*. *J. Cell Biol.* **116**, 597–604
- Andrade, M. A., Ponting, C. P., Gibson, T. J. and Bork, P. (2000) Homology-based method for identification of protein repeats using statistical significance estimates. *J. Mol. Biol.* **298**, 521–537
- Das, A. K., Cohen, P. W. and Barford, D. (1998) The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* **17**, 1192–1199

- 36 Groves, M. R. and Barford, D. (1999) Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.* **9**, 383–389
- 37 Van der Spuy, J., Kana, B. D., Dirr, H. W. and Blatch, G. L. (2000) Heat shock cognate protein 70 chaperone-binding site in the co-chaperone murine stress-inducible protein 1 maps to within three consecutive tetratricopeptide repeat motifs. *Biochem. J.* **345**, 645–651
- 38 Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. and Moarefi, I. (2000) Structure of TPR domain–peptide complexes: critical elements in the assembly of the Hsp70–Hsp90 multichaperone machine. *Cell* **101**, 199–210
- 39 Nishikawa, S. and Endo, T. (1997) The yeast JEM1p is a DnaJ-like protein of the endoplasmic reticulum membrane required for nuclear fusion. *J. Biol. Chem.* **272**, 12889–12892
- 40 Nabi, I. R., Watanabe, H. and Raz, A. (1992) Autocrine motility factor and its receptor: role in cell locomotion and metastasis. *Cancer Metastasis Rev.* **11**, 5–20
- 41 Shimizu, K., Tani, M., Watanabe, H., Nagamachi, Y., Niinaka, Y., Shiroishi, T., Ohwada, S., Raz, A. and Yokota, J. (1999) The autocrine motility factor receptor gene encodes a novel type of seven transmembrane protein. *FEBS Lett.* **456**, 295–300
- 42 Lorick, K. L., Jensen, J. P., Fang, S., Ong, A. M., Hatakeyama, S. and Weissman, A. M. (1999) RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11364–11369
- 43 Wang, H. J., Benlimame, N. and Nabi, I. R. (1997) The AMF-R tubule is a smooth ilimaquinone-sensitive subdomain of the endoplasmic reticulum. *J. Cell Sci.* **110**, 3043–3053
- 44 Bordallo, J. and Wolf, D. H. (1999) A RING-H2 finger motif is essential for the function of Der3/Hrd1 in endoplasmic reticulum associated protein degradation in the yeast *Saccharomyces cerevisiae*. *FEBS Lett.* **448**, 244–248
- 45 Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O. and Eisenberg, D. (1999) Detecting protein function and protein–protein interactions from genome sequences. *Science* **285**, 751–753
- 46 Burns, K., Clatworthy, J., Martin, L., Martinon, F., Plumpton, C., Maschera, B., Lewis, A., Ray, K., Tschopp, J. and Volpe, F. (2000) Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. *Nat. Cell Biol.* **2**, 346–351
- 47 Neuwald, A. F. (1997) Barth syndrome may be due to an acyltransferase deficiency. *Curr. Biol.* **7**, R465–R466
- 48 Plemper, R. K., Bordallo, J., Deak, P. M., Taxis, C., Hitt, C. and Wolf, D. H. (1999) Genetic interactions of Hrd3p and Der3p/Hrd1p with Sec61p suggest a retro-translocation complex mediating protein transport for ER degradation. *J. Cell Sci.* **112**, 4123–4134
- 49 Hampton, R. Y., Gardner, R. G. and Rine, J. (1996) Role of 26 S proteasome and *HRD* genes in the degradation of 3-hydroxy-3-methylglutaryl-CoA reductase, an integral endoplasmic reticulum membrane protein. *Mol. Biol. Cell* **7**, 2029–2044
- 50 Grant, B. and Greenwald, I. (1996) The *Caenorhabditis elegans sel-1* gene, a negative regulator of *lin-12* and *glp-1* encodes a predicted extracellular protein. *Genetics* **143**, 237–247
- 51 Ponting, C. P., Aravind, L., Schultz, J., Bork, P. and Koonin, E. V. (1999) Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* **289**, 729–745
- 52 Travers, K. J., Patil, C. K., Wodicka, L., Lockhart, D. J., Weissman, J. S. and Walter, P. (2000) Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell* **101**, 249–258
- 53 Casagrande, R., Stern, P., Diehn, M., Shamu, C., Osario, M., Zuniga, M., Brown, P. O. and Ploegh, H. (2000) Degradation of proteins from the ER of *S. cerevisiae* requires an intact unfolded protein response pathway. *Mol. Cell* **5**, 729–735
- 54 Shamu, C. E. and Walter, P. (1996) Oligomerization and phosphorylation of the Ire1p kinase during intracellular signalling from the endoplasmic reticulum to the nucleus. *EMBO J.* **15**, 3028–3039
- 55 Xia, Z. X., Dai, W. W., Xiong, J. P., Hao, Z. P., Davidson, V. L., White, S. and Matthews, F. S. (1992) The three-dimensional structures of methanol dehydrogenase from two methylotrophic bacteria at 2.6 Å resolution. *J. Biol. Chem.* **267**, 22289–22297
- 56 Russell, R. B., Sasieni, P. D. and Sternberg, M. J. E. (1998) Supersites within superfolds. Binding site similarity in the absence of homology. *J. Mol. Biol.* **282**, 903–918
- 57 Fülöp, V., Böcskei, Z. and Polgár, L. (1998) Prolyl oligopeptidase: an unusual β -propeller domain regulates proteolysis. *Cell* **94**, 161–170

Received 19 May 2000/20 July 2000; accepted 9 August 2000