ANTIOXIDANTS & REDOX SIGNALING
Volume 5, Number 4, 2003

Forum Original Research Communication

Defining the Domain Boundaries of the Human Protein Disulfide Isomerases

H.I. ALANEN,[1] K.E.H. SALO,[1] M. PEKKALA,[1] H.M. SIEKKINEN,[1] A. PIRNESKOSKI,[2] and L.W. RUDDOCK[1]

ABSTRACT

The protein disulfide isomerase (PDI) family of folding catalysts are constructed from combinations of redox-active and redox-inactive domains, all of which are probably based on the thioredoxin fold. To understand the function of each domain in the variety of catalytic reactions that each family member can perform (to differing extents), the domain boundaries of each family member must be known. By using a technique based on sequence alignments and the known structure of the a and b domains of human PDI, we generated a large number of domain constructs for all six redox-active human PDIs: PDI, PDIp, ERp72, ERp57, P5, and PDIr. The ability to generate significant amounts of soluble protein in *E. coli* from most of these domain constructs strongly indicates that the domain boundaries are correct. The implications for these domain boundaries on the tertiary structure of the human PDIs are discussed. *Antioxid. Redox Signal.* 5, 367–374.

INTRODUCTION

PROTEIN FOLDING IN THE ENDOPLASMIC RETICULUM (ER) is often associated with the formation of native disulfide bonds, which stabilize the tertiary and/or quaternary structure of the protein. The formation of native disulfide bonds is a complex process and is the rate-limiting step in the biogenesis of many secreted proteins (for reviews, see 9, 10). It is not suprising therefore that eukaryotic organisms have developed an extensive machinery designed to facilitate this process that possesses the flexibility required to aid the folding of a wide range of protein substrates.

The archetypal protein involved in disulfide bond formation is protein disulfide isomerase (PDI) (for reviews, see 8, 12). PDI was the first catalyst of protein folding to be identified, 40 years ago (13), but many questions remain unanswered about its mechanism of action. The problem is difficult because of the wide range and complexity of the physiological substrates and the absence of a complete three-dimensional structure for any PDI. PDI comprises four structural domains, **a**, **b**, **b′**, and **a′** plus a C-terminal acidic extension. The **e** domain, which was postulated to exist based on homology with an estrogen receptor (10), is not a separate domain, but is instead part of the **a** and **b** domains. The homologous **a** and **a′** domains of PDI, which contain the redox-active motif -WCGHC-, show significant sequence identity to thioredoxin, a small protein involved in many cytoplasmic redox functions (18, 19). To date, the structures of the **a** and **b** domains of PDI have been solved (21–23) and reveal the characteristic thioredoxin fold that is seen in all single-domain members of the thioredoxin superfamily (see 27).

Although multiple gene products play a role in disulfide bond formation *in vivo* (for example, in yeast the gene products of *PDI1*, *ERO1*, *EUG1*, *MPD1*, *MPD2*, and *EPS1* have all been implicated in the pathway for native disulfide bond formation; (for review, see 9), it is clear that not all of these are members of the PDI family. To date, there are six reported members of the PDI family in higher eukaryotes, which in-

[1]Biocenter Oulu and Department of Biochemistry, P.O. Box 3000, FIN-90014 University of Oulu, Oulu, Finland.
[2]Department of Biosciences, University of Kent, Canterbury Kent CT2 7NJ, U.K. Current address is Biocenter Oulu and Department of Biochemistry, P.O. Box 3000, FIN-90014 University of Oulu, Oulu, Finland.

clude a redox-active domain, PDI, PDIp (6), ERp72 (28), ERp57 (1), P5 (26), and PDIr (16). All six are multidomain proteins, being comprised of redox-active (**a**-like) and redox-inactive (**b**-like) domains.

From studies on domain fragments of human PDI (4, 24), it appears that the domains act synergistically, with the **a** and **a′** domains providing the ability to catalyze redox reactions and the **b′** domain providing the principle protein/peptide binding site. No clear function has yet been assigned to the **b** domain, but only when all four domains are present and in the correct order is significant catalysis of complex isomerization observed (*i.e.,* isomerization of disulfide bonds in a protein that must be accompanied by a significant conformational change in the substrate; 4). Many studies have been undertaken in order to understand the contributions of different parts of the PDIs to their overall catalytic abilities and in order to understand the differences observed between different PDI family members. These have included the generation of domain constructs (for example, see 4), chimeric proteins based on intraprotein domain shuffling (for example, see 33), and chimeric proteins based on interprotein domain shuffling (for example, see 30). In order to assess fully the results of any such studies, the domain boundaries of all of the members of the PDI family must be known. In an excellent review article by Ferrari and Söling (8), domain boundaries for all the human PDIs were suggested, but no information was given on how these were determined nor was any experimental evidence provided to confirm them.

To address the question of the domain boundaries of the human PDI family, a comparison was made between multiple sequence alignments of PDI family member domains and combined with the structural information available from NMR studies on the **a** and **b** domains of human PDI. Based on this information domain constructs were generated and tested for significant levels of soluble expression in *E. coli.* The results allow for the confirmation of the domain boundaries of the human PDI family.

MATERIALS AND METHODS

Gene inserts for expression vectors for mature human P5 (Leu20-Leu440), mature human PDIr (Ser22-Leu519), and mature human ERp72 (Val20-Leu645) were generated by PCR from a human liver cDNA library (Clontech), using primers that included an in-frame NdeI site 5′ to the first codon of the gene and either a BamHI site (ERp72) or a SalI site (PDIr, P5) after a TAA stop codon at the 3′-end. The inserts were cloned into pLWRP51, a modified version of pET23b (Novagen), which encodes for an N-terminal his-tag in frame with the cloned gene. The resulting gene products included the sequence MHHHHHHM- prior to the first amino acid of the mature sequence. Gene inserts for mature human PDI (Asp18-Leu508), for the **a** domain of human PDI (Asp18-Ala137), and for mature human ERp57 (Ser25-Leu505, with a silent mutation to remove the internal NdeI site) were subcloned as NdeI/BamHI fragments from existing vectors (3, 34) into pLWRP51. A gene insert for the **bb′a′c** domain construct of human PDI was generated previously (4) and subcloned as an NdeI/BamHI fragment into pET23b (Novagen). A gene insert for mature human PDIp (Gln22-Leu507) was subcloned by PCR from an existing vector (31) into pLWRP51. Silent mutations were made in amino acids R27, P29, R148, R149, and R150 to increase the final yield of expressed protein. Domain constructs for PDI domain N1–2, N1–3, and N1–4 were generated by introducing a stop codon into the plasmid encoding full-length PDI using the QuikChange mutagenesis kit (Stratagene). The remaining domain constructs for the PDI family members were generated by PCR using plasmids containing gene inserts encoding for mature proteins as templates and primers that included an in-frame NdeI site 5′ to the first codon of the gene and either a BamHI site (PDI, ERp57, ERp72) or a SalI site (PDIr, P5) after a TAA stop codon at the 3′-end. The inserts were cloned into pLWRP51. All plasmids generated were sequenced to ensure there were no errors in the cloned genes (see Table 2 for plasmid names, intermediate plasmids are not included).

RESULTS

Alignments of the amino acid sequences of the redox-active domains of the PDI family

Multiple alignments were done of the 14 catalytic domains of the human PDI family [PDI (NP_000909), [PDIp XP_007939), ERP57 (NP_005304), ERP72 (NP_004902), P5 (NP_005733), and PDIr (NP_006801)] using CLUSTAL W (17), T-COFFEE (29), Match-Box (5), and Multialin (2). The results from all four multiple alignments were different, and they were combined into a single alignment based on consensus alignments and using the known structure of the **a** domain of PDI to move gaps into loop regions and to align corresponding helix and strand regions without misaligning regions of good sequence similarity. The final alignment is shown in Fig. 1. All gaps/insertions are located in loop regions of the structure, with the exception of the gap in the second catalytic domains of PDI and PDIp at a position analogous to A91 of PDI, which is located in α 3.

Definition of the domain boundaries of the redox-active domains

Although the alignments of the redox-active domains can be used by themselves to define the domain boundaries, a more accurate method of making the definition is to combine it with known structural information. The nuclear magnetic resonance (NMR) structure of the **a** domain of human PDI has been solved (21). As this domain is representative of all of the redox-active domains, the structural boundaries from the NMR structure can be used in combination with the alignments to define the domain boundaries of the other domains.

The domain construct used for the NMR studies of the **a** domain was human PDI D18-A136 (21), structurally, however the first amino acid of β1 of the thioredoxin fold is V26 and the last amino acid of α 4 is T133. Hence, V26-T133 defines structurally the **a** domain of PDI. The N-termini of the other redox-active domains can be readily determined from the alignment (see Table 1). The C-termini are less easy to define as the similarity between domains is very low in the structure terminating α 4 region. There are two possible ways

TABLE 1. DOMAIN BOUNDARIES OF THE HUMAN PDIS FROM THE FERRARI AND SÖHLING REVIEW (8) AND FROM THIS STUDY*

Protein		*Ferrari and Söhling*	*This study*
PDI	Domain 1 (**a**)	23–130	(H) V26-T133 (G)
	Domain 2 (**b**)	135–230	(A) A137-N232 (Q)
	Domain 3 (**b′**)	235–344	(N) Q233-P368 (V)
	Domain 4 (**a′**)	368–471	(P) V369-S472 (G)
ERp57	Domain 1 (**a**)	25–129	(D) V27-A132 (G)
	Domain 2 (**b**)	134–237	(A) S136-F241(G)
	Domain 3 (**b′**)	241–354	(F) G242-P377 (V)
	Domain 4 (**a′**)	377–481	(P) V378-N486 (P)
PDIp	Domain 1 (**a**)	ND†	(G) I44-V151 (G)
	Domain 2 (**b**)	ND	(S) A155-H254 (S)
	Domain 3 (**b′**)	ND	(H) S255-P389 (V)
	Domain 4 (**a′**)	ND	(P) V390-N493 (G)
ERp72	Domain 1 (**a°**)	61–166	(G) V64-Q169 (P)
	Domain 2 (**a**)	175–280	(V) T179-S283 (G)
	Domain 3 (**b**)	286–390	(P) S287-L394 (P)
	Domain 4 (**b′**)	395–503	(L) P395-P525 (V)
	Domain 4 (**a′**)	526–632	(P) V527-L638 (S)
P5	Domain 1 (**a°**)	24–128	(D) V27-L139 (G)
	Domain 2 (**a**)	159–266	(D) V162-A273 (P)
	Other	275–386	See text
PDIr	Other	28–136	S25-G140
	Domain 2 (**a°**)	150–257	(V) V154-N258 (P)
	Domain 3 (**a**)	275–380	(S) V278-N381 (P)
	Domain 4 (**a′**)	396–502	(V) L400-E507 (G)

*The flanking amino acids of each Domain are indicated for the boundaries from this study. The **b′** domains listed from this study include interdomain regions.

†ND, not determined.

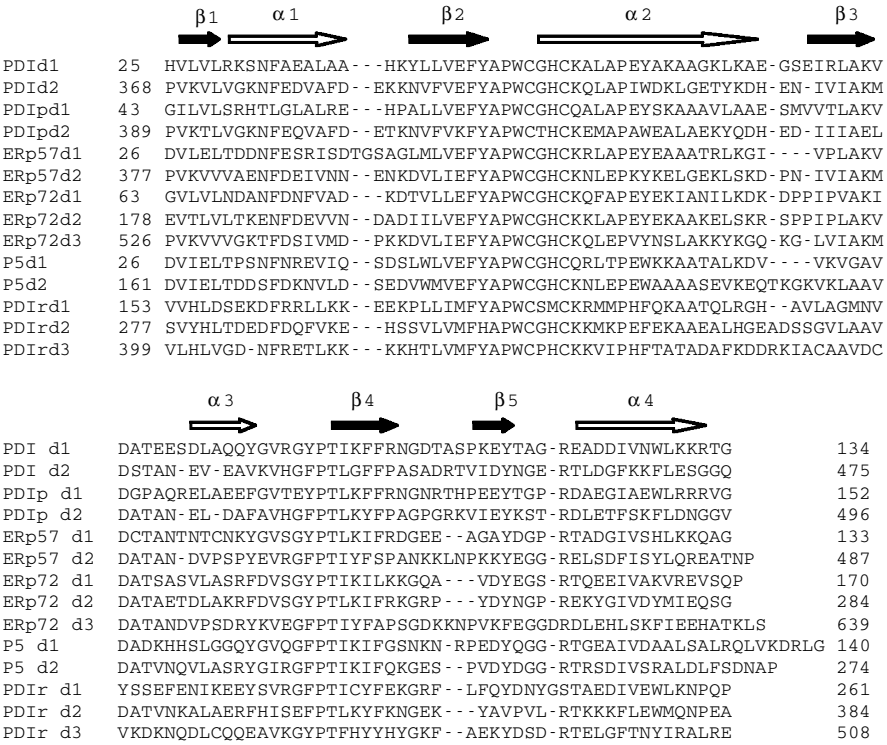

FIG. 1. Combined multiple alignments of the catalytic domains of the human PDI family. Mutiple alignments generated by CLUSTAL W (17), T-COFFEE (29), Match-Box (5), and Multialin (2) were combined based on consensus alignments and using the known structure of the **a** domain. The secondary structural elements of the **a** domain of PDI are also indicated. Numbering is for the full-length protein.

to define the C-termini, by simple alignment (counting amino acids from the last similar/identical amino acid if necessary) or by looking for helix-breaking residues (especially Gly, Pro, and Ser) in this region. As the alignments of the **a** and **a′** domains of human PDI, combined with the position of the structure-breaking sequence Gly473 Gly474 in the **a′** domain, indicate that different redox-active domains may have a variable-length α 4 the latter option was used. This gives the domain boundaries for the redox-active domains listed in Table 1. It should be noted that this method suggests that α 4 for both redox-active domains of P5 are slightly longer than for other family members and that the C-terminus of the last redox-active domain of ERp72 is more problematic to define.

Alignments of the non–redox-active domains of the PDI family

Multiple alignments were done of the four **b** domains of the human PDI family (PDI, PDIp, ERp57, and ERp72) using CLUSTAL W (17), T-COFFEE (29), Match-Box (5), and Multialin (2). The results from all four multiple alignments were different, and they were combined into a single alignment based on consensus alignments and using the known structure of the **b** domain of PDI to move gaps into loop regions and to align corresponding helix and strand regions without misaligning regions of good sequence similarity. All gaps/insertions are located in loop regions of the structure. The final alignment is shown in Fig. 2. When the first domain of PDIr and/or the putative **b**-like domain from P5 (8) were included in these multiple alignments, no significant similarity/identity could be observed (data not shown).

Definition of the domain boundaries of the redox-inactive domains

The NMR structure of the **b** domain of human PDI has been solved (22), and as this domain shows good sequence similarity/identity with the domains from ERp72, ERp57, and PDIp, this can be used to define the domain boundaries in these proteins. The domain construct used for the NMR studies of the **b** domain was human PDI A136-A245 (22); structurally, β1 of the thioredoxin fold starts with A137 and α 4 ends with N232. Hence, A137-N232 defines structurally the **b** domain of PDI. The N- and C-termini of the other **b** domains can be determined from this alignment (see Table 1).

Although the **b′** domain of human PDI has sequence similarity to the **b** domain to an extent where it is commonly believed to have the same fold (8, 12), the similarity/identity is low and insufficient to define accurately the domain boundaries. However, as this domain lies between the **b** and **a′** domains, the boundaries of the **b′** domain plus interdomain linkers can be deduced for PDI, PDIp, ERp57, and ERp72. These (see Table 1) were used to generate **b′** domain constructs.

Human PDIr and P5 clearly have at least one domain in addition to their redox-active domains. That for PDIr lies at the N-terminus of the protein. Although no significant sequence similarity can be found by multiple alignments with either the **b** or **b′** domains of the other human PDI family members, domain boundaries can be identified based on the N-terminus of the mature protein and on the existence of conserved interdomain linker region in this protein (see Discussion). Using these, the domain boundaries for PDIr domain 1 are S22-G140. P5 has at least one non–redox-active domain at the C-terminus of the protein, part of which has been reported to show similarity to the **b** domain of PDI (8). We find no significant sequence similarity to either the **b** or **b′** domains of the other human PDI family members, and a purified C-terminal fragment of human P5 shows a significantly different far-UV circular dichroism spectrum from other purified domains (Alanen and Ruddock, unpublished observations), suggesting it might not share the thioredoxin fold.

Generation and expression of domain constructs

The definition of any domain boundaries by alignment is prone to potential error, and, in addition, the positioning of an N-terminal amino function or C-terminal carboxyl function within the domain construct may cause folding and/or structural problems. Accordingly, many of the domain constructs were generated to include flanking amino acids, which would

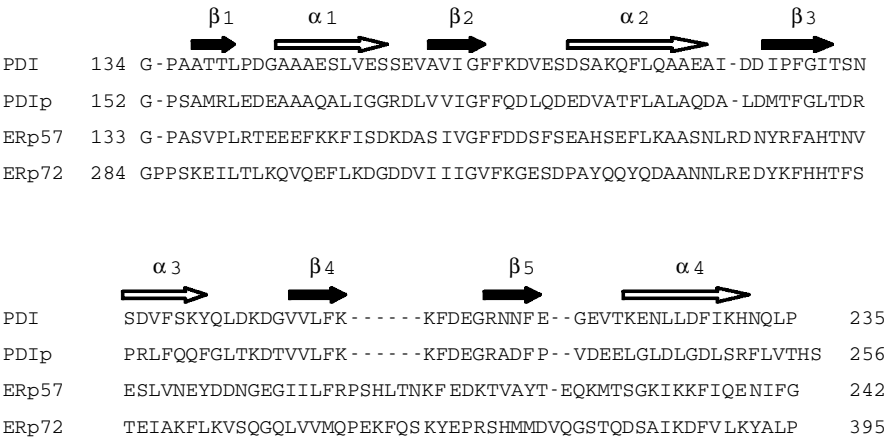

FIG. 2. Combined multiple alignments of the b domains of the human PDI family. Mutiple alignments generated by CLUSTAL W (17), T-COFFEE (29), Match-Box (5), and Multialin (2) were combined based on consensus alignments and using the known structure of the **b** domain. The secondary structural elements of the **b** domain of PDI are also indicated. Numbering is for the full-length protein.

TABLE 2. PLASMIDS USED IN THIS STUDY THAT GENERATE SIGNIFICANT LEVELS OF SOLUBLE PROTEIN*

Protein produced†		*Plasmid name*
PDI	Mature (D18-L508)	pLWRP64
	Domain N1 (D18-A136)	pLWRP69
	Domain 2+ (A136-A245)	pOLR48
	Domain 3 (K230-P368)	pLWRP64
	Domain +4+ (D365-G479)	pLWRP69
	Domain +4C (D365-L508)	pLWRP68
	Domain N1–2 (D18-L234)	pKEHS68
	Domain N1–3 (D18-P368)	pKEHS135
	Domain N1–4 (D18-G474)	pKEHS137
	Domain N1–4+ (D18-G479)	pKEHS136
	Domain 1–4C (D24-L508)	pKEHS44
	Domain 2–4C (A136-L508)	pLWRP16
	Domain 3–4C (L234-L508)	pLWRP32
ERp57	Mature (S25-L505)	pKEHS69
	Domain 1 (S25-P134)	pLWRP70A
	Domain 3 (G242-V378)	pLWRP73A
	Domain 1–3 (S25-V378)	pLWRP76
	Domain 1–4+ (S25-I490)	pOLR42
	Domain 2–4C (S136-L505)	pOLR57
	Domain 3–4C (G242-L505)	pOLR58
	Domain 2–3 (S136-V378)	pLWRP78
PDIp	Mature (Q22-L525)	pAP32
	Domain +1 (E32-G152)	pMPP1
	Domain 2 (P153-R257)	pMPP2
	Domain 4+ (P389-P498)	pMPP3
	Domain N1–2 (Q22-R257)	pKEHS70
	Domain N1–3 (Q22-P389)	pKEHS89
	Domain +3–4C (V252-L525)	pLWRP5
ERp72	Mature (V20-L645)	pHIA64
	Domain N1 (V20-P170)	pHIA54
	Domain 1 (N62-P170)	pHIA7
	Domain 2 (V178-P285)	pHIA80
	Domain 3 (P286-L394)	PHIA81
	Domain 5 (V527-L638)	pHIA82
	Domain 5C (V527-L645)	pHIA83
	Domain 4–5C (P395-L645)	pHIA47
	Domain 1–5C (N62-L645)	pHIA72
	Domain 2–3 (V178-L394)	pHIA52
P5	Mature (L20-L440)	pHIA34
	Domain N1 (L20-G140)	pHIA87
	Domain N1+ (L20-K160)	pHIA88
	Domain 2 (K160-P274)	pHIA89
	Domain 3 (P274-L440)	pHIA91
PDIr	Mature (S22-L519)	pHIA44
	Domain 1 (S22-G140)	pHMS1
	Domain 1+ (S22-K151)	pHMS2
	Domain 2 (D152-P261)	pOLR138
	Domain 2+ (D152-G276)	pOLR139
	Domain 3 (S277-P382)	pOLR140
	Domain 3+ (S277-T397)	pOLR141
	Domain 4C (S398-L519)	pOLR143

*Visualization by Coomassie staining of SDS-PAGE gels of total and soluble fractions after expression in the cytoplasm of *E. coli*. All proteins are N-terminally hexa-his tagged (except pLWRP16) with the numbering being for the natural full-length protein. Domains are listed in the order that they appear in the protein. Plasmids pLWRP5 and pLWRP6 have previously been reported (25), but the numbering for human PDIp has changed on publication of the full-length sequence (XP_007939).

†The prefix N indicates that this includes the N-terminal extension of the mature protein, and the prefix + that the construct's N-terminal extends more than three amino acids than that defined for the boundaries. The suffix C indicates that this version of the final domain includes the C-terminal extension of the mature protein; and the suffix + that the construct's C-terminus goes more than three amino acids than that defined for the boundaries.

be predicted to be within interdomain loops. After construction and verification by sequencing, each domain construct was expressed in the cytoplasm of *E. coli* [usually BL21 (DE3) containing the pLysS plasmid] by isopropyl β-D-1-thiogalactoside induction. After induction, the cells were pelleted and lysed, and total and soluble fractions from the lysates were run on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gels to test for soluble expression. The results are shown in Table 2. All of the individual domains of all six redox-active human PDIs were tested (except for the **a′** domain of ERp57); only three individual domains could not be expressed solubly in the cytoplasm of *E. coli*; these were the **b** domain of ERp57 (S136-F241) and the **b′** domains of ERp72 (P395-P526) and PDIp (M256-P389).

DISCUSSION

The definition of domain boundaries for multidomain proteins is often done by partial proteolysis of the mature protein (see 11 as example of human PDI). However, this technique is dependent on the location of suitable cleavage sites at the domain boundaries, the accessibility of these to the protease, and the requirement for the domain fragment generated to be relatively resistant to proteolysis. Alternatively, domain boundaries may be defined theoretically based on alignments between family members sharing homologous domains. This approach was adopted here and combined with information obtained on the structural limits of the **a** and **b** domains of human PDI as determined by NMR. Fortunately, these two are good representatives of the redox-active and redox-inactive domain modules from which the human PDI family is constructed.

By using these domain boundaries (see Table 1 for summary), a series of constructs were made and tested for expression in *E. coli*. As a domain is best defined as an independent structural unit that is able to fold in isolation from the rest of the protein and as only proteins that fold are made solubly in significant quantities when expressed in the cytoplasm of *E. coli*, any construct for which expression was observed must comprise a domain (plus possible N- and C-terminal extensions). Although some of the constructs were generated guided by previously reported domain constructs and show considerable overlap between adjacent domains (most notably some of the PDI constructs), many show no overlap between adjacent domains and hence must just define the domain plus possible short interdomain linkers.

The vast majority of the constructs generated show soluble expression in *E. coli* after induction that can easily be visualized by Coomassie staining of a soluble cell extract run on an SDS-PAGE gel. Hence, it is probable that the domain boundaries defined here are correct. Only three individual domains (of the 23 tested) cannot be expressed solubly in the cytoplasm of *E. coli*, these are the **b** domain of ERp57 and the **b′** domains of ERp72 and PDIp. Although it is possible that this represents inaccuracies in the definition of the domain boundaries, it should be noted that six different PDIp **b′** constructs have been generated, all of which are expressed insolubly (Salo and Ruddock, unpublished observations), and multidomain constructs including these domains, *e.g.,* ERp72 **b′a′c** and PDIp **b′a′c**, can be expressed solubly (see Table 1) and do not show significant degradation (Alanen and Ruddock, unpublished observations), which would be expected if the **b′** domain were not correctly folded. Instead this inability to generate soluble protein probably represents something related to structure/function of these domains in the protein (see below).

Although the structure of two representative domains of human PDI have been solved and others are in the process of being solved by NMR (*e.g.,* 7, 32), we currently have very little information on the organization of domains within the tertiary structure of any of the family members [we know that the active sites of PDI can get within 16 Å of each other (15)]. What is clear from the domain boundaries defined here is that between many of the domains there is only a very short linker region, *i.e.,* that the C-terminus of one domain and the N-terminus of another domain are probably linked by a small turn. This implies that there will be very little interdomain flexibility. The exceptions to this are between the **b′** and **a′** domains of PDI, ERp72, ERp57, and PDIp, between the first and second catalytic domains of ERp72 and P5, and between all of the domains of PDIr.

Freedman *et al.* (11) first postulated an interdomain linker between the **b′** and **a′** domains based on limited proteolysis of bovine PDI, an idea that is supported by the theoretical model proposed by Ferrari and Söling (8). Unfortunately, the homology between the **b** and **b′** domains of PDI is limited, and hence it is difficult to define where the **b′** domain ends and the interdomain linker begins, but there clearly is a longer linker region between **b′** and **a′** of up to 19 amino acids in length, and hence more potential flexibility than between the other domains. These linkers in PDI, PDIp, ERp57, and ERp72 show similarities (see Fig. 3). Apparently, there is a similar length linker between the first two catalytic domains of P5, but this shows no similarity to those found between the **b′** and **a′** domains (see Fig. 3). The other grouping of interdomain linkers are between eight and 15 amino acids long and are found between all four domains of PDIr and between the first two catalytic domains of ERp72. All are acidic, proline-rich, and contain a single tryptophan (see Fig. 3). All of these

```
PDI     350 KIKPHLMSQELPEDWDKQP 368
ERp57   360 NLKRYLKSEPIPESNDG-P 377
PDIp    356 QVKPYLLSQEIPPDWDQRP 374
ERp72   509 KLKPVIKSQPVPKN-NKGP 526


P5      140 GRSGGYSSGKQGRSDSSSKK 160


ERp72   170 PDWTPPPE         177
PDIr 1  141 PPLWEEDPGAK      151
PDIr 2  261 PQPQVPETPWADGG   276
PDIr 3  385 PPPPEPTWEEQQT    397
```

FIG. 3. Sequences of the interdomain linkers of the human PDI family. Group 1 consists of the putative linker found between the **b′** and **a′** domains of PDI, PDIp, ERp57, and ERp72. Group 2 consists of the linkers found between the first and second catalytic domains of ERp72 and all of the PDIr domains. The linker between the first and second catalytic domain of P5 clusters with neither group. Numbering is for the full-length protein.

linkers are predicted not to be in regular secondary structure and would possibly permit significant movement of domains with respect to each other.

The general phenomenom of short interdomain linkers implies that the domains probably remain in the same relative orientation with respect to each other, and it is likely that there are very significant interdomain contacts between domains, especially involving those domains located in the center of the sequence, *e.g.,* **b** and **b′** for PDIp. Often interdomain contacts in proteins are based on the burial of hydrophophic surfaces, and it is noteworthy that the three individual domains tested that could not be expressed solubly are the **b** domain of ERp57 and the **b′** domains of ERp72 and PDIp, and that these domains have a more hydrophobic character than the other domains found in the same proteins.

In addition to combinations of redox-active and -inactive thioredoxin-like modules, the human PDIs all appear to have N- and C-terminal extensions and a highly acidic region. Whereas the ER-retention motif at the very C-terminus of all six proteins has a clear function, the function of the other parts of the extensions are unknown. It is clear that these extensions are not required for correct folding (see Table 2 for examples) or for the PDI-like activity of these family members (for example, see 4; also Ruddock *et al*., unpublished observations). It is also apparent that many of these extensions are highly acidic (PDI C-terminus, ERp72 N-terminus, PDIp C- and N-termini, P5 C-terminus) and that of the human PDIs only ERp57 lacks regions of high negative charge (the interdomain linkers of PDIr are all acidic). Although the C-terminal extension of PDI is a putative calcium-binding domain (8), the physiological function of this acidic region is unknown, but it is noteworthy that the periplasmic thiol-disulfide oxidoreductase DsbA has a conserved acidic patch on the opposite face to its active site (14, 20).

The definition of the domain boundaries of the human PDIs should allow us to dissect out the function of individual domains and allow us to see more clearly the synergy that arises through domain–domain interactions in this important class of protein folding catalysts.

ACKNOWLEDGMENTS

The work was supported by Biocenter Oulu and Biotechnology and Biological Sciences Research Council grant 96/C13322.

ABBREVIATIONS

ER, endoplasmic reticulum; NMR, nuclear magnetic resonance; PAGE, polyacrylamide gel electrophoresis; PDI, protein disulfide isomerase; SDS, sodium dodecyl sulfate.

REFERENCES

1. Bennett CF, Balcarek JM, Varrichio A, and Crooke ST. Molecular cloning and complete amino-acid sequence of form-I phosphoinositide-specific phospholipase C. *Nature* 334: 268–270, 1988.
2. Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16: 10881–10890, 1988.
3. Darby NJ and Creighton TE. Characterization of the active site cysteine residues of the thioredoxin-like domains of protein disulfide isomerase. *Biochemistry* 34: 16770–16780, 1995.
4. Darby NJ, Penka E, and Vincentelli R. The multi-domain structure of protein disulfide isomerase is essential for high catalytic efficiency. *J Mol Biol* 276: 239–247, 1998.
5. Depiereux E, Baudoux G, Briffeuil P, Reginster I, De Bolle X, Vinals C, and Feytmans E. Match-Box_server: a multiple sequence alignment tool placing emphasis on reliability. *Comput Appl Biosci* 13: 249–256, 1997.
6. DeSilva MG, Lu J, Donadel G, Modi WS, Xie H, Notkins AL, and Lan MS. Characterization and chromosomal localization of a new-protein disulfide-isomerase, PDIp, highly expressed in human pancreas. *DNA Cell Biol* 15: 9–16, 1996.
7. Dijkstra K, Karvonen P, Pirneskoski A, Koivunen P, Kivirikko KI, Darby NJ, van Straaten M, Scheek RM, and Kemmink J. Assignment of ^{1}H, ^{13}C and ^{15}N resonances of the a′ domain of protein disulfide isomerase. *J Biomol NMR* 14: 195–196, 1999.
8. Ferrari DM and Söling H-D. The protein disulfide-isomerase family: unravelling a string of folds. *Biochem J* 339: 1–10, 1999.
9. Frand AR, Cuozzo JW, and Kaiser CA. Pathways for protein disulfide bond formation. *Trends Cell Biol* 10: 203–210, 2000.
10. Freedman RB, Hirst TR, and Tuite MF. Protein disulfide isomerase: building bridges in protein folding. *Trends Biochem Sci* 19: 331–336, 1994.
11. Freedman RB, Gane PJ, Hawkins HC, Hlodan R, McLaughlin SH, and Parry JW. Experimental and theoretical analyses of the domain architecture of mammalian protein disulfide-isomerase. *Biol Chem* 379: 321–328, 1998.
12. Freedman RB, Klappa P, and Ruddock LW. Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. *EMBO Rep* 3: 136–140, 2002.
13. Goldberger RF, Epstein CJ, and Anfinsen CB. Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. *J Biol Chem* 238: 628–635, 1963.
14. Guddat LW, Bardwell JC, and Martin JL. Crystal structures of reduced and oxidized DsbA: investigation of domain motion and thiolate stabilization. *Structure* 6: 757–767, 1998.
15. Hawkins HC, de Nardi M, and Freedman RB. Redox properties and cross-linking of the dithiol/disulfide active sites of mammalian protein disulfide-isomerase. *Biochem J* 275: 349–353, 1991.
16. Hayano T and Kikuchi M. Molecular cloning of the cDNA encoding a novel protein disulfide isomerase-related protein (PDIR). *FEBS Lett* 372: 210–214, 1995.
17. Thompson JD, Higgins DG, and Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific

gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680, 1994.
18. Holmgren A. Thioredoxin and glutaredoxin systems. *J Biol Chem* 264: 13963–13966, 1989.
19. Holmgren A. Antioxidant function of thioredoxin and glutaredoxin systems. *Antioxid Redox Signal* 2: 811–820, 2000.
20. Hu SH, Peek JA, Rattigan E, Taylor RK, and Martin JL. Structure of TcpG, the DsbA protein folding catalyst from *Vibrio cholerae*. *J Mol Biol* 268: 137–146, 1997.
21. Kemmink J, Darby NJ, Dijkstra K, Nilges M, and Creighton TE. Structure determination of the N-terminal thioredoxin-like domain of protein disulfide-isomerase using multidimensional heteronuclear C-13/N-15 NMR-spectroscopy. *Biochemistry* 35: 7684–7691, 1996.
22. Kemmink J, Darby NJ, Dijkstra K, Nilges M, and Creighton TE. The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules. *Curr Biol* 7: 239–245, 1997.
23. Kemmink J, Darby NJ, Dijkstra K, Nilges M, and Creighton TE. The structure in solution of the b domain of protein disulfide isomerase. *J Biomol NMR* 13: 357–368, 1999.
24. Klappa P, Ruddock LW, Darby NJ, and Freedman RB. The b′ domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins. *EMBO J* 17: 927–935, 1998.
25. Klappa P, Freedman RB, Langenbuch M, Lan MS, Robinson GK, and Ruddock LW. The pancreas-specific protein disulfide-isomerase PDIp interacts with a hydroxyl group in ligands. *Biochem J* 354: 553–559, 2001.
26. Lundstrom-Ljung J, Birnbach U, Rupp K, Soling HD, and Holmgren A. Two resident ER-proteins, CaBP1 and CaBP2, with thioredoxin domains, are substrates for thioredoxin reductase: comparison with protein disulfide isomerase. *FEBS Lett* 357: 305–308, 1995.
27. Martin JL. Thioredoxin—a fold for all reasons. *Structure* 3: 245–250, 1995.
28. Mazzarella RA, Srinivasan M, Haugejorden SM, and Green M. ERp72, an abundant luminal endoplasmic reticulum protein, contains three copies of the active site sequences of protein disulfide isomerase. *J Biol Chem* 265: 1094–1101, 1990.
29. Notredame C, Higgins DG, and Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302: 205–217, 2000.
30. Pirneskoski A, Ruddock LW, Klappa P, Freedman RB, Kivirikko KI, and Koivunen P. Domains b′ and a′ of protein disulfide isomerase fulfill the minimum requirement for function as a subunit of prolyl 4-hydroxylase. The N-terminal domains a and b enhance this function and can be substituted in part by those of ERp57. *J Biol Chem* 276: 11287–11293, 2001.
31. Ruddock LW, Freedman RB, and Klappa P. Specificity in substrate binding by protein folding catalysts: tyrosine and tryptophan residues are the recognition motifs for the binding of peptides to the pancreas-specific protein disulfide isomerase PDIp. *Protein Sci* 9: 758–764, 2000.
32. Silvennoinen L, Karvonen P, Koivunen P, Myllyharju J, Kivirikko K, and Kilpelainen I. Assignment of ^{1}H, ^{13}C and ^{15}N resonances of the a′ domain of Erp57. *J Biomol NMR* 20: 385–386, 2001.
33. Xiao R, Solovyov A, Gilbert HF, Holmgren A, and Lundstrom-Ljung J. Combinations of protein-disulfide isomerase domains show that there is little correlation between isomerase activity and wild-type growth. *J Biol Chem* 276: 27975–27980, 2001.
34. Zapun A, Darby NJ, Tessier DC, Michalak M, Bergeron JJ, and Thomas DY. Enhanced catalysis of ribonuclease B folding by the interaction of calnexin or calreticulun with ERp57. *J Biol Chem* 273: 6009–6012, 1998.

Address reprint requests to:
Dr. L.W. Ruddock
Biocenter Oulu and Department of Biochemistry
P.O. Box 3000
FIN-90014 University of Oulu
Oulu, Finland

E-mail: lloyd.ruddock@oulu.fi

Received for publication February 14, 2003; accepted May 9, 2003.

This article has been cited by:

1. Chao Wang , Wei Li , Jinqi Ren , Jingqi Fang , Huimin Ke , Weimin Gong , Wei Feng , Chih-chen Wang . Structural Insights into the Redox-Regulated Dynamic Conformations of Human Protein Disulfide Isomerase. *Antioxidants & Redox Signaling*, ahead of print. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links] [Supplemental material]
2. Adam M. Benham . 2012. The Protein Disulfide Isomerase Family: Key Players in Health and Disease. *Antioxidants & Redox Signaling* **16**:8, 781-789. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]
3. Anna-Kaisa Lappi, Lloyd W. Ruddock. 2011. Reexamination of the Role of Interplay between Glutathione and Protein Disulfide Isomerase. *Journal of Molecular Biology* **409**:2, 238-249. [CrossRef]
4. Van Dat Nguyen, Mirva J. Saaranen, Anna-Riikka Karala, Anna-Kaisa Lappi, Lei Wang, Irina B. Raykhel, Heli I. Alanen, Kirsi E.H. Salo, Chih-chen Wang, Lloyd W. Ruddock. 2011. Two Endoplasmic Reticulum PDI Peroxidases Increase the Efficiency of the Use of Peroxide during Disulfide Bond Formation. *Journal of Molecular Biology* **406**:3, 503-515. [CrossRef]
5. Guennadi Kozlov, Pekka Määttänen, David Y. Thomas, Kalle Gehring. 2010. A structural overview of the PDI family of proteins. *FEBS Journal* **277**:19, 3924-3936. [CrossRef]
6. Guennadi Kozlov, Simon Azeroual, Angelika Rosenauer, Pekka Määttänen, Alexey Yu. Denisov, David Y. Thomas, Kalle Gehring. 2010. Structure of the Catalytic a0a Fragment of the Protein Disulfide Isomerase ERp72. *Journal of Molecular Biology* **401**:4, 618-625. [CrossRef]
7. Anna-Riikka Karala, Lloyd W. Ruddock. 2010. Bacitracin is not a specific inhibitor of protein disulfide isomerase. *FEBS Journal* **277**:11, 2454-2462. [CrossRef]
8. Anna-Riikka Karala, Anna-Kaisa Lappi, Lloyd W. Ruddock. 2010. Modulation of an Active-Site Cysteine pKa Allows PDI to Act as a Catalyst of both Disulfide Bond Formation and Isomerization. *Journal of Molecular Biology* **396**:4, 883-892. [CrossRef]
9. Mirva J. Saaranen , Anna-Riikka Karala , Anna-Kaisa Lappi , Lloyd W. Ruddock . 2010. The Role of Dehydroascorbate in Disulfide Bond Formation. *Antioxidants & Redox Signaling* **12**:1, 15-25. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]
10. Anne Katrine Wallis, Ateesh Sidhu, Lee J. Byrne, Mark J. Howard, Lloyd W. Ruddock, Richard A. Williamson, Robert B. Freedman. 2009. The ligand-binding bâ² domain of human protein disulphide-isomerase mediates homodimerization. *Protein Science* **18**:12, 2569-2577. [CrossRef]
11. Feras Hatahet , Lloyd W. Ruddock . 2009. Protein Disulfide Isomerase: A Critical Evaluation of Its Function in Disulfide Bond Formation. *Antioxidants & Redox Signaling* **11**:11, 2807-2850. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]
12. Van Dat Nguyen, Katrine Wallis, Mark J. Howard, Antti M. Haapalainen, Kirsi E.H. Salo, Mirva J. Saaranen, Ateesh Sidhu, Rik K. Wierenga, Robert B. Freedman, Lloyd W. Ruddock, Richard A. Williamson. 2008. Alternative Conformations of the x Region of Human Protein Disulphide-Isomerase Modulate Exposure of the Substrate Binding b' Domain. *Journal of Molecular Biology* **383**:5, 1144-1155. [CrossRef]
13. I. Raykhel, H. Alanen, K. Salo, J. Jurvansuu, V. D. Nguyen, M. Latva-Ranta, L. Ruddock. 2007. A molecular specificity code for the three mammalian KDEL receptors. *The Journal of Cell Biology* **179**:6, 1193-1204. [CrossRef]
14. Anna–Riikka Karala , Lloyd W. Ruddock . 2007. Does S-Methyl Methanethiosulfonate Trap the Thiol–Disulfide State of Proteins?. *Antioxidants & Redox Signaling* **9**:4, 527-531. [Abstract] [Full Text PDF] [Full Text PDF with Links]
15. Agnes Görlach , Peter Klappa , Dr. Thomas Kietzmann . 2006. The Endoplasmic Reticulum: Folding, Calcium Homeostasis, Signaling, and Redox Control. *Antioxidants & Redox Signaling* **8**:9-10, 1391-1418. [Abstract] [Full Text PDF] [Full Text PDF with Links]

16. Christian W. Gruber, Maša #emažar, Begoña Heras, Jennifer L. Martin, David J. Craik. 2006. Protein disulfide isomerase: the structure of oxidative folding. *Trends in Biochemical Sciences* **31**:8, 455-464. [CrossRef]
17. Peter Morin Nissom, Siaw Ling Lo, Jennifer Chi Yi Lo, Peh Fern Ong, Justin Wee Eng Lim, Keli Ou, Rosa Cynthia Liang, Teck Keong Seow, Maxey Ching Ming Chung. 2006. Hcc-2, a novel mammalian ER thioredoxin that is differentially expressed in hepatocellular carcinoma. *FEBS Letters* **580**:9, 2216-2226. [CrossRef]
18. H.I. Alanen , K.E.H. Salo , A. Pirneskoski , Dr. L.W. Ruddock . 2006. pH Dependence of the Peptide Thiol-Disulfide Oxidase Activity of Six Members of the Human Protein Disulfide Isomerase Family. *Antioxidants & Redox Signaling* **8**:3-4, 283-291. [Abstract] [Full Text PDF] [Full Text PDF with Links]
19. Adam M. Benham . 2005. Oxidative Protein Folding: An Update. *Antioxidants & Redox Signaling* **7**:5-6, 835-838. [Citation] [Full Text PDF] [Full Text PDF with Links]
20. Marcus Mayer, Stephan Frey, Peppi Koivunen, Johanna Myllyharju, Johannes Buchner. 2004. Influence of the Oxidoreductase ERp57 on the Folding of an Antibody Fab Fragment. *Journal of Molecular Biology* **341**:4, 1077-1084. [CrossRef]
21. A.K. Lappi, M.F. Lensink, H.I. Alanen, K.E.H. Salo, M. Lobell, A.H. Juffer, L.W. Ruddock. 2004. A Conserved Arginine Plays a Role in the Catalytic Cycle of the Protein Disulphide Isomerases. *Journal of Molecular Biology* **335**:1, 283-295. [CrossRef]
22. Adam Benham . 2003. Oxidative Protein Folding: Recent Advances and Some Remaining Challenges. *Antioxidants & Redox Signaling* **5**:4, 355-357. [Citation] [Full Text PDF] [Full Text PDF with Links]