Identifying and Characterizing a Structural Domain of Protein Disulfide Isomerase

Nigel J. Darby, Johan Kemmink, and Thomas E. Creighton*

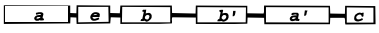
European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany

Received April 1, 1996[⊗]

ABSTRACT: Protein disulfide isomerase (PDI) appears on the basis of its primary structure to be a multidomain protein, but the number and nature of the domains has been uncertain. Two of the domains, *a* and *a*′, which are homologous to thioredoxin and active in catalysis of disulfide bond formation, have been identified and characterized previously. Sections of the N-terminal half of the PDI sequence have been expressed and the limits of their folded structures delineated by limited proteolysis. In addition to the *a*-domain, the boundaries of a domain with no activity on thiol/disulfide groups, designated *b*, have been identified. This domain has been produced independently; its cooperative unfolding transition and its CD and NMR spectra confirm that it is an autonomously folded structure in isolation and when part of PDI. Fusion of the *b*-domain to the *a*-domain, as occurs naturally in the first half of PDI, did not alter substantially the catalytic activity of the *a*-domain. It still catalyzes only a subset of the thiol/disulfide exchange reactions of intact PDI and has a reduced ability to catalyze protein disulfide rearrangements. The *a*- and *b*-domains account structurally for virtually all of the first half of the PDI polypeptide chain, and it is very unlikely that there exists a proposed third domain homologous to the estrogen receptor. The *b*-domain exhibits some sequence homology to calsequestrin, a calcium binding protein from the sarcoplasmic reticulum of muscle.

Protein disulfide isomerase (PDI)[1] is a 55 kDa multifunctional protein of the endoplasmic reticulum that is involved in the folding of disulfide-containing proteins (Freedman, 1992, 1995; Freedman et al., 1994). It catalyzes the formation, rearrangement, and breakage of disulfide bonds, which are often rate-limiting processes during protein folding (Givol et al., 1964; Creighton et al., 1980, 1993; Darby et al., 1994), using two active site -Cys-Gly-His-Cys- sequences that are similar to those of thioredoxin and a group of related proteins (Edman et al., 1985; Hawkins & Freedman, 1991; Vuori et al., 1992; La Mantia & Lennarz, 1993; Lyles & Gilbert, 1994). PDI is also an essential subunit of prolyl-4-hydroxylase (Pihlajaniemi et al., 1987) and of the microsomal triglyceride transfer complex (Wetterau et al., 1990, 1991); in both cases, it appears to have primarily structural rather than catalytic roles.

Elucidation of the primary structure of the protein gave the first indication that it probably consists of multiple domains (Edman et al., 1985). Examination of intron−exon boundaries, sequence homologies, both internally and to other proteins, and some limited proteolysis studies suggested the following model of the domain structure of PDI (Freedman et al., 1994).



(1)

The *a*- and *a*′-domains are homologous to thioredoxin and contain the active site -Cys-Gly-His-Cys- sequences. The central section of the molecule consists of two portions of sequence that are homologous to each other, which were designated the *b*- and *b*′-domains (Edman et al., 1985). Subsequently, the sequence between the *a*- and *b*-domains, residues 101−144, was reported to be similar to part of the ligand binding domain of the human estrogen receptor and designated the *e*-domain (Tsibris et al., 1989; Freedman et al., 1994). At the C terminus of the molecule is a sequence of predominantly acidic residues, the *c*-domain, which is a common feature of proteins that bind calcium (Lucero et al., 1994), and a KDEL sequence that controls retention of the protein in the endoplasmic reticulum (Haugejorden et al., 1991).

The structure and function of PDI is being dissected by preparing the putative domains using protein engineering methods. Initially, versions of the *a*- and *a*′-domains were prepared that contained residues Asp1 to Ala120 and Asp348 to Gly462, respectively (Darby & Creighton, 1995a). Both proteins were folded, although the *a*′-domain was less stable than expected. Nevertheless, both individual proteins had the active site properties expected from studies on the intact protein (Darby & Creighton, 1995b), although they catalyzed only a subset of the reactions catalyzed by PDI (Darby & Creighton, 1995a).

The structure of the *a*-domain has now been determined by NMR (Kemmink et al., 1995, 1996), and it raises questions about the proposed model of the PDI domain structure, in particular the significance of the putative *e*-domain. The last 20 residues of the polypeptide fragment

* Author to whom correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* July 15, 1996.

[1] Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; CD, circular dichroism; GSH and GSSG, glutathione in its reduced and disulfide forms respectively; P_{2SH}^{27SH} and P_S^S, peptide substrate with cysteine residues at positions 2 and 27 in the dithiol and disulfide forms, respectively; P_{2OH}^{27SH}, peptide substrate with a single cysteine residue at position 27 and a Ser residue at position 2; the mixed disulfide forms of the cysteine residues of these peptides with glutathione are denoted by -SSG sub- and superscripts; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; PDI, protein disulfide isomerase; PDI.C380S,C383A, PDI with Cys380 and Cys383 changed to Ser and Ala, respectively; PMSF, phenylmethanesulfonyl fluoride; SDS−PAGE, sodium dodecyl sulfate−PAGE; TFA, trifluoroacetic acid; TLCK, N_α-*p*-tosyl-L-lysine chloromethyl ketone; TOCSY, total correlation spectroscopy.

S0006-2960(96)00763-5 CCC: $12.00

containing the *a*-domain correspond to the first half of the putative *e*-domain, yet most of them appear to be an intrinsic part of the *a*-domain folded structure, forming an α-helix that packs against the central core of the molecule, just as in the homologous structure of thioredoxin. This makes it unlikely that the segment denoted the *e*-domain could be an autonomous structural unit.

The domain structure of PDI has important implications for understanding its structure and functions. Studies on the PDI *a*- and *a*′-domains have indicated that other parts of PDI are required for the full range of PDI activities to be manifest, in particular its activity in isomerizing protein disulfide bonds (Darby & Creighton, 1995a). Further studies on engineered proteins with various combinations of PDI domains should determine what other aspects of the PDI molecule are required for this. The role of PDI in other functions not related to catalysis of thiol−disulfide interchange, such as in assembly of prolyl-4-hydroxylase, in the triglyceride transferase complex, and in calcium binding, probably also depend upon the non-thioredoxin-like parts of the PDI sequence. Finally, given the lack of structural information for intact PDI, the identification, expression and structure determination by NMR of isolated PDI domains is one route toward determining the structure of this molecule. To date the only detailed structural data available has come from this approach (Kemmink et al., 1995, 1996).

In this report the relationship between the proposed *a*-, *e*-, and *b*-domains of PDI is clarified using the combined approach of protein engineering and limited proteolysis. The data obtained define the limits of the *b*-domain and suggest that the putative *e*-domain (Tsibris et al., 1989) is not a structural entity and instead forms parts of the *a*- and *b*-domains. On the basis of the boundaries of these two domains, a polypeptide chain corresponding almost solely to the *b*-domain has been prepared. It is shown to be useful for further structural and biochemical analysis.

MATERIALS AND METHODS

Materials. The cloned gene for human PDI (Pihlajaniemi et al., 1987) was kindly provided as cDNA clone S-138 by K. Kivirikko (University of Oulu, Finland). BPTI was a generous gift of Bayer AG. The preparation of intact human PDI and the PDI *a*-domain have been described previously (Darby & Creighton, 1995a).

Construction of the Expression Vectors for PDI Domains and Domain Combinations. The amino acid sequences of each polypeptide construct used in this study are defined in Table 1. Each is depicted by the hypothetical domains that it was expected to contain, followed by a subscript giving the residue numbers of the sequence of mature human PDI it contains. Construction of the expression vector for A_{1-120} using cDNA clone S-138 has been described previously (Darby & Creighton, 1995a). An identical approach was used for preparation of the DNA for all the constructs described here. This produced expression vectors in which the target protein was expressed directly with an initiating methionine residue under control of the pET12a T7 promoter (Studier et al., 1990). For reasons that were not established, polypeptide $B_{148-257}$ could not be expressed in this way, however, so an alternative expression system using the same general approach was developed. In this instance, the DNA insert was prepared by PCR from S-138 using two primers that allowed insertion of flanking *Bam*HI sites. The primer complementary to the N-terminus of the protein also included an initiating methionine codon, while the other primer carried a stop codon. The insert was cloned into the *Bam*HI site of pET3a (Studier et al., 1990), and recombinants of the correct orientation were selected by PCR between internal and flanking primers. This construct encodes a protein that contains the first 12 residues of the highly expressed T7 major capsid protein, followed by the $B_{148-257}$ coding sequence. As the latter sequence contains no methionine residues, the fusion sequence could be removed by cleavage with cyanogen bromide.

Table 1: PDI Fragments Expressed in This Study[a]

fragment	residues of PDI included
A_{1-120}	∗Asp1 to Ala120
$E_{100-157}$	(∗)Thr100 to Gln157
AE_{1-157}	∗Asp1 to Gln157
$EB_{100-257}$	Thr100 to Asn257
$B_{148-257}$	Glu148 to Asn257
$B_{119-228}$	(∗)Ala119 to Ala228
AEB_{1-257}	∗Asp1 to Ala257

[a] The numbering scheme is based on the mature sequence of PDI. It does not include the initiating N-terminal methionine residue of each expressed protein, which was present in the purified protein when indicated by an asterisk; parentheses indicate that the methionine residue was present in only a fraction of the molecules.

PDI.C380S,C383A, a version of the PDI molecule in which the active site sequence of the *a*′-domain is changed from *Cys*-Gly-His-*Cys* to *Ser*-Gly-His-*Ala*, was prepared by oligonucleotide-mediated mutagenesis of the target sequence in S-138 (Sarkar & Sommer, 1990), followed by recloning of the appropriate DNA fragment into the PDI expression vector via two unique restriction sites. The identities of all expression vectors were verified by nucleotide sequencing.

Gene Expression. Protein production was carried out in *Escherichia coli* strain BL21 (DE3), which also contained the pLysS plasmid to control leak through expression, as described previously (Darby & Creighton, 1995a). Cells were grown in LB medium at 37 °C unless otherwise stated. In each case, expression of the constructs was readily apparent from the increase in amount of a protein of about the expected molecular weight on SDS−PAGE. The identities of all polypeptide chains prepared in this way were confirmed by mass spectrometry.

Purification of PDI.C380S,C383A. PDI.C380S,C383A was purified as previously described for the normal protein (Darby & Creighton, 1995a).

Purification of Polypeptide $E_{100-157}$. Cells were grown in minimal media containing ammonium chloride and glucose as the nitrogen and carbon sources, respectively. The pH of the cell lysate was adjusted to pH 8.5 and loaded onto a Q-Sepharose column equilibrated with 50 mM Tris (pH 8.5). The protein was eluted with a gradient from 0 to 0.15 M sodium chloride. The pooled fractions were dialyzed against 50 mM Tris (pH 7.4) and then stored at −20 °C.

Purification of Polypeptide AE_{1-157}. This polypeptide chain was purified in the manner described previously for the A_{1-120} domain (Darby & Creighton, 1995a).

Purification of Polypeptide $EB_{100-257}$. Ammonium sulfate was added to the cell lysate at 4 °C to 40% of saturation and the precipitated protein removed by centrifugation. To the supernatant, further ammonium sulfate was added to reach 65% of saturation. The precipitated protein was

recovered by centrifugation, resuspended in 20 mM Tris (pH 8), and dialyzed against 50 mM Tris (pH 8). The sample was then loaded onto a 1.6 × 15 cm Q-Sepharose column equilibrated in the same buffer and eluted with a gradient from 0 to 0.5 M NaCl. Ammonium sulfate was added to 65% of saturation to the fractions containing the $EB_{100-257}$ polypeptide, and the precipitated protein was recovered by centrifugation. It was dissolved in 4 mL of 50 mM sodium phosphate (pH 7.4) and loaded onto a 1.6 × 10 cm column of Octyl Sepharose equilibrated with the same buffer containing ammonium sulfate at 50% saturation. The protein was eluted by a gradient from this buffer to the same one without ammonium sulfate. The pooled fractions were dialyzed against 50 mM sodium phosphate (pH 7.4) and stored frozen at −80 °C until use.

Purification of Polypeptide $B_{148-257}$. Overexpression of $B_{148-257}$ was apparent by SDS−PAGE only when its N-terminus was fused to a 12-residue leader peptide from the T7 capsid major coat protein; the product was a soluble protein. Taurine was added to the cell lysate to a final concentration of 4% (w/v) to improve the subsequent chromatography of the protein, presumably by reducing the effects of aggregation. The solution was loaded onto a 1.6 × 20 cm column of Q-Sepharose equilibrated with 50 mM Tris (pH 7.4), containing 4% (w/v) taurine, and eluted with a gradient from 0 to 0.5 M NaCl. HCl was added to 0.1 M to the pooled fractions, and an approximate 100-fold excess of cyanogen bromide over protein was added. After overnight incubation at room temperature, the cleavage to release the PDI sequence was essentially complete as judged by SDS−PAGE. Cyanogen bromide was removed by dialysis against 50 mM Tris-acetate (pH 8) and the protein stored at −20 °C.

Purification of Polypeptide $B_{119-228}$. Cells were grown in minimal media supplemented with BME vitamin mix (Gibco-BRL) and using ammonium sulfate labeled with ^{15}N to the extent of 20% or 100% as the sole nitrogen source. Ammonium sulfate was added to 35% saturation to the cell lysate at 4 °C and the precipitate removed by centrifugation. Further ammonium sulfate was added to 80% saturation. The protein was recovered by centrifugation, resuspended in 15 mL of 50 mM sodium phosphate (pH 7.4), and loaded onto a 1.6 × 10 cm column of Phenyl Sepharose equilibrated with the same buffer containing ammonium sulfate at 80% saturation. The column was washed with a further 50 mL of the same solution, and the protein was eluted by a linear gradient to the absence of ammonium sulfate in the same buffer. The pooled fractions were dialyzed against 50 mM Tris (pH 7.4) and loaded onto a 1.6 × 20 cm column of Q-Sepharose equilibrated with the same buffer containing 0.1 M sodium chloride. The protein was eluted with a gradient from 0.1 to 0.3 M NaCl. The pooled fractions were dialyzed against 10 mM sodium phosphate (pH 6.5) and stored at −80 °C until use.

Purification of Polypeptide AEB_{1-257}. Ammonium sulfate was added to 45% saturation to the cell lysate at 4 °C. The precipitate was removed by centrifugation and further ammonium sulfate added to the supernatant to a saturation of 70%. The protein was recovered by centrifugation, resuspended in 15 mL of 50 mM sodium phosphate (pH 7.4), 2 mM DTT, and 1 mM EDTA, and loaded onto a 1.6 × 15 cm column of Phenyl Sepharose equilibrated with the same buffer containing 70% saturated ammonium sulfate. The protein was eluted by a gradient of decreasing ammonium sulfate concentration. The pooled fractions were dialyzed against 20 mM Tris (pH 8), then adjusted to 25 mM Tris (pH 7.5) and 2 mM DTT, and loaded onto a 1.6 × 15 cm column of Q-Sepharose equilibrated with the same buffer. The protein was eluted with a gradient from 0 to 0.4 M NaCl. Pooled fractions were dialyzed against 50 mM Tris (pH 7.4) and stored at −80 °C until use.

Limited Proteolysis of PDI Constructs. Limited proteolysis was carried out in 50 mM Tris (pH 8) at 37 °C using trypsin, V8 protease, or thermolysin at the protease concentrations described in the text. Proteolysis was stopped by addition to 5 mM of either PMSF or TLCK (for trypsin) or EDTA (for thermolysin), followed after a 2 min incubation by either acidification to pH 2 (for analysis by HPLC) or addition of electrophoresis sample buffer and boiling (for electrophoretic analysis). Proteolysis with V8 protease was terminated solely by either acidification or addition of sample buffer and boiling. Samples were analyzed by SDS−PAGE or by reverse-phase HPLC on a Vydac C-18 column using gradients of acetonitrile in 0.1% (v/v) TFA. Proteolytic fragments isolated by HPLC were lyophilized and resuspended in 0.1% (v/v) TFA for sequence analysis and mass spectrometry or 50 mM Tris (pH 8)/8 M urea for electrophoresis on urea gradient gels.

Concentration Measurements. The concentrations of PDI and A_{1-120} were determined using the previously reported molar absorbance coefficients of 56 399 and 19 060 cm^{-1} M^{-1}, respectively (Darby & Creighton, 1995a). PDI.C380S,-C383A was assumed to have the same molar absorbance coefficient as normal PDI. Molar absorbance coefficients for the other PDI polypeptides were calculated by the method of Gill and Von Hippel (1989); $E_{100-157}$, 5690 cm^{-1} M^{-1}; AE_{1-157}, 19 120 cm^{-1} M^{-1}; $EB_{100-257}$, 8270 cm^{-1} M^{-1}; $B_{148-257}$, 2560 cm^{-1} M^{-1}; $B_{119-228}$, 1280 cm^{-1} M^{-1}; AEB_{1-257}, 21 620 cm^{-1} M^{-1}.

Electrophoresis. SDS−PAGE was performed using 15% (w/v) polyacrylamide gels in the system of Laemmli (1970). Native PAGE was performed using the high pH gel system of Davis (1964). Urea gradient gel electrophoresis was carried out as described by Creighton and Shortle (1994).

Gel Filtration. Gel filtration analysis was carried out on Superose 12 (Pharmacia), TSK 2000SWXL, or TSK 4000SWXL columns in either 50 mM sodium phosphate (pH 7.4)/0.1 M KCl or 50 mM Tris (pH 7.4)/0.1 M NaCl.

Spectral Analysis. Circular dichroism spectra were measured with a Jasco J710 spectrometer in a 1 mm path length cell at 20 °C; spectra were averages of five scans. The protein was dissolved in 20 mM Tris (pH 7.4), also containing 0.2 mM DTT when cysteine residues were present in the protein. NMR measurements were performed on 2 mM samples of the $EB_{100-257}$ and $B_{119-228}$ polypeptides dissolved in 90% H_2O/10% D_2O that was buffered at pH 6.5 by 10 mM sodium phosphate. All spectra were recorded on a Brucker AMX600 spectrometer at 300 K.

Catalytic Assays. The effects of catalysts on the disulfide refolding of reduced BPTI and on disulfide bond and glutathione mixed disulfide formation in two model peptides were determined as described previously (Darby & Creighton, 1995a). In all cases the thiol−disulfide exchange reactions with 0.5 mM GSSG and 2 mM GSH were carried out at 25 °C in 0.1 M Tris (pH 7.4), 0.2 M KCl, and 1 mM EDTA.

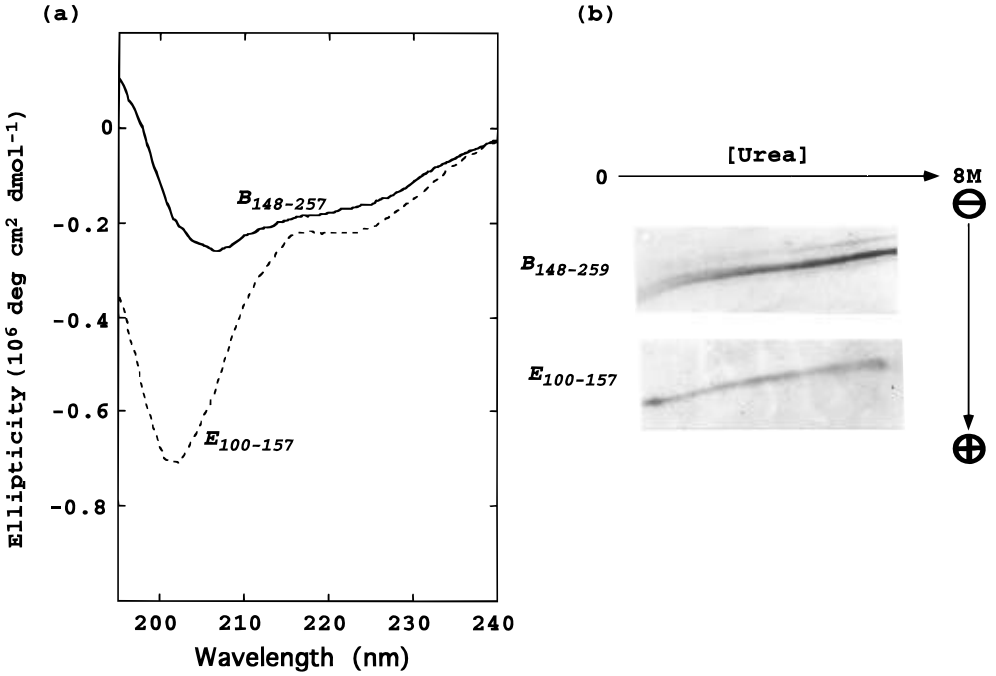


FIGURE 1: Far-UV CD spectra (a) and urea gradient gel electrophoresis (b) of the $B_{148-257}$ and $E_{100-157}$ polypeptide chains. The gradual decrease of electrophoretic mobility with increasing urea concentration observed in (b) was typical of other proteins that underwent no unfolding transition on similar gels.

RESULTS

Characterization of the Structural Domains of PDI with Fragments of Its N-Terminal Half. The polypeptide fragments of human PDI that were used in this study are listed in Table 1. The nomenclature adopted is based upon the hypothesized domain structure of PDI (eq 1), with the hypothetical domains present in capital letters and the residues of mature PDI in subscripts. The polypeptide chains produced were chosen to contain all of the residues of each putative domain, plus all flanking polar residues. This strategy helps to ensure that the polypeptide chain will be soluble and that it will have all the residues necessary for stability of the folded domain. Thus the *e*-domain was proposed to consist of residues 100−144 (Tsibris et al., 1989; Freedman et al., 1994) but was prepared here as residues 100−157, $E_{100-157}$. Similarly, polypeptide $B_{148-257}$ included all the residues proposed by Edman et al. (1985) and Freedman et al. (1994) to form the *b*-domain of PDI, 154−225, but with additional residues at each end.

The aim of this work was to prepare autonomously folded units of PDI suitable for further structural and functional analysis. The criteria were that the proteins (i) had obvious signs of folded structure by NMR and CD, (ii) undergo cooperative folding transitions by urea gradient gel electrophoresis, (iii) not form multiple aggregated forms, as judged by gel filtration, and (iv) be suitable for structure determination by multidimensional NMR analysis.

Two PDI fragments were made initially with sequences based upon the proposed *e*- and *b*-domains, $E_{100-157}$ and $B_{148-257}$, respectively. $E_{100-157}$ was expressed directly as a soluble protein and purified as a mixture of two with and without the N-terminal initiating methionine residue. The $B_{148-257}$ fragment was released by CNBr cleavage of a fusion protein.

Both the $B_{148-257}$ and $E_{100-157}$ polypeptide chains gave CD spectra without substantial indications of structure (Figure 1). There were some slight indications of structure in the fusion protein for $B_{148-257}$, but gel filtration analysis demonstrated that this protein was highly aggregated (data not shown). Urea gradient gel analysis indicated that neither $E_{100-157}$, $B_{148-257}$, nor the $B_{148-257}$ fusion protein underwent cooperative unfolding transitions, as there was no substantial change in electrophoretic mobility at any point along the urea gradient. A fully folded protein normally has about 3 times greater electrophoretic mobility than its unfolded form, and the transition between the two occurs abruptly in a cooperative unfolding transition.

The covalent structures of these polypeptide chains were verified by mass spectrometry, and the most likely explanation for their unfolded nature was that the domain boundaries chosen were incorrect. Because of the highly cooperative nature of protein stability, omitting even a few residues at the termini of a protein can result in it not adopting a fully native state [e.g., Taniuchi et al. (1970, 1973)]. One approach to explore this is to add residues gradually to the termini to determine at what point a folded structure is adopted. An alternative approach was used here, however, in which polypeptide chains comprising two or three of the expected *a*-, *e*-, and *b*-domains were prepared and then subjected to limited proteolysis, to remove the surplus residues.

All three of these proteins, AE_{1-157}, $EB_{100-257}$, and AEB_{1-257}, contained folded structures. All three were soluble, eluted as a single, included peak on gel filtration analysis, gave a CD spectrum typical of a folded protein, and exhibited a cooperative folding transition on urea gradient gel electrophoresis (Figure 2). Polypeptide AEB_{1-257} gave a single unfolding transition at about the same urea concentrations as observed with AE_{1-157} and $EB_{100-257}$. As AEB_{1-257} is expected to have both of the structural domains that are present individually in AE_{1-157} and $EB_{100-257}$, these observations indicate that the individual domains have very similar stabilities to unfolding by urea and that they unfold

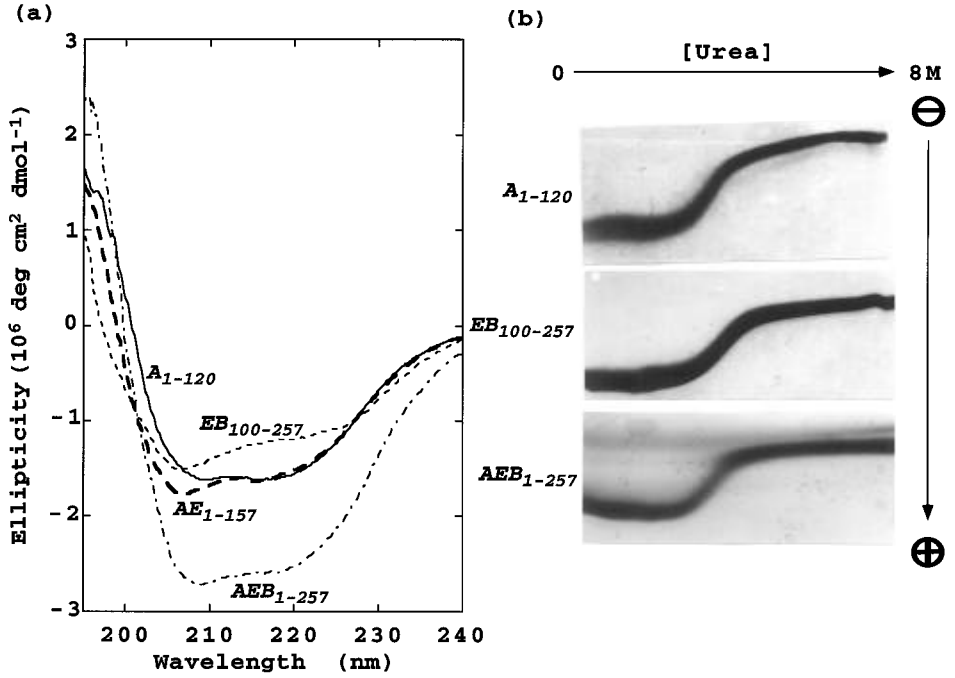


FIGURE 2: Far-UV CD spectra (a) and urea gradient gel electrophoresis (b) of the A_{1-120}, AE_{1-157}, $EB_{100-257}$, and AEB_{1-257} polypeptide chains. The slight perturbations in mobility at very low and very high urea concentrations were due to imperfections in the gels and are not indicative of further unfolding transitions.

independently in AEB_{1-257}. That $EB_{100-257}$ was folded was confirmed by NMR analysis, but it was apparent from the NMR spectra that there were also substantial amounts of unfolded polypeptide chain (see Figure 7a, below), which would hamper any structure determination by NMR spectroscopy. To define more accurately the limits of folded structure within these polypeptide chains, limited proteolysis studies were carried out to remove any flexible, nonstructured residues.

The conditions of proteolysis were selected initially using SDS−PAGE separations of proteins subjected to digestions with varying amounts of trypsin, V8 protease, or thermolysin for varying lengths of time (Figure 3). The most stable large fragments were isolated by HPLC and identified by mass spectrometry and, when required, by N-terminal sequence analysis (Table 2). Whether there was cooperatively folded structure in each fragment was assessed by urea gradient gel electrophoresis. When proteolysis removes only unstructured residues from the termini, the resulting fragment should exhibit an unfolding transition at the same urea concentration as the original protein, but with a larger relative change in electrophoretic mobility upon unfolding, because a greater proportion of the polypeptide chain is in the folded conformation. This was generally observed with the fragments isolated here (Figure 4).

The major points of proteolysis in AE_{1-157} and $EB_{100-257}$ that resulted in fragments that retained stable, cooperatively folded structures are shown in Figure 5. Comparison of the patterns of proteolysis in the two proteins suggests the location of the boundaries of *a*-domain to within about 15 residues. V8 protease and thermolysin cut the $EB_{100-257}$ polypeptide chain after Glu104 and Trp111, respectively, whereas these sites were not cleaved rapidly in AE_{1-157}. The difference in susceptibility of the sites in the two constructs presumably relates to the fact that these residues are within folded structure in AE_{1-157}, but not in $EB_{100-257}$. No cleavage at the Thr112−Leu113 peptide bond, a potential thermolysin cleavage site, was observed in either construct. This is presumably because the leucine is followed by a proline, which is known to reduce susceptibility to proteolysis by thermolysin (Matsubara et al., 1969). The pattern of

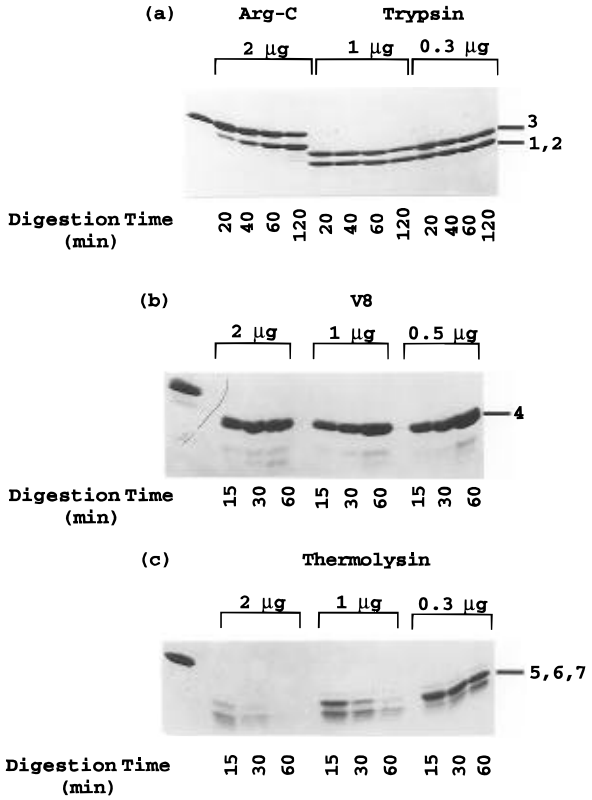


FIGURE 3: Proteolytic trimming of polypeptide $EB_{100-257}$ using (a) Arg-C endopeptidase and trypsin, (b) V8 protease, and (c) thermolysin monitored by SDS−PAGE. The initial polypeptide $EB_{100-257}$ is in the first lane of each gel. The amounts of protease added to 100 μL of 25 μM polypeptide $EB_{100-257}$ and the period of proteolysis are indicated. The numbered bands correspond to the fragments listed in Table 2; some bands contained more than one fragment.

Table 2: Proteolytic Trimming of the $EB_{100-257}$ Polypeptide[a]

fragment[b]	protease	primary structure[c]	molecular mass	
			measured	expected
(1)[b]	trypsin	(K)K[114]RTG to QTAPK[230](I)	12 826.5	12 827.3
(2)		(K)R[115]TGP to QTAPK[230](I)	12 698.1	12 699.1
(3)		(R)T[116]GPA to QTAPK[230](I)	12 542.3	12 541.5
(4)	V8	(E)A[104]DDI to LVIE[222](F)	13 493.8	13 493.0
(5)	thermolysin	(W)L[112]KKR to QTAPK[230](I)	13 067.9	13 068.7
(6)		(W)L[112]KKR to PLVIE[222](F)	12 165.1	12 165.7
(7)		(W)L[112]KKR to HNQLP[218](L)	11 710.5	11 711.1

[a] Proteolytic digestions of polypeptide $EB_{100-257}$ were carried out as described in Materials and Methods, and the resulting peptides were purified by reverse-phase HPLC. Substrate to protease ratios were 50:1 for trypsin (60 min digestion), 150:1 for thermolysin (15 min digestion), and 125:1 for V8 protease (60 min digestion). The fragments were identified from their molecular weights determined by mass spectrometry and knowledge of the enzyme specificity, except for the thermolysin fragments, for which the N-terminal sequence was also determined. The fragment numbering refers to the sequence of mature PDI. [b] The numbers refer to the proteolytic fragments separated by SDS−PAGE (Figure 4). [c] The residues in the primary structure flanking the fragment are shown in brackets.

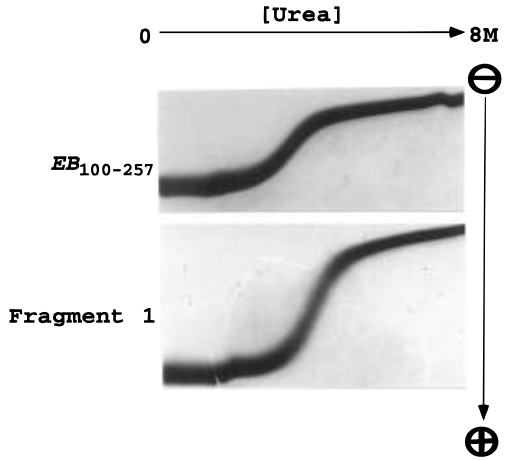


FIGURE 4: Urea gradient gel electrophoresis of the initial polypeptide $EB_{100-257}$ and of the fragment 1 generated by trypsin cleavage. Fragment 1 (Figure 3) contains only residues 114−230 (Table 2); it was purified by HPLC prior to electrophoresis.

proteolysis indicates the presence of a folded domain in the AE_{1-157} polypeptide chain consisting of no more than residues 1−130, consistent with the presence of the folded *a*-domain, which has been established by NMR to consist of residues 3−116 (Kemmink et al., 1996).

Cleavage occurred at Glu130 and Ser131 in AE_{1-157}, but not in $EB_{100-257}$, suggesting that these residues lie within structured regions only in the latter. The greatest extent of proteolysis at the N-terminus of $EB_{100-257}$ occurred after Arg115. At the C-terminus of $EB_{100-257}$, the shortest fragment generated was from cleavage after Pro218. Potential cleavage sites before this residue in the sequence were not utilized. These data indicate that the folded portion of the $EB_{100-257}$ polypeptide chain begins between residues 116−130 and terminates in the region of residue 218; this folded structure is designated as the *b*-domain.

Proteolysis of the AEB_{1-257} polypeptide chain by either trypsin or thermolysin resulted in only small decreases in the size of the protein. As neither of these proteases appeared to generate stable fragments from AE_{1-157} by cleavage at the N-terminus, it seems likely that proteolysis of AEB_{1-257} under these conditions occurred mainly at the C-terminus. Residues 116−130, which are likely to link the two domains, do not contain any sites that are readily cleaved by thermolysin or trypsin.

Preparation of the Isolated b-Domain of PDI. The above proteolysis data led to the design of a new polypeptide chain of residues Ala119 to Ala228, to contain only the *b*-domain, $B_{119-228}$. The residues immediately preceding Ala119 are disordered in the NMR structure of the A_{1-120} construct, are predicted to have a high probability of being in a turn conformation, and would be very close to the folded *a*-domain, so they were considered unlikely to be part of the next domain. The C-terminus of $B_{119-228}$ was chosen to allow the polypeptide chain to terminate in a number of relatively polar residues, for solubility reasons. Polypeptide $B_{119-228}$ was expressed as a soluble protein that was a single molecular species on gel filtration analysis. The N-terminal methionine residue was removed from a substantial proportion of the protein, but some of the unprocessed form was also present. The CD spectrum of this protein is compared to that of $EB_{100-257}$ in Figure 6. Subtraction of the spectrum of $B_{119-228}$ from that of $EB_{100-257}$ yielded a spectrum characteristic of an unfolded protein (cf. Figure 1). This is consistent with the limited proteolysis data indicating that there are unstructured regions at each end of polypeptide $EB_{100-257}$. The $B_{119-228}$ protein gave a cooperative folding transition on urea gradient gel electrophoresis (Figure 6b), with an unfolding transition at about the same urea concentration as observed with the larger fragment $EB_{100-257}$ (Figure 2b). The CD spectrum of AEB_{1-257} was very similar to the sum of the spectra of A_{1-120} and $B_{119-228}$ (Figure 6),

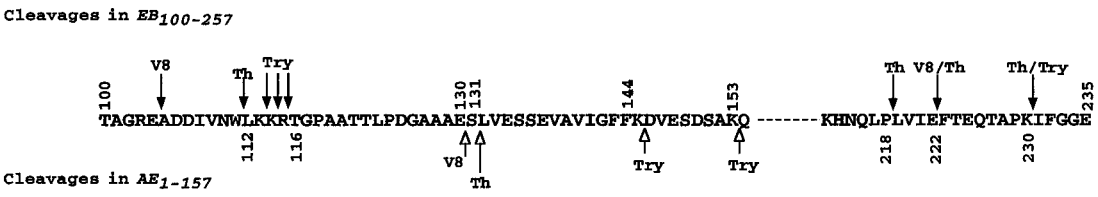


FIGURE 5: Sites of proteolytic cleavage in $EB_{100-257}$ and AE_{1-157}. Cleavage of $EB_{100-257}$ at the sites shown at the top resulted in fragments of the central portion of the primary structure that maintained a stable cooperatively folded structure, as determined by urea gradient gel electrophoresis. Major sites of proteolysis identified in the same area of the sequence of polypeptide AE_{1-157} are also shown for comparison. The residue numbers refer to the mature PDI sequence. The proteases are designated as follows: Th, thermolysin; Try, trypsin; V8, V8 protease.

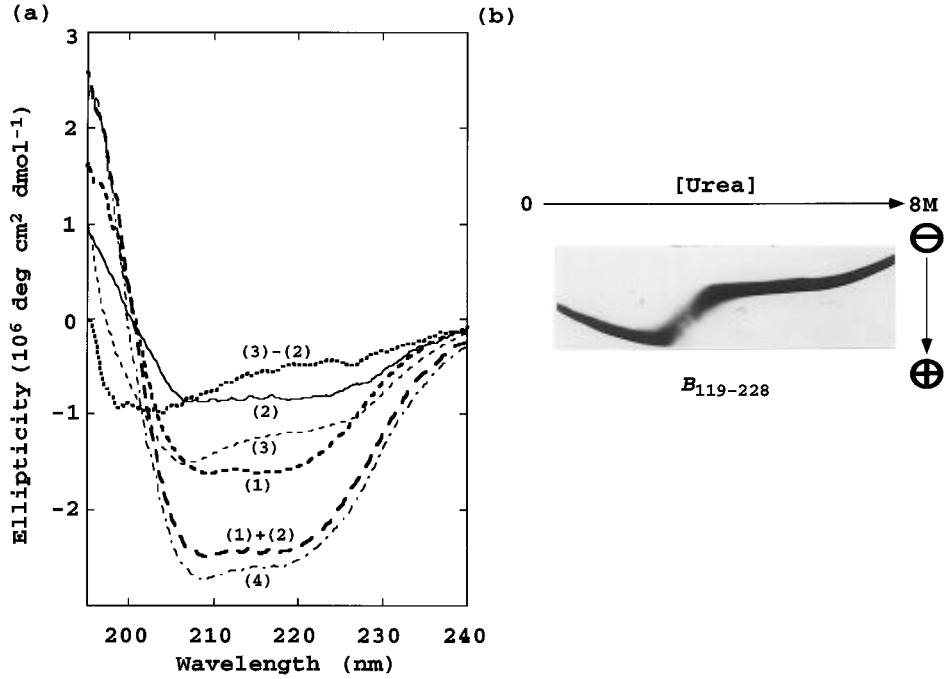


FIGURE 6: (a) Far-UV CD spectra of the (1) A_{1-120}, (2) $B_{119-228}$, (3) $EB_{100-257}$, and (4) AEB_{1-257} polypeptide chains. (b) Urea gradient gel electrophoresis of $B_{119-228}$.

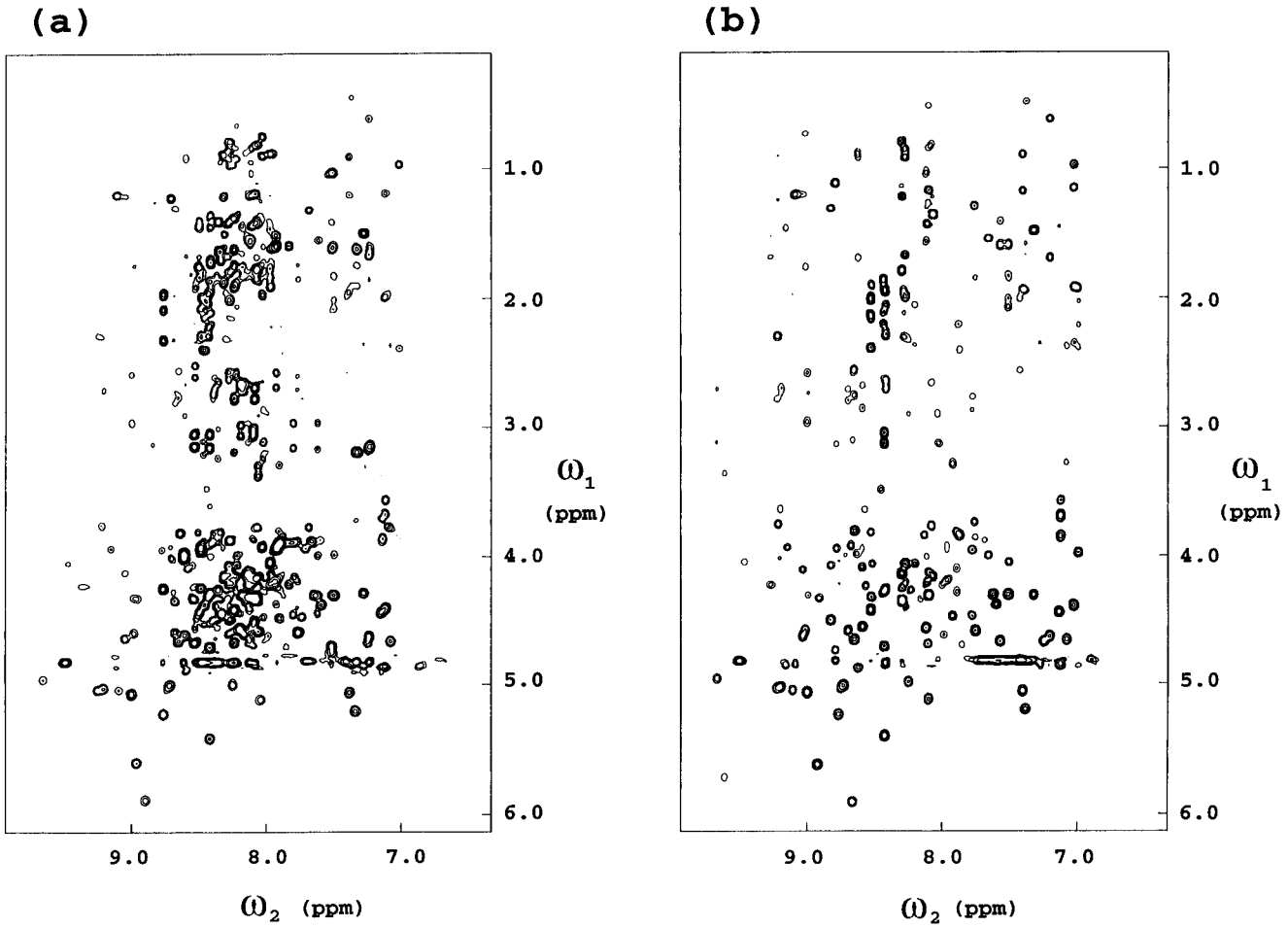


FIGURE 7: Portions of the 600 MHz ^{1}H-NMR TOCSY spectra (τ_m = 60 ms) of polypeptides (a) $EB_{100-257}$ and (b) $B_{119-228}$, showing intraresidue cross peaks between the NH protons and those of its $C^\alpha H$ and side chains. A spectral width of 8196.7 Hz was used in both dimensions. The final resolution of the processed spectrum was 8.0 Hz/pt in both ω_1 and ω_2.

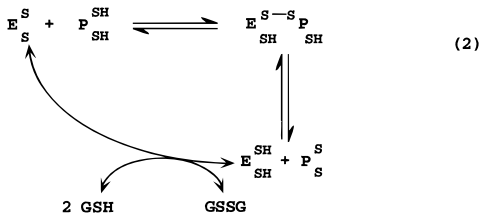
consistent with the same two folded structures, i.e., both the *a*- and the *b*-domains, being present in the two cases.

The presence of very similar folded structure in both the $EB_{100-257}$ and $B_{119-228}$ polypeptide chains was apparent from

the presence of dispersed resonances in their 1D NMR spectra (data not shown). Such spectra are, however, not sensitive to the presence of unfolded sections of polypeptide chain. Resonances arising from unfolded residues normally do not show large dispersion of their chemical shifts and tend to give sharp peaks as a result of increased flexibility. Such resonances are more readily discerned in TOCSY spectra (Figure 7). Such spectra of both $EB_{100-257}$ and $B_{119-228}$ contain well dispersed peaks characteristic of a folded protein, but that of $EB_{100-257}$ also has a considerable number of intense peaks clustered around 8 ppm in ω_2, characteristic of unfolded residues. Most of these peaks are absent from the spectrum of $B_{119-228}$, confirming that this polypeptide chain has fewer residues not part of the folded conformation.

Catalytic Activities. In view of previous findings that the A_{1-120} domain alone catalyzes only a subset of the activities catalyzed by PDI, the catalytic properties of the AEB_{1-257} polypeptide chain were examined to determine if addition of the *b*-domain influenced in any way the catalytic activity of the *a*-domain. To make this comparison meaningful relative to PDI, these catalytic activities were compared to that of a form of PDI in which the *a*′-domain was inactivated by replacing its active site -Cys-Gly-His-Cys- sequence with -*Ser*-Gly-His-*Ala*-. The *b*-domain contains no cysteine residues and would be expected to be inactive in thiol/disulfide exchange reactions.

Catalysis of the formation of disulfide bonds was examined in two unfolded 28-residue model peptides, P_{2SH}^{27SH} and P_{2OH}^{27SH}, which are identical apart from having respectively two and one cysteine residues. They are useful models of unfolded proteins with which to examine the process of disulfide bond formation, as all the possible thiol and disulfide species can be resolved by HPLC in each case (Darby et al., 1994; Darby & Creighton, 1995c). Catalytic amounts of PDI have been shown to catalyze each step in the reaction between the peptide and GSH and GSSG, except for the interchange of the glutathione moiety between the two cysteine residues (Darby et al., 1994). The likely mechanism of catalysis involves the direct transfer of an active site disulfide bond from the catalyst (designated here as E_S^S) via an intermediate enzyme−peptide mixed disulfide. The resulting E_{SH}^{SH} form of the catalyst is reoxidized by GSSG, and this step limits the overall reaction velocity (Darby & Creighton, 1995c; unpublished data).



(2)

The apparent catalysis of formation of the mixed disulfide P_{SH}^{SSG} very likely results from the reaction of the intermediate enzyme−peptide mixed disulfide with GSH; when P_{2OH}^{27SH} is the substrate, this is the only product possible.

The A_{1-120} domain of PDI has about half the catalytic activity of intact PDI in catalyzing disulfide formation in P_{SH}^{SH} and in P_{OH}^{SH}; the *a*′-domain accounts for the other half (Darby & Creighton, 1995a). This was confirmed here with the observation that PDI.C380S,C383A had an activity very similar to that of A_{1-120} alone and about half that of intact PDI (Table 3). Polypeptide AEB_{1-257} had about the same activity as PDI.C380S,C383A and as A_{1-120} alone. These observations indicate that the individual *a*- and *a*′-domains contribute individually and independently to the catalysis of this step by each protein. Further, the presence of the *b*-domain has little effect on this catalytic activity of the *a*-domain, either in isolation in AEB_{1-257} or together in PDI.C380S,C383A.

PDI also catalyzes the second step in forming the disulfide bond of P_S^S, the intramolecular step in both of the glutathione mixed disulfides, and each mixed disulfide was used at similar rates (Darby et al., 1994; Darby & Creighton, 1995a). The catalytic activity of PDI.C380S,C383A was quite similar (Table 4). In contrast, A_{1-120} catalyzed this step much less and utilized P_{2SSG}^{27SH} in preference to P_{2SH}^{27SSG}, as previously reported (Darby & Creighton, 1995a). AEB_{1-257} had catalytic activity that was similar in magnitude to that of A_{1-120}, but the difference in utilization of the mixed disulfides was not so marked. The catalytic activities of the individual PDI domains were clearly distinguished from those of intact PDI, and some other aspect of the PDI structure is involved in the function of making the intramolecular disulfide bond. That other aspect is not solely the *b*-domain, which had little effect.

The disulfide folding pathway of BPTI can be used to provide a comprehensive overview of the properties of catalysts of thiol−disulfide interchange, as it is well understood and includes events that involve disulfide bond

Table 3: Effects of PDI and Its Isolated Domains on the Rate of Disulfide Formation in the Model Peptides P_{2SH}^{27SH} and P_{2OH}^{27SH} [a]

	rate of disappearance (10^{-2} s^{-1})	
catalyst	P_{2SH}^{27SH}	P_{2OH}^{27SH}
A_{1-120}	1.9 ± 0.3	0.8 ± 0.06
AEB_{1-257}	1.6 ± 0.3	0.8 ± 0.06
PDI.C380S,C383A	1.8 ± 0.3	1.1 ± 0.20
PDI	2.9 ± 0.2	1.5 ± 0.20
none	0.14 ± 0.02	0.12 ± 0.02

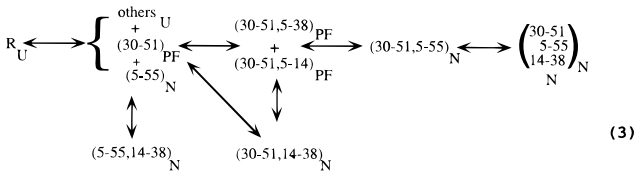
[a] The rates of disappearance of P_{2SH}^{27SH} and P_{2OH}^{27SH} were measured in the presence of 0.5 mM GSSG, 2 mM GSH, and 1 μM of each of the catalysts. In the case of P_{2SH}^{27SH}, the products of the reaction were P_S^S plus the two single mixed-disulfide species, P_{2SH}^{27SSG} and P_{2SSG}^{27SH}, and the double mixed disulfide species. In the case of P_{2OH}^{27SH}, the only product was P_{2OH}^{27SSG}.

Table 4: Effects of PDI and Its Domains on the Rate of Forming the Disulfide Bond in the Mixed Disulfides between Glutathione and Each of the Peptide Cysteine Residues[a]

	rate produced by each catalyst (s^{-1}/μM catalyst)	
catalyst	P_{2SH}^{27SSG}	P_{2SSG}^{27SH}
A_{1-120}	0.034	0.09
AEB_{1-257}	0.042	0.058
PDI.C380S,C383A	0.43	0.21
PDI	0.46	0.32

[a] The initial rate of the reaction $P_{SH}^{SSG} \rightarrow P_S^S$ was measured with each of the mixed disulfide forms of the peptide and varying concentrations of the catalysts, which had been reduced prior to their use. The linear dependence of the observed rate upon the catalyst provided the values given here. The observed rates in the absence of catalysts for P_{2SH}^{27SSG} and P_{2SSG}^{27SH} were 0.007 and 0.009 s^{-1}, respectively.

formation, breakage, and rearrangement (Creighton et al., 1996).



(3)

R is the fully reduced and unfolded protein with six free thiol groups. The disulfide species are depicted by the residue numbers of the cysteine residues paired in disulfide bonds. Those that adopt unfolded conformations are designated with the subscript U, partly-folded conformations by PF, and fully folded conformations by N. Forming the first disulfide bond in R is essentially random, but it then undergoes rapid disulfide interchange. At neutral pH as used in these studies, two one-disulfide species predominate: $(30-51)_{PF}$ and $(5-55)_N$. As a result, subsequent disulfide formation is largely restricted to these species. The quasi-native species $(5-55,14-38)_N$ and $(30-51,14-38)_N$, with two native disulfide bonds, predominate at neutral pH because they are relatively stable and because they are blocked in further disulfide formation. The most productive folding process occurs via intramolecular disulfide rearrangements of the two-disulfide intermediates with nonnative second disulfide bonds, (30−51,5−14) and (30−51,5−38), to $(30-51,5-55)_N$ (Darby et al., 1995), which can rapidly form the final disulfide bond. The effects of catalysts can be determined using either the entire pathway or the individual isolated intermediates, in the presence or absence of reagents for making and breaking protein disulfide bonds.

The effects of the various catalysts on the folding pathway of BPTI were examined in the presence of 0.5 mM GSSG and 2 mM GSH, which approximates the redox conditions of the endoplasmic reticulum where PDI usually functions (Hwang et al., 1992). The formation of the native three-disulfide form of the protein, N-BPTI, is very slow at neutral pH, and the kinetically trapped $(5-55,14-38)_N$ and $(30-51,14-38)_N$ species predominate. The effects of catalysts can be judged simply by following the rates of appearance and disappearance of these three species.

PDI increases the rates of all the steps in the BPTI pathway, as well as making direct formation of the 5−55 disulfide bond significant in $(30-51)_{PF}$ and $(30-51,14-38)_N$ (Creighton et al., 1980, 1993). In the experiments here in the presence of PDI (Figure 8), there was an increase in the general rate of disulfide bond formation, as shown by the increase in the rate of formation of $(5-55,14-38)_N$ and $(30-51,14-38)_N$. These species also disappeared more quickly, due to the combined effects of increased rates of disulfide bond rearrangements and direct insertion of the third native disulfide bond. PDI.C380S,C383A was only slightly less active than PDI in each aspect, consistent with other observations that the *a*- and *a*′-domains function relatively independently in catalyzing disulfide formation (Vuori et al., 1992; Lyles & Gilbert, 1994; Darby & Creighton, 1995a).

The A_{1-120} polypeptide chain increased the rate of initial disulfide formation, as demonstrated by the increased rate of appearance of the two-disulfide species (Figure 8). Their slow rates of disappearance, however, suggest that A_{1-120} was relatively inactive in catalyzing disulfide rearrangements.

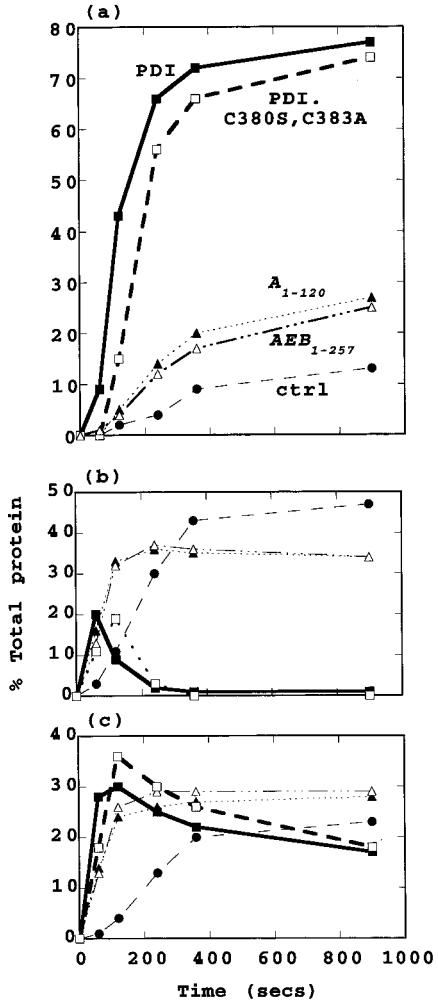


FIGURE 8: Effects of each catalyst on the time course of refolding of reduced BPTI with 0.5 mM GSSG and 2 mM GSH. HPLC separations of the acid trapped species were used to measure the relative amounts of each of the species as a function of time. The time course is presented of the appearance and disappearance of (a) fully refolded three-disulfide form of BPTI, (b) $(30-51,14-38)_N$, irrespective of whether there is a mixed disulfide between the free thiol groups and glutathione, and (c) $(5-55,14-38)_N$. The catalyst was (●) none; (■) 4 μM PDI; (□) PDI.C380S,C383A; (▲) 4 μM A_{1-120}; (△) 4 μM AEB_{1-257}.

The minor increase in the rates of their disappearance, plus the increased rate of formation of N-BPTI (Figure 8a), may be due to the direct insertion of the 5−55 disulfide bond into $(30-51)_{PF}$ and $(30-51,14-38)_N$. Polypeptide AEB_{1-257} had catalytic properties very similar to those of the A_{1-120}.

These conclusions were confirmed by direct examination of the catalysis of disulfide bond formation and rearrangement in $(30-51,14-38)_N$ by the various PDI derivatives. At pH 7.4, the half-time for spontaneous rearrangement of this intermediate to $(5-55,14-38)_N$ and $(30-51,5-55)_N$ is about 100 min; the latter species tends to be air oxidized to N-BPTI. At a substrate concentration of 10 μM, this process is accelerated about 200-fold by 2 μM PDI and 120-fold by 2 μM PDI.C380S,C383A. In contrast, 2 μM A_{1-120} or AEB_{1-257}, only accelerated the process by factors of 2- and 2.5-fold, respectively; such a small rate enhancement could be due to air oxidation of the catalyst and direct transfer of this disulfide bond.

Although $(30-51,14-38)_N$ normally forms the 5−55 disulfide bond extremely slowly in the presence of disulfide compounds such as GSSG, its formation by direct transfer

of a PDI or PDI.C380S,383A active site disulfide bond occurs rapidly (Darby & Creighton, 1995a). Accurate estimates of the rate of this process is complicated by the competing rearrangement of the intermediate, which becomes more significant as the quantity of reduced catalyst increases (Darby & Creighton, 1995a). The very low rate of catalysis of rearrangement by A_{1-120} and AEB_{1-257}, however, allowed the direct rate of the reaction between the BPTI intermediate and the disulfide form of A_{1-120} and AEB_{1-257} to be measured, giving second order rate constants of 1088 and 1406 s^{-1} M^{-1}, respectively. These rates are about 5-fold lower than those estimated for direct oxidation by PDI (Darby & Creighton, 1995a).

DISCUSSION

The multidomain nature of PDI was first proposed on the basis of its primary structure (Edman et al., 1985). Some aspects of the proposed domain model have been confirmed with the preparation of two of the individual domains of PDI, *a* and *a*′, which are homologous to thioredoxin (Darby & Creighton, 1995a,b). The *a*- and *a*′-domains could be engineered in a straightforward manner using the homology to thioredoxin to define the extents of each domain, although each domain was longer than proposed by Edman et al. (1985) and Freedman et al. (1994). To identify further domains was much more difficult, as the homology to known proteins was very uncertain. Initial attempts at designing polypeptide chains based on the existing model of eq 1 of the PDI domain structure ($E_{100-157}$ and $B_{148-257}$) did not produce any useful, folded proteins. An alternative strategy was adopted in which longer polypeptide chains containing two or three of the putative domains were prepared and then trimmed by proteases to isolate primarily the cooperatively folded structure. This could be conveniently assessed by using urea gradient gel electrophoresis as a sensitive method to examine the conformational properties of the proteolytic fragments. This method also has the advantage that the protein need not be purified to homogeneity before its state of folding can be examined. Removing only extraneous residues does not change the stability of the folded structure and increases the relative change in hydrodynamic volume upon unfolding at high urea concentrations (Figure 4).

Earlier models of the domain structure of PDI (eq 1) proposed that the thioredoxin-like *a*- and *a*′-domains are separated in sequence by three other domains, *e*, *b*, and *b*′. The identification of the *e*-domain as residues 101−144 by homology to the ligand binding domain of the estrogen receptor was attractive, as it provided an explanation for the, albeit modest, inhibition of PDI activity by estrogens (Tsibris et al., 1989). On the other hand, many other proteins of the data base show greater sequence similarities to PDI (Kemmink et al. 1995), and part of the *b*-domain was also found to have comparable sequence similarity to the estrogen receptor, but the two segments of sequence have reversed orders in the two proteins. Recently, determination of the structure of A_{1-120}, which contains the first 20 residues of the *e*-domain, cast further doubt that these residues could form part of an individual domain (Kemmink et al., 1995, 1996). Most of them have the same structural role in the *a*-domain as the corresponding residues in thioredoxin. As this *a*-domain appeared to be complete structurally, it was unlikely that 20 of its residues could also be part of another domain.

The present study goes further in casting doubt upon the existence of an autonomous *e*-domain. Expression of the proposed *e*-domain plus flanking polar residues, $E_{100-157}$, did not yield a folded protein. It could be argued that it is so small, with only 58 residues, that stabilizing interactions with other domains are required for it to attain a folded structure. When it was expressed joined to either the *a*-domain or the putative *b*-domain, however, the residues of the putative *e*-domain did not adopt folded conformations and were susceptible to extensive proteolysis. The proteolysis data were consistent with a model in which the first 16 residues of the proposed *e*-domain (Thr100 to Thr116) are part of the *a*-domain, but that the next structural domain starts somewhere between Thr116 and Glu130. This inference is confirmed by the finding that $B_{148-257}$, which lacked those residues prior to residue 148, was unfolded. Addition of residues 119−147, in $B_{119-228}$, resulted in a folded protein. Therefore, the residues of the first half of the putative *e*-domain are part of the *a*-domain, whereas at least a majority of the residues of the second half are part of the *b*-domain.

The structural evidence suggests, therefore, that an independent domain homologous to the ligand binding site of the estrogen receptor is unlikely to be present in the first half of the PDI polypeptide chain. It is not possible, however, to explain the report that estrogens inhibited PDI activity, with selectivity for estrogen over other steroid hormones (Tsibris et al., 1989). It is, of course, possible that estrogen binds in a less specific manner than suggested by Tsibris et al. (1989) and that the apparent difference between it and the other steroid hormones tested relates to some physical property of the molecule itself, such as an aromatic A-ring. Further studies are needed to resolve this issue.

The present results indicate that the N-terminal half of human PDI comprises only two folded domains, designated *a* and *b*, as originally proposed by Edman et al. (1985), but substantially greater in length. The corresponding AEB_{1-257} polypeptide chain was relatively resistant to proteolysis by V8 protease, trypsin, or thermolysin, with the main sites of cleavage appearing to be in the unfolded section of polypeptide chain at the C-terminus, after the *b*-domain, which was also cleaved by these enzymes in the $EB_{100-257}$ polypeptide. The *a*- and *b*-domains therefore account for virtually all of the first 222 residues of PDI, which is nearly half of the entire polypeptide chain. The C-terminal half of the polypeptide chain consists primarily of segments *b*′ and *a*′, plus a very acidic C-terminal segment (Edman et al., 1985). Therefore, the structural organization of this half of the polypeptide chain is likely to be similar to that of the first half, although the role of the C-terminal segment of acidic residues remains unknown. The b′-domain is likely to start near Glu222, the end of the *b*-domain, and to end near Asp348, which is approximately where the *a*′-domain begins.

The unfolding transitions observed here for the A_{1-120}, $EB_{100-257}$, $B_{119-228}$, and AEB_{1-257} constructs each occurred at similar urea concentrations. There was no indication with AEB_{1-257}, which contains both the *a*- and the *b*-folded domains, of the kind of complex unfolding transitions that are often seen with multidomain proteins, where each domain may be observed to undergo an individual unfolding transition. It appears that the folded structures of the individual *a*- and *b*-domains have very similar stabilities and that there are no substantial interactions between them in PDI or in

AEB_{1-257}. Intact PDI is observed to unfold by urea gradient gel electrophoresis in two steps at urea concentrations similar to those at which the *a*- and *b*-domains unfolded in the proteins studied here (unpublished observations). A previous report has indicated that unfolding of intact PDI by urea approximated to a single cooperative transition with a midpoint at 4.8 M, while more complex behavior was observed with guanidinium chloride (Morjana et al., 1993).

Although the boundaries of the *b*-domain of PDI have now been defined more clearly, the possible functions of it and the *b*′-domain remain unknown. They might be involved in assembly of the complexes of prolyl-4-hydroxylase and of the microsomal triglyceride transferase or in the binding of substrate proteins, but the peptide binding site of PDI has been reported to be close to the C-terminus of the protein, possibly within the *c*-domain (Noiva et al., 1993).

In view of previous findings that the *a*- and *a*′-domains catalyze only a subset of the reactions catalyzed by PDI, those involving insertion of disulfide bonds into unfolded proteins or peptides, it was of interest to examine if the addition of the *b*-domain to the *a*-domain sequence resulted in a more PDI-like spectrum of catalytic activities. The results obtained with the half molecule AEB_{1-257} were compared to a mutant PDI, PDI.C380S,C383A, in which the active site sequence of the *a*′-domain had been mutated, so that only a single active site, that of the *a*-domain, is present. This allows the properties of the active sites in the various constructs to be compared more directly. As expected from previous studies on PDI mutants (Vuori et al., 1992; Lyles & Gilbert, 1994), removal of the *a*′ active site resulted only in quantitative changes in the activity of the protein. PDI.C380S,C383A behaved very similarly to both the A_{1-120} and AEB_{1-257} constructs in catalyzing disulfide bond formation in model peptides, as expected from previous studies that suggested these activities of PDI could be quantitatively accounted for by the *a*- and *a*′-domains alone, and with no synergy between them (Darby & Creighton, 1995a). PDI.C380S,C383A also resembled PDI in that it catalyzed disulfide rearrangements in the quasinative states of BPTI, $(30-51,14-38)_N$ and $(5-55,14-38)_N$. In contrast, the AEB_{1-257} half molecule was virtually inactive and like the *a*-domain alone, A_{1-120}, in this respect. Clearly, further components of the PDI sequence beyond the *a*- and *b*-domains are required to be present for the full range of activities of PDI to be manifest.

It has been suggested that efficient catalysis of protein disulfide rearrangements might require the presence of two thioredoxin-like active sites on the catalyst (Zapun et al., 1995), but the efficient catalysis of rearrangements by PDI.C380S,C383A, with only a single such active site, appears to rule out this proposal. PDI does have a tendency to dimerize, but with a dissociation constant of about 10^{-4} M (N.J.D., S. J. Harding, and T.E.C., unpublished observations), so it will be primarily monomeric at the much lower concentrations used here. Some other aspect of the PDI structure that is present in PDI.C380S,C383A, but not in AEB_{1-257}, appears to be required for the full range of activities of PDI to be manifest.

With the boundaries of the *b*-domain now more clearly defined, it is worth reconsidering if this domain shows homology to any other proteins in the data base. Sonnhamer and Kahn (1994) developed an automated procedure to identify autonomous domains in proteins on the basis of sequence homology, which was applied to all sequences in the SwissProt amino acid sequence data base and compiled into a protein domain data base, ProDom. As expected, this data base catalogs the known homology between PDI, thioredoxin, and related proteins such as CaBP1 (P5 protein), CaBP2 (ERp72), and ERp60 (Freedman et al., 1994). Of more interest, however, is the presence in ProDom family 527 of homology between part of the *b*-domain and members of the calsequestrin family of calcium binding proteins of muscle sarcoplasmic reticulum. This homology is relatively limited, consisting of only about 40 residues, but this is due in part to the very stringent criteria that were used in compiling this data base, with the chance occurrence of homology less than 10^{-6}. Outside these 40 residues, further sequence similarities can be detected, possibly encompassing the *a*- and putative *b*′-domains. Unfortunately, no structures of calsequestrin or the *b*-domain are available, so it is not yet possible to look for the expected structural homology. It is intriguing from a biological point of view, however, as calsequestrin and PDI and the PDI-like proteins CaBP1 and CaBP2 are low-affinity calcium stores of the endoplasmic reticulum (Lytton & Nigam, 1992; Nguyen Van et al., 1993; Füllekrüg et al., 1994). Another hint of a relationship between calsequestrin and PDI is the finding that a 58 kDa protein first identified as a calcium storage protein from sea urchin eggs (Oberdorf et al., 1988) is homologous to PDI and has PDI-like activity (Lucero et al., 1994). This protein also exhibited immunological cross-reactivity with the cardiac muscle form of calsequestrin, although not the skeletal muscle form (Oberdorf et al., 1988). Some unusual chromatographic properties of calsequestrin are also shared by the sea urchin egg PDI, although calcium had some different effects on the conformational properties of the two proteins (Lebeche & Kaminer, 1992).

Given the previous misidentification of PDI as being involved in a range of cellular functions (Freedman, 1993) and the possible pitfalls that can occur in assigning protein function and relationships on the basis of limited sequence similarity alone, the suggested relationship between PDI and calsequestrin must be treated with caution (Freedman, 1992). This issue cannot be clearly resolved until the structures of both proteins are known. Fortunately, calsequestrin has now been crystallized (Hayakawa et al., 1994), and the structure of $B_{119-257}$ is being determined by NMR. In any case, the latter structure will lead to further progress in understanding the structure of PDI and the roles of its non-thioredoxin-like domains.

ACKNOWLEDGMENT

We thank K. Kivirikko for generously supplying the cloned gene of human PDI, Elke Penka and Renaud Vincentelli for excellent technical assistance, R. Jacob for peptide synthesis, Bayer AG for the gift of BPTI, and Robert Freedman, Toby Gibson, and Chris Sander for helpful discussion. Mass measurements and sequence analysis were carried out by members of the EMBL Peptide and Protein Group.

REFERENCES

Creighton, T. E., & Shortle, D. (1994) *J. Mol. Biol. 242*, 670−682.

Creighton, T. E., Hillson, D., & Freedman, R. B. (1980) *J. Mol. Biol. 142*, 43−62.
Creighton, T. E., Bagley, C. J., Cooper, L., Darby, N. J., Freedman, R. B., Kemmink, J., & Sheikh, A. (1993) *J. Mol. Biol. 232*, 1176−1196.
Creighton, T. E., Darby, N. J., & Kemmink, J. (1996) *FASEB. J. 10*, 100−118.
Darby, N. J., & Creighton, T. E. (1995a) *Biochemistry 34*, 11725−11735.
Darby, N. J., & Creighton, T. E. (1995b) *Biochemistry 34*, 16770−16780.
Darby, N. J., & Creighton, T. E. (1995c) *Biochemistry 34*, 3576−3587.
Darby, N. J., Freedman, R. B., & Creighton, T. E. (1994) *Biochemistry 33*, 7937−7947.
Darby, N. J., Morin, P. E., Talbo, G., & Creighton, T. E. (1995) *J. Mol. Biol. 249*, 463−477.
Davis, B. J. (1964) *Ann. N.Y. Acad. Sci. 121*, 404−427.
Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A., & Rutter, W. J. (1985) *Nature 317*, 267−270.
Freedman, R. B. (1992) in *Protein Folding* (Creighton, T. E., Ed.) pp 455−539, W. H. Freeman, New York.
Freedman, R. B. (1995) *Curr. Opin. Struct. Biol. 5*, 85−91.
Freedman, R. B., Hirst, T. R., & Tuite, M. F. (1994) *Trends Biochem. Sci. 19*, 331−336.
Füllekrüg, I., Sönnichsen, B., Wünsch, U., Arseven, K., Nguyen Van, P., Söling, H.-D., & Mieskes, G. (1994) *J. Cell. Sci. 107*, 2719−2727.
Gill, S. C., & von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
Givol, D., Goldenberger, R. F., & Anfinsen, C. B. (1964) *J. Biol. Chem. 239*, 3114−3116.
Haugerjorden, S. M., Srinivasan, M., & Green , M. (1991) *J. Biol. Chem. 266*, 6015−6018.
Hawkins, H. C., & Freedman, R. B. (1991) *Biochem. J. 275*, 335−339.
Hayakawa, K., Swenson, L., Baksh, S., Wei, Y., Michalak, M., & Derewenda, Z. S. (1994) *J. Mol. Biol. 235*, 357−360.
Hwang, C., Sinskey, A. J., & Lodish, H. F. (1992) *Science 257*, 1496−1502.
Kemmink, J., Darby, N. J., Dijkstra, K., Scheek, R. M., & Creighton, T. E. (1995) *Protein Sci. 4*, 2587−2593.
Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M., & Creighton, T. E. (1996) *Biochemistry 35*, 7684−7691.
Laemmli, U. K. (1970) *Nature 27*, 680−685.
LaMantia, M., & Lennarz, W. J. (1993) *Cell 74*, 899−908.
Lebeche, D., & Kaminer, B. (1992) *Biochem. J. 287*, 741−747
Lucero, H. A., Lebecke, E., & Kaminer, B. (1994) *J. Biol. Chem. 269*, 23112−23119.
Lyles, M. M., & Gilbert, H. F. (1994) *J. Biol. Chem. 269*, 30946−30952.
Lytton, J., & Nigam, S. J. (1992) *Curr. Opin. Cell. Biol. 4*, 220−226.
Martin, J. L. (1995) *Structure 3*, 245−250.
Matsubara, H., Singer, A., & Sasaki, R. M. (1969) *Biochem. Biophys. Res. Commun. 34*, 719−724.
Morjana, N. A., Mc Keone, B. J., & Gilbert, H. F. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2107−2111.
Nguyen Van, P., Rupp, K., Lampen, A., & Söling, H.-D. (1993) *Eur. J. Biochem. 213*, 785−789.
Noiva, R., Freedman, R. B., & Lennarz, W. J. (1993) *J. Biol. Chem. 268*, 19210−19217.
Oberdorf, J. A., Lebeche, D., Head, J. F., & Kaminer, B. (1988) *J. Biol. Chem. 263*, 6806−6809.
Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllylä, R., Huhtala, M.-L., Koivu, J., & Kivirriko, K. I. (1987) *EMBO. J. 6*, 643−649.
Sarkar, G., & Sommer, S. S. (1990) *BioTechniques 8*, 289−297.
Sonnhamer, E. L. L., & Kahn, D. (1994) *Protein Sci. 3*, 482−492.
Studier, F. W., Rosenberg, A. H., & Dunn, J. W. (1990) *Methods Enzymol 185*, 60−89.
Taniuchi, H. (1970) *J. Biol. Chem. 245*, 5459−5468.
Taniuchi, H. (1973) *J. Biol. Chem. 248*, 5164−5174.
Tsibris, J. C. M., Hunt, L. T., Ballejo, G., Barker, W. C., Toney, L. J., & Spellacy, W. N. (1989) *J. Biol. Chem. 264*, 13967−13970.
Vuori, K., Myllylä, R., Pihlajaniemi, T., & Kivirikko, K. I. (1992) *J. Biol. Chem. 267*, 7211−7214.
Wetterau, J. R., Combs, K. A., Spinner, S. N., & Joiner, B. J. (1990) *J. Biol. Chem. 265*, 9800−9807.
Wetterau, J. R., Combs, K. A., Spinner, S. N., & McLean, L. R. (1991) *Biochemistry 30*, 9728−9735.
Zapun, A., Missiakas, D., Raina, S., & Creighton, T. E. (1995) *Biochemistry 34*, 5075−5089.

BI960763S