Journal of Biomolecular NMR, **13:** 357–368, 1999.
KLUWER/ESCOM

The structure in solution of the *b* domain of protein disulfide isomerase*

Johan Kemmink[a,b,**], Klaas Dijkstra[c], Matteo Mariani[c], Ruud M. Scheek[c], Elke Penka[a], Michael Nilges[a] & Nigel J. Darby[a,d]

[a]*European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, D-69012 Heidelberg, Germany;* [b]*Biocenter and Department of Biochemistry, University of Oulu, FIN-90571 Oulu, Finland;* [c]*The Groningen Biomolecular Science and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands;* [d]*Present address: Astra Hässle AB, S-43183 Mölndal, Sweden*

Received 14 October 1998; Accepted 17 December 1998

Key words: calsequestrin, disulfide bond, protein disulfide isomerase, protein folding, thioredoxin

Abstract

Protein disulfide isomerase (PDI) is a multifunctional protein of the endoplasmic reticulum, which catalyzes the formation, breakage and rearrangement of disulfide bonds during protein folding. It consists of four domains designated ***a***, ***b***, ***b***′ and ***a***′. Both ***a*** and ***a***′ domains contain an active site with the sequence motif -Cys-Gly-His-Cys- involved directly in thiol-disulfide exchange reactions. As expected these domains have structures very similar to the ubiquitous redox protein thioredoxin. A low-resolution NMR structure of the ***b*** domain revealed that this domain adopts a fold similar to the PDI ***a*** domain and thioredoxin [Kemmink, J., Darby, N.J., Dijkstra, K., Nilges, M. and Creighton, T.E. (1997) *Curr. Biol.*, **7**, 239–245]. A refined ensemble of solution structures based on the input of 1865 structural restraints shows that the structure of PDI ***b*** is well defined throughout the complete protein except for about 10 residues at the C-terminus of the sequence. ^{15}N relaxation data show that these residues are disordered and not part of this structural domain. Therefore the domain boundaries of PDI can now be fixed with reasonable precision. Structural comparison of the PDI ***b*** domain with thioredoxin and PDI ***a*** reveals several features important for thiol-disulfide exchange activity.

Abbreviations: PDI, protein disulfide isomerase; (h)TRX, (human)thioredoxin; PFG, pulsed-field gradient; 2D, 3D, two-, three-dimensional; NOE, nuclear Overhauser effect; ppm, parts per million; TSP, trimethyl-silylpropionic acid; rms, root mean square; HSQC, heteronuclear single quantum coherence. The residue numbering is based on the sequence of intact PDI.

Introduction

Protein disulfide isomerase (PDI) is a 55 kDa multifunctional protein, which is mainly involved in disulfide bond mediated folding of proteins translocated into the endoplasmic reticulum (Freedman et al., 1994). PDI catalyzes the formation, breakage and rearrangement of disulfide bonds, which are often rate limiting steps during folding (Creighton et al., 1993). In addition, PDI is a subunit of two other enzymes: prolyl-4-hydroxylase, which has an $\alpha_2\beta_2$ quaternary structure (Pihlajaniemi et al., 1987), and the $\alpha\beta$ heterodimeric triglyceride transfer protein complex (Wetterau et al., 1990). In these enzymes PDI appears to have a more structural role, preventing the catalytic α subunits from aggregating.

The three-dimensional structure of the intact PDI has not been solved yet, although the protein has been known for over 30 years (Goldberger et al., 1963).

*Coordinates of the ***b*** domain of protein disulfide isomerase have been deposited in the Brookhaven Protein Data Bank (Filename: 2BJX.PDB).

**To whom correspondence should be addressed. E-mail: johan.kemmink@oulu.fi

Supplementary information: a table of the ^{1}H, ^{13}C, and ^{15}N assignments of PDI ***b*** domain is available from the corresponding author on request.

A major breakthrough in understanding the properties of PDI was the determination of its amino acid sequence (Edman et al., 1985). Two segments of the sequence that comprise about half the protein, designated ***a*** and ***a***′, are clearly homologous to each other (47% identity in part of their sequences) and to the ubiquitous redox protein thioredoxin (TRX). Each of these domains contains a -Cys-Gly-His-Cys- active-site sequence motif, which is very similar to the TRX active site (-Cys-Gly-Pro-Cys-) and cycles between the dithiol and the disulfide form during catalysis (Kallis and Holmgren, 1980; Hawkins and Freedman, 1991; Lundström and Holmgren, 1993; Holmgren, 1995; Darby and Creighton, 1995a,b). Studies on protein constructs comprising only the ***a*** or ***a***′ domain have provided direct insight into the active-site properties of PDI (Darby et al., 1995b). Structural studies using NMR methods confirmed the expected structural similarity of the isolated ***a*** domain with that of thioredoxin (Kemmink et al., 1995, 1996).

Previous work on the ***a*** and ***a***′ domain revealed that these domains alone can account for at least some of the involvement of PDI in catalyzing disulfide bond formation in unfolded (poly-) peptides (Darby and Creighton, 1995a; Darby et al., 1996). The isolated ***a*** and ***a***′ domains, however, are less efficient catalysts of disulfide bond reduction and rearrangements, particularly in (partly) folded substrate molecules (Darby and Creighton, 1995a; Darby et al., 1996). These observations suggest that other segments of PDI are required for its full activity.

Until recently, the structure and functions of the middle section of the PDI sequence were not well characterized. Edman et al. (1985) suggested that the section between the ***a*** and ***a***′ domains consists of two segments that are homologous to each other and designated these as the ***b*** and ***b***′ domains. Studies combining protein engineering, limited proteolysis and NMR methods confirmed the original domain assignment of Edman et al. (1985), although with different limits of the domain boundaries (Darby et al., 1996; Kemmink et al., 1996). A subsequent modification suggested by Tsibris et al. (1989) was ruled out. Different multi-domain constructs, combining the catalytic centers of ***a*** or ***a***′ domains with the other domains, were characterized with respect to their disulfide bond reduction and rearrangement activities by Darby et al. (1998). It appeared that the ***b***′ domain plays a particularly important role although only the ***a-b-b***′***-a***′ construct had an activity that was fully comparable to native PDI. The acidic C-terminus of PDI (Asp^{463}-Leu^{491}) as well as the two unpaired cysteine residues in the ***b***′ domain (Cys^{295} and Cys^{326}) are not involved in the catalytic process (Darby et al., 1998). Cross-linking studies revealed that the ***b***′ domain is an essential and sufficient component for the binding of small peptides. For the binding of larger peptides or misfolded proteins the presence of additional domains is required (Klappa et al., 1998).

In the absence of 3D structural information the mechanism of substrate binding remains unknown. Preliminary NMR structural studies of the ***b*** domain of PDI showed that this domain also adopts a fold very similar to that of TRX (Kemmink et al., 1997). Considering the very limited sequence homology between both the ***a*** and ***a***′ domains and the ***b*** domain, this is a very surprising result and indicates that PDI is essentially constructed of linked active and inactive thioredoxin modules (Kemmink et al., 1997).

The previous NMR studies of the ***a*** and ***b*** domain of PDI (Kemmink et al., 1995, 1996, 1997) revealed that these domains are connected with a short linker that consists of only residues between Thr^{116} and Ala^{120}. Whereas the exact boundaries of the ***a***′ domain can be derived from the existing sequence homology between the ***a*** and ***a***′ domains, the boundary between the ***b*** and ***b***′ domains is less certain on the basis of previous studies. A substantial segment located at the C-terminus appeared to be unfolded in a preliminary structural study of a ***b*** domain construct comprising residues Ala^{119}–Ala^{228} of mature PDI (Kemmink et al., 1997). Here we present structural models on the basis of additional experimental restraints combined with results from relaxation studies to assess the structural properties and the dynamical behaviour of the ***b*** domain of PDI in more detail.

Materials and methods

Materials

The cloned gene for human PDI (Pihlajaniemi et al., 1987) was kindly provided as cDNA clone S-138 by K.I. Kivirikko (University of Oulu, Finland). The PDI ***b*** domain was produced as described elsewhere (Darby et al., 1996).

Sample preparation

The methods to produce uniformly ^{15}N-labeled and $^{13}C/^{15}N$-labeled PDI ***b*** domain were similar to those used previously for the PDI ***a*** domain (Kemmink et

Table 1. Acquisition parameters of the NMR experiments

Experiment	ω_1					ω_2					ω_3				ns	Spectrometer
	Nucleus	Carrier	sw	t_{max}	Mod	Nucleus	Carrier	sw	t_{max}	Mod	Nucleus	Carrier	sw	t_{max}		
Spectral assignments																
CBCANH	$^{13}C^{\alpha/\beta}$	48.5	12000	5.3	TP	^{15}N	121.4	3000	21.3	TP	$^{1}H^{N}$	H_2O	8205	62.4	8	Varian
CBCA(CO)NH	$^{13}C^{\alpha/\beta}$	48.5	12000	5.3	TP	^{15}N	121.4	3000	17.8	TP	$^{1}H^{N}$	H_2O	8205	62.4	4	Varian
HBHA(CBCACO)NH	$^{1}H^{\alpha/\beta}$	3.17	4103	15.6	TP	^{15}N	121.4	3000	22.3	TP	$^{1}H^{N}$	H_2O	8205	62.4	4	Varian
H(C)CH-COSY	^{1}H	2.51	5000	25.6	TP	^{13}C	48.1	12000	5.3	SE	^{1}H	2.51	5000	102.4	4	Varian
HC(C)H-TOCSY	^{1}H	2.51	5000	19.2	TP	^{13}C	48.1	12000	5.8	TP	^{1}H	2.51	5000	128.0	8	Varian
(HB)CB(CGCD)HD						$^{13}C^{\beta}$	35.7	6000	8.0	ST	$^{1}H^{\delta}$	H_2O	8008	64.0	128	Varian
(HB)CB(CGCDCE)HE						$^{13}C^{\beta}$	35.7	6000	8.0	ST	$^{1}H^{\varepsilon}$	H_2O	8008	64.0	512	Varian
HB(CBCGCD)HD						$^{1}H^{\beta}$	H_2O	8000	64.0	ST	$^{1}H^{\delta}$	H_2O	8008	64.0	96	Varian
HC(C)H-TOCSY	$^{1}H^{aro}$	H_2O	4004	32.0	TP	$^{13}C^{aro}$	128.5	6000	10.7	TP	$^{1}H^{aro}$	H_2O	8008	64.0	4	Varian
HNCO	^{15}N	117.0	2500	25.6	TP	$^{13}C'$	174.9	4000	16.0	TP	$^{1}H^{N}$	H_2O	8008	64.0	4	Varian
Dihedral restraints																
HNHA	^{1}H	H_2O	8197	19.5	TP	^{15}N	121.6	1786	22.4	TP	$^{1}H^{N}$	H_2O	8197	62.5	8	Bruker
HNHB	^{1}H	H_2O	8008	14.9	TP	^{15}N	117.0	2500	25.6	TP	$^{1}H^{N}$	H_2O	8008	63.9	4	Varian
^{15}N-TOCSY-HSQC	^{1}H	H_2O	8197	19.5	TP	^{15}N	121.6	1786	22.4	TP	$^{1}H^{N}$	H_2O	8187	62.5	8	Bruker
Distance restraints																
^{15}N-NOESY-HSQC	^{1}H	H_2O	8197	19.5	TP	^{15}N	121.6	1786	22.4	TP	$^{1}H^{N}$	H_2O	8187	62.5	8	Bruker
^{13}C-NOESY-HSQC	^{1}H	H_2O	8008	16.0	TP	^{13}C	39.0	12000	9.3	SE	^{1}H	H_2O	8008	63.9	4	Varian
Relaxation																
^{1}H-^{15}N NOE						^{15}N	117.0	2500	12.8	ST	^{1}H	H_2O	8008	64.0	80	Varian

Indicated are the type of nucleus detected, the position of the radio-frequency carrier in ppm or coinciding with the water frequency (H_2O), the size of the spectral window sw in Hz, the maximum acquisition time t_{max} in ms, mode of indirect detection (TP, TPPI; ST, States-TPPI; SE, sensitivity enhanced; see Cavanagh et al., 1996), the number of scans ns accumulated and the type of spectrometer used (Varian Unity Inova 600; Bruker AMX 600).

Table 2. Structural statistics of 24 PDI ***b*** domain structures analyzed

Distance restraints	
Unambiguous	1512
Intra-residue	409
Sequential	440
Non-sequential ($\|i - j\| \geq 2$ and $\|i - j\| \leq 4$)	296
Non-sequential ($\|i - j\| \geq 5$)	367
Ambiguous	272
Dihedral angle restraints	
ϕ	52
χ^1	29
Deviations from experimental restraints[a]	
Distance restraints (pm)[b]	
Unambiguous NOEs	2.00 (0.02)
Ambiguous NOEs	0.8 (0.2)
All NOEs	1.87 (0.01)
Dihedral angle restraints (°)[c]	
All	0.27 (0.05)
Deviations from idealized covalent geometry	
Bonds (pm)	0.19 (0.06)
Angles (°)	0.499 (0.005)
Impropers (°)	0.404 (0.004)
E_{repel}(kJ/mol)	46 (7)
Quality control	
Procheck (residues 120–215)	
Residues in most favored regions	80%
Residues in additional allowed regions	18%
Residues in generously allowed regions	2%
Residues in disallowed regions	0%
Deviations from the average structure (nm)	
Residues 120–215	
Backbone	0.047 (0.007)
All heavy atoms	0.082 (0.006)
Residues 120–159, 166–199 and 204–215	
Backbone	0.041 (0.009)
All heavy atoms	0.069 (0.007)

[a]The standard deviations are in parentheses.
[b]No NOE violation > 0.05 nm.
[c]No dihedral angle violation > 5°.

al., 1995, 1996). Both samples were dissolved to concentrations of 1–2 mM in 93% H_2O/7% D_2O (v/v) containing 10 mM sodium phosphate buffer at pH 6.5.

NMR spectroscopy

1. General

All NMR spectra were recorded at 300 K on a Varian Unity INOVA 600 (GBB) or a Bruker AMX600 spectrometer (EMBL). Both spectrometers were equipped with self-shielded triple-resonance z-gradient probes. In all experiments gradients were employed for suppression of artifacts and the solvent resonance (Bax and Pochapsky, 1992). In some experiments PFGs were used for gradient selection of the desired coherence pathway, using the sensitivity enhancement schemes described by Cavanagh et al. (1991), Palmer et al. (1991), and Kay et al. (1992).

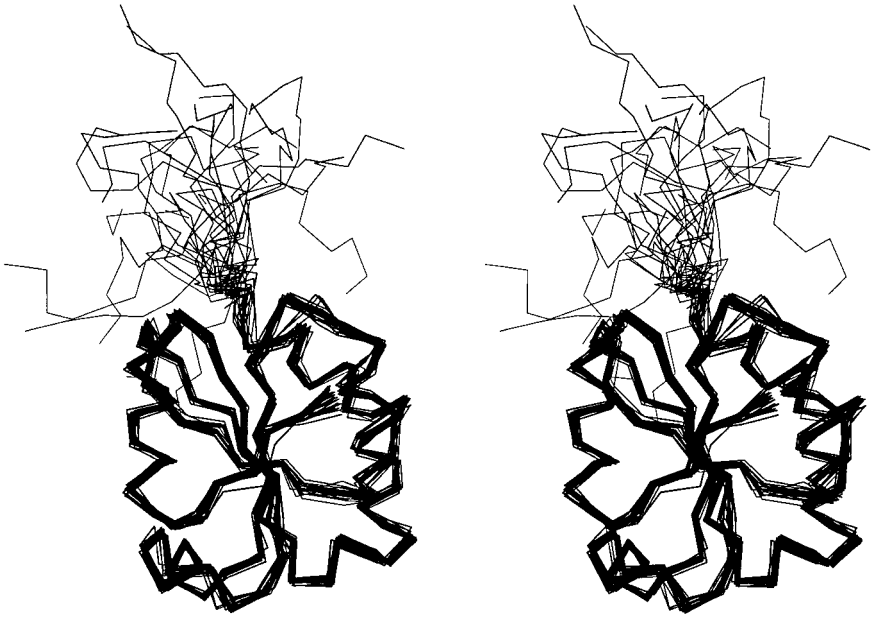

Figure 1. Stereoscopic representation showing the C^{α} traces of the ensemble of 24 structures of PDI ***b*** selected for structural analysis. The structures were superimposed with respect to the backbone N, C^{α}, and C' atoms of the ordered regions in the complete ensemble (Nilges et al., 1987). The orientation of the traces is the same as in Figure 2A.

The NMR data collected for the resonance assignments and structure determination of PDI ***b*** were processed and analyzed using the program Snarf v0.8.9 (Frans van Hoesel, University of Groningen). Linear prediction was used to extend the number of time domain points in the indirectly detected dimensions of 3D experiments. Before Fourier transformation, the resulting time domain data were multiplied by a suitable window function (Lorentz-Gauss or shifted sine bell). The $\{^1H\}$-^{15}N heteronuclear NOE data were processed with the program NMRPipe (Delaglio et al., 1995) and analyzed with the program NMRView v3 (Johnson and Blevins, 1994).

Sequential resonance assignments of PDI ***b*** were accomplished using a limited number of heteronuclear 2D and 3D experiments. Structural restraints were derived from ^{13}C- and ^{15}N-resolved 3D NOE spectra as well as spectra providing coupling constant information. The acquisition parameters of all the experiments are compiled in Table 1. The ^1H carrier was always positioned at the water resonance unless indicated otherwise. Carrier frequencies of ^{13}C and ^{15}N are given in ppm relative to TSP (Bax and Subramanian, 1986) or to liquid NH_3 (Live et al., 1984), respectively.

2. Spectral assignments

For assignment of the frequencies belonging to the $^1H^N$, ^{15}N, $^{13}C^{\alpha/\beta}$ and $^1H^{\alpha/\beta}$ nuclei, 3D CBCANH (Grzesiek and Bax, 1992a), CBCA(CO)NH (Grzesiek and Bax, 1992b) and HBHA(CBCACO)NH (Grzesiek and Bax, 1993) spectra were recorded. These three experiments were almost sufficient for unambiguous assignment of most backbone and C^{β}/H^{β} nuclei. Assignment of more complicated aliphatic spin-systems was accomplished by interpretation of a modified version of the H(C)CH-COSY (Bax et al., 1990) and the HC(C)H-TOCSY (Kay et al., 1993) experiments using the previous assignments of $^{13}C^{\alpha/\beta}$ and $^1H^{\alpha/\beta}$ nuclei as starting points. Coherence mixing in the HC(C)H-TOCSY experiment was accomplished by the use of a 24 ms DIPSI-3 mixing sequence (Shaka et al., 1988).

The aromatic side chains were assigned using 2D experiments, which correlate the $^{13}C^{\beta}/^1H^{\beta}$ resonances with the protons of the aromatic moiety, plus a 3D HC(C)H-TOCSY with the ^{13}C carrier placed in the aromatic region in order to connect the aromatic spin-systems. To correlate the $^{13}C^{\beta}$ resonances with the $^1H^{\delta/\epsilon}$ resonances 2D (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE (Yamazaki et al., 1993) spectra were recorded. Correlation of the aromatic δ protons with the β protons was accomplished with the 2D HB(CBCGCD)HD experiment, a modified ver-

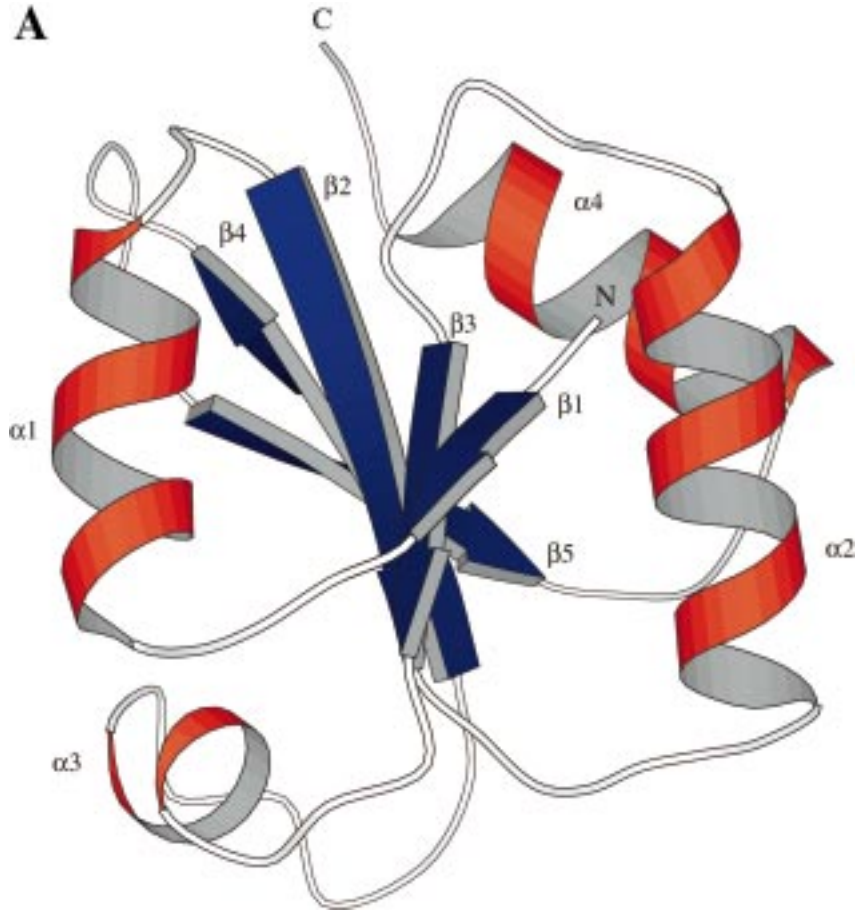

Figure 2. MOLSCRIPT (Kraulis, 1991) representations of one of the structures of the ensemble of 24 structures of PDI ***b***. The sequential order in the sequence of the β-strands (blue) and the α-helices (red) is indicated. The disordered C-terminal region of the protein is omitted from the drawing. The views in A and B are rotated by 90° with respect to each other.

sion of the original (HB)CB(CGCD)HD experiment (Yamazaki et al., 1993). Coherence mixing in the HC(C)H-TOCSY experiment was accomplished using a 15.6 ms DIPSI-3 mixing sequence.

Finally, the chemical shifts of the backbone carbonyl resonances were obtained from an HNCO (Grzesiek and Bax, 1992c) experiment using the previously determined backbone ^{15}N and ^{1}H^{N} assignments.

3. Coupling constants

Vicinal $^{3}J_{\mathrm{HNH}\alpha}$ coupling constants were measured from a 3D HNHA experiment (Kuboniwa et al., 1994). Side-chain dihedral angles χ_1, together with β-methylene proton stereospecific assignments, were determined for a number of residues by estimation of $^{3}J_{\mathrm{H}\alpha\mathrm{H}\beta}$- and $^{3}J_{\mathrm{NH}\beta}$-coupling constants from a 3D ^{15}N-TOCSY-HSQC experiment recorded with a short coherence mixing time (τ_m = 30 ms; Clore et al., 1991), and a 3D HNHB experiment (Archer et al., 1991), respectively. Coherence mixing in the ^{15}N-TOCSY-HSQC experiment was accomplished by a modified MLEV16 mixing scheme (Kadkhodaei et al., 1993).

4. Collection of NOE data

^{15}N-NOESY-HSQC data sets (Marion et al., 1989; Zuiderweg and Fesik, 1989) were recorded with two different mixing times (τ_m = 100, 150 ms). Furthermore, a ^{13}C-NOESY-HSQC experiment (Muhandiram et al., 1993) was recorded (τ_m = 100 ms).

5. Structural restraints

Structural restraints were derived from two sources: ^{3}J-coupling constants and NOEs. Analysis of the 3D HNHA experiment resulted in 52 ϕ angle restraints.

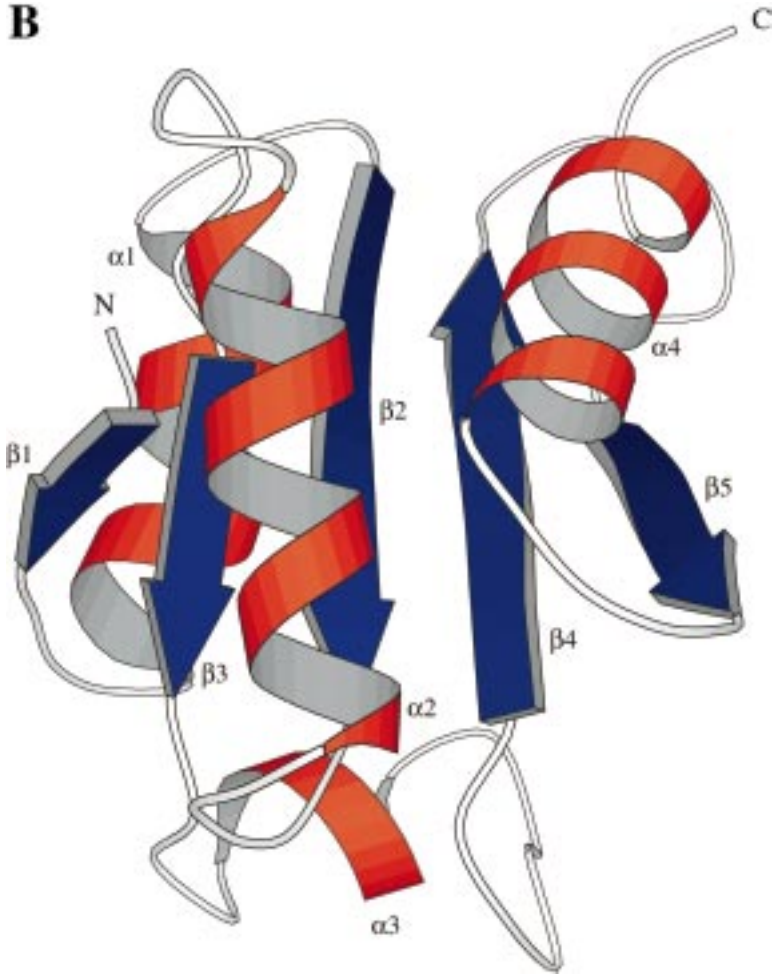

Figure 2. (continued).

The angle ϕ was restrained to -60 (± 30)° if $^{3}J_{\mathrm{HNH}\alpha} < 5$ Hz or to -120 (± 30)° if $^{3}J_{\mathrm{HNH}\alpha} > 8$ Hz.

Analysis of the ^{15}N-TOCSY-HSQC and HNHB spectra resulted in a total of 29 β-methylene proton stereospecific assignments and χ_1 dihedral angles restraints. Depending on the relative values of $^{3}J_{\mathrm{H}\alpha\mathrm{H}\beta}$ and $^{3}J_{\mathrm{NH}\beta}$, χ_1 dihedral angles were restrained to -60 (± 30)°, 60 (± 30)° or 180 (± 30)°.

NOE distance restraints were collected from three different spectra: a ^{15}N-NOESY-HSQC and a ^{13}C-NOESY-HSQC spectrum recorded with mixing times of 100 ms and a ^{15}N-NOESY-HSQC spectrum recorded with a mixing time of 150 ms. Assignment of NOEs was accomplished by comparison of peak positions against a database containing the spectral assignments as implemented in the program Snarf. Interpretation of ambiguous restraints was handled automatically using the ARIA protocol (Nilges, 1995; Nilges et al., 1997). The distances derived from the ^{15}N-NOESY-HSQC spectra were re-calibrated in each iteration of the ARIA procedure by comparison with all the ^{1}H-^{1}H distances of the models calculated in the previous cycle. Cross-peaks in the ^{13}C-NOESY-HSQC spectrum were classified as strong, medium or weak yielding distance bounds of 0.18–0.25, 0.18–0.35/0.38 and 0.18–0.50/0.55 nm, respectively. The slightly higher upper bounds were used for the limited number of ambiguous NOEs derived from this spectrum.

6. NMR relaxation experiments

The steady-state {^{1}H}-^{15}N heteronuclear NOEs were determined by recording spectra with and without saturation of the amide protons during the relaxation delay of 3.5 s (Kay et al., 1989). Errors in the NOEs

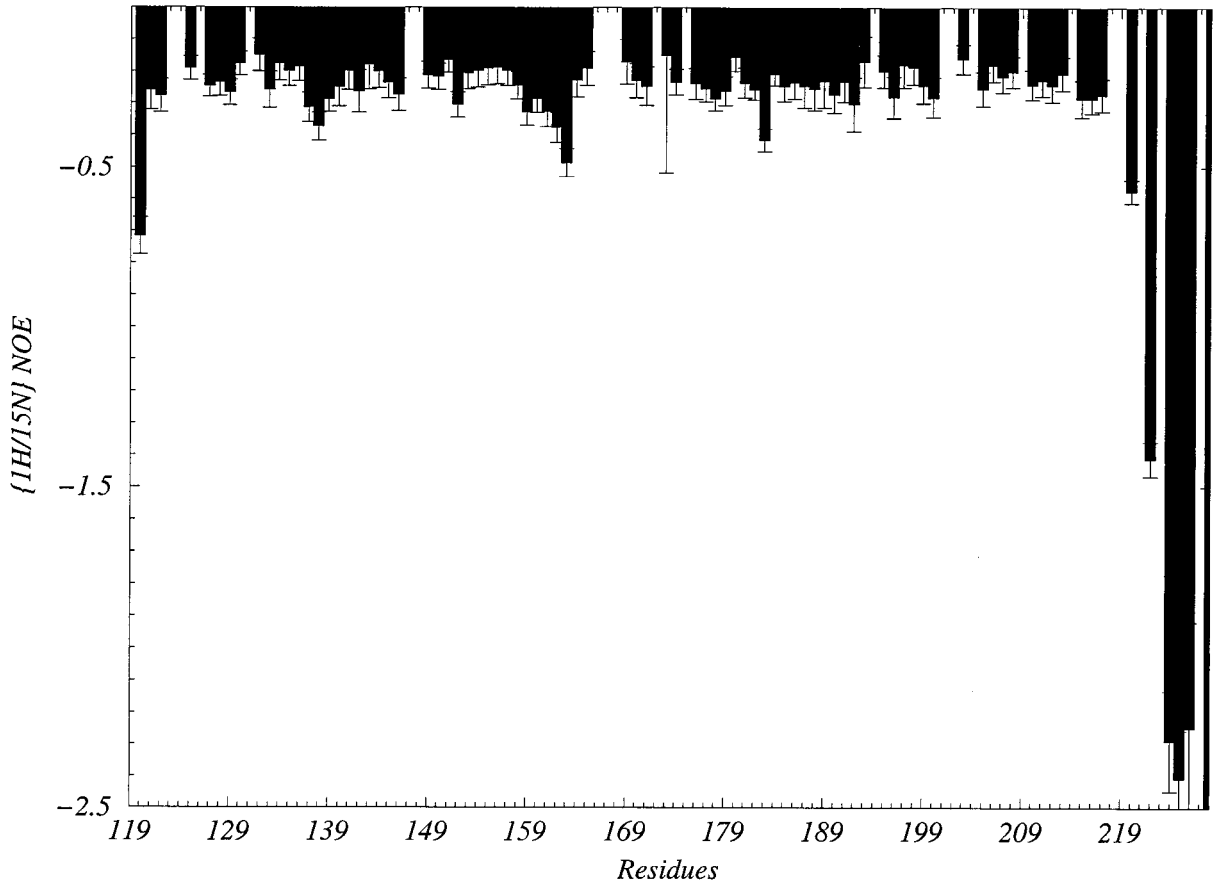

Figure 3. The backbone steady-state $\{^1H\}$-^{15}N heteronuclear NOEs measured for PDI ***b***. Error bars indicate the estimated standard deviations.

were estimated from the propagation of the noise on the measured peak intensities. Only peaks that show no overlap were considered.

Structure calculations

First, a set of 50 structures was generated on the basis of unambiguously assigned NOE distance restraints, derived from the 100 ms $^{13}C/^{15}N$-NOESY-HSQC spectra using a hybrid distance geometry/simulated annealing approach (Nilges et al., 1988) with an extended version of X-PLOR (Brünger, 1992). Floating assignments were used for prochiral groups. Subsequently, the initial number of distance restraints was augmented by an automated iterative assignment scheme as described previously for PDI ***a*** (Kemmink et al., 1996). Ambiguous distance restraints were derived from the 150 ms ^{15}N-NOESY-HSQC spectrum and the 100 ms ^{13}C-NOESY-HSQC spectra. In this stage also the dihedral angle restraints were included in the calculations. The iterative assignment procedure resulted in a final restraint list of 1512 unambiguous and 272 ambiguous distance restraints (Table 2). Finally, the unambiguous distance restraints were used to generate 400 structures with the hybrid distance geometry/simulated annealing approach. The ambiguous NOE restraints and the dihedral angle restraints were included in the refinement part of this round of calculations.

Results

The domain studied here (PDI ***b***) comprises residues Ala^{119}–Ala^{228} of human PDI, plus an initiating methionine residue at the N-terminus. The protein was expressed in *Escherichia coli* in unlabeled, ^{15}N-labeled and $^{13}C/^{15}N$-labeled forms. The NMR spectra of PDI ***b*** were indicative of a folded protein in its na-

tive state and its structure could be determined at pH 6.5 at 300 K.

Resonance assignments

The NMR assignments of the backbone $^{1}H^{N}$, $^{13}C^{\alpha}$, ^{15}N, and the side-chain $^{13}C^{\beta}$ atoms of PDI ***b*** were accomplished almost exclusively on the basis of the CBCANH and the CBCA(CO)NH experiments (vide supra). A few ambiguities were resolved by interpretation of other spectra recorded on PDI ***b***. Most of the other nuclei were assigned using the HC(C)H-TOCSY and H(C)CH-COSY experiments. The aromatic side-chain protons were assigned using experiments, which correlate these protons with intra-residue $^{1}H^{\beta}$ and $^{13}C^{\beta}$ frequencies, plus an HC(C)H-TOCSY experiment with the ^{13}C radiofrequency carrier placed in the aromatic region. Using these spectra complete assignments of the aromatic side chains could be obtained except for Phe^{189}, Phe^{192}, and Phe^{211}. A table with all the assigned ^{1}H, ^{13}C and ^{15}N resonance frequencies of PDI ***b*** is made available as supplementary information.

Structural calculations

The final set of refined structures of the PDI ***b*** domain was calculated from 1512 unambiguous and 272 ambiguous NOE restraints plus 81 dihedral angle restraints. Twenty-four structures with the lowest total energy were selected for structural analysis. The structural statistics are compiled in Table 2. The C^{α} traces of these 24 structures superimposed with respect to the backbone atoms are shown in Figure 1. The structures are well defined along the complete polypeptide chain, except for the C-terminal region ranging from residues Gln^{216} to Ala^{228}. The average rms deviation for the well-ordered part with respect to the average structure amounts to 0.047 ($\pm$0.007) nm for the backbone N, C^{α} and C' atoms and 0.082 ($\pm$0.006) nm for all heavy atoms. Slightly lower values of 0.041 ($\pm$0.009) nm and 0.069 ($\pm$0.007) nm, respectively, were obtained when two loop regions (Glu^{160}–Ile^{165} and Glu^{200}–Val^{203}) were excluded from the comparison (Table 2). Ramachandran plot analysis using Procheck-NMR (Laskowski et al., 1996) revealed that virtually all the residues (98%) in the well-ordered segment (Ala^{120}–Asn^{215}) are within the boundaries of the most favoured and the additionally allowed regions (Table 2).

Description of the structure

The three-dimensional fold of the PDI ***b*** domain is clearly very similar to that of TRX and the PDI ***a*** domain (Kemmink et al., 1997). It consists of a central core made up of a five-stranded β-sheet surrounded by four α-helices (Figure 2A). The structure can be divided into two sub-domains: The N-terminal sub-domain (βα)βαβ comprising residues Ala^{120}–Ser^{171} and the C-terminal sub-domain ββα comprising residues Asp^{184}–Asn^{215}. The two domains are connected via the short helix α3 (Figure 2B). The initial β-strand and α-helix (βα) are not considered part of the general thioredoxin motif, as they are not present in other members of this family (Martin, 1995).

Relaxation data

The $\{^{1}H\}$-^{15}N heteronuclear NOEs were determined for approximately 85% of the residues observable in the ^{15}N-^{1}H HSQC spectrum. The results are compiled in Figure 3. The data reveal that the internal backbone flexibility of PDI ***b*** is restricted and uniform over the whole protein except for the C-terminal region. The data points obtained starting at residue Val^{220} towards the C-terminus indicate increasingly unrestricted motions with respect to the protein core and true disorder. In addition to the C-terminal region also regions around residues Val^{138}, Ile^{162}, Lys^{183} in the ordered part of PDI ***b*** show a slight increase in flexibility. These residues are close to the linker regions α1/β2, α2/β3, and α3/β4, respectively. A few residues at the N-terminus are more flexible compared to average as indicated by the NOE obtained for Ala^{120}.

Discussion and conclusions

The sequence of PDI ***b*** comprises only about 20% out of the complete sequence of human PDI. Yet it behaves as an independently folded domain in the PDI structure sharing all the elements of secondary structure and the global fold of the PDI ***a*** domain and TRX. The structure of PDI ***b*** could be determined at a high level of precision and it confirms the features found in the low-resolution structure obtained earlier (Kemmink et al., 1997).

The domain boundaries as determined from limited proteolysis studies (Darby et al., 1996) and the solution structure of PDI ***a***, which precedes the ***b*** domain

in the PDI sequence, are consistent with the results obtained from this study. The first element of secondary structure (strand β1) of PDI ***b*** starts at residue Ala^{120}. The C-terminal helix α4 of PDI ***a*** terminates at residue Thr^{116} (Kemmink et al., 1996) leaving only a very short linker region between domain ***b*** and domain ***a***. The intervening residues -Gly^{117}-Pro^{118}-(Ala^{119})- are typically found in β-hairpin type conformations.

The C-terminus of PDI ***b*** appears to be unstructured (Figure 1) as indicated by a lack of medium- and long-range NOEs. Inspection of the {^{1}H}-^{15}N heteronuclear NOE data proves that the absence of such NOEs is caused by real disorder and is not due to trivial circumstances like spectral overlap. The limited proteolysis studies performed on the fragment Thr^{100}–Asn^{257} (Darby et al., 1996) revealed cleavage sites before Leu^{219} and Phe^{223}, which can be rationalized on the basis of these data. Moreover, this observation defines the limits of the ***b***′ domain of PDI with significant precision. Considering the ***a*** and ***a***′ domains as internal sequence repeats the ***b***′ domain must be comprised in the sequence between Leu^{219} and Gln^{350}.

It is surprising that structurally the PDI ***b*** can be considered as a sequence repeat of the preceding PDI ***a*** domain despite the very low level of identity (11%) in the primary sequence. In contrast the PDI ***a*** and ***a***′ are classical sequence repeats sharing 47% of identical amino acids in part of their sequences (Edman et al., 1985). This observation and the limited sequence homology between ***b*** and ***b***′ made us speculate that mature PDI is constructed of active and inactive thioredoxin modules adapted during evolution to provide PDI with its complete spectrum of catalytic activities (Kemmink et al., 1997; Darby et al., 1998).

Although the PDI ***b*** domain preserved the thioredoxin fold, it lost a number of amino acid residues that are typical for members of the thioredoxin family. Not only the characteristic active-site sequence motif -Cys-Xxx-Yyy-Cys- has been deleted from the sequence, but also the conserved *cis*-proline residue at the start of strand β4. The *cis*-proline is located near the active site cysteines (Figure 4) and stabilized by tertiary interactions. The PDI ***b*** domain adopts a very similar structure without a *cis* peptide bond present. This strongly indicates that the *cis* proline is an important feature of proteins belonging to the thioredoxin family defining the properties of their active sites.

Another conserved feature in the thioredoxin family is the presence of a proline residue in helix α2, altering its regular helical structure. In the PDI ***b*** sequence it is replaced by a lysine residue. The active site sequences of PDI ***a*** (-Cys-Gly-His-Cys-) and hTRX (-Cys-Gly-Pro-Cys-) are located at the N-terminus of this helix, which will affect the electrostatic properties of the active site determining its redox potential (Gane et al., 1994; Kortemme and Creighton, 1995). In the β-strand preceding these active sites a conserved acidic residue (Glu in PDI ***a*** and Asp in hTRX) is replaced by a Gly in PDI ***b***. In the PDI ***a*** structure (Kemmink et al., 1996) the side chain of this residue is buried and close in space to the active site. Its pK_a is perturbed as at the experimental pH (5.1) it appeared to be the only acidic amino acid in a protonated, uncharged form (Kemmink et al., unpublished results). Similar observations have been made with thioredoxins (Jeng and Dyson, 1996; Qin et al., 1996).

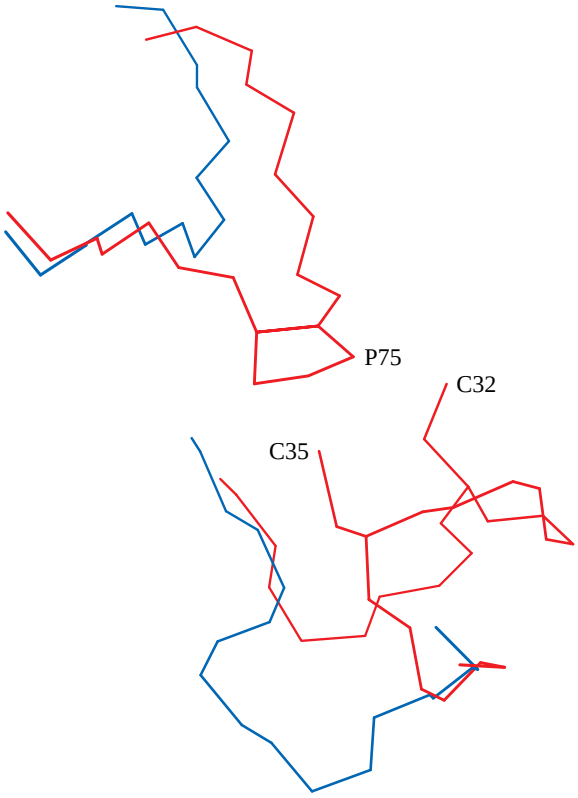

Figure 4. Alignment of the backbone of the active-site region of hTRX (mutant protein with Cys^{73} replaced by Ser; Brookhaven Protein Data Bank accession code: 1erv) in red and the corresponding region of PDI ***b*** in blue. The side chains of the active-site cysteines Cys^{32} and Cys^{35} plus the cis-proline Pro^{75} of hTRX are shown as well. The alignment was made using the program DALI (Holm and Sander, 1993, 1994).

PDIs and thioredoxins transfer disulfide bonds to and from substrate protein molecules. They react more rapidly with protein substrates than with comparable small thiol and disulfide molecules (Darby and Creighton, 1995a,b). Non-covalent, non-specific interactions are believed to be responsible for these rate enhancements. Possible interaction sites consisting of hydrophobic patches located around the solvent exposed reactive cysteine have been identified in both hTRX (Qin et al., 1995) and PDI ***a*** (Kemmink et al., 1996). The molecular surface of PDI ***b*** does not show a distinct hydrophobic patch in a region comparable to hTRX or PDI ***b***. However, other potential hydrophobic binding sites on the protein surface of PDI ***b*** can be distinguished. Such sites might be involved in substrate binding, but they might also be involved in interactions with other PDI domains. The refined structure of the ***b*** domain provides interesting insights into the evolution of PDI structure and function and serves as a basis for answering some remaining questions concerning substrate binding and the arrangement of domains in PDI. It is another example of how far two sequences can diverge and yet retain the same 3D structure.

At a functional level duplication of the thioredoxin module has had important functional consequences, but the exact molecular basis for this remains unclear. Interesting in this respect is a recent report of Wang et al. (1998) describing the crystal structure of rabbit skeletal calsequestrin, the major Ca^{2+} storage protein of muscle. Previously a significant sequence homology between calsequestrin and PDI ***b*** had been identified (Darby et al., 1996) suggesting that calsequestrin may also contain thioredoxin-like domains. This is confirmed by the 3D structure of intact calsequestrin, which consists of three highly negatively charged thioredoxin-like domains, which are connected via short linkers. Calsequestrin appears from this study (Wang et al., 1998) to be another example of linked thioredoxin modules adapted to perform specific functions other than thiol-disulfide exchange reactions alone. Its domain architecture is reminiscent of that of PDI (Kemmink et al., 1997) and may provide insight into the 3D arrangement of domains in PDI. The availability of the refined structures of calsequestrin and of the ***b*** domain plus its detailed characterization by NMR should provide an important basis for answering questions about the structure and function of PDI by allowing us to map sites of domain-domain and domain-substrate interactions.

The ***b*** domain is an independently folding domain of PDI. A refined ensemble of solution structures based on the input of 1865 structural restraints shows that its structure is well defined throughout the complete protein except for about 10 residues at the C-terminus of the sequence. These appear to be disordered as assessed by the results of ^{15}N relaxation data. This allows us to define the boundaries of the structural domains of PDI with high accuracy. Structurally the PDI ***b*** domain is a copy of the PDI ***a*** domain. During evolution it retained the thioredoxin fold, but it lost a number of features essential for the thiol-disulfide exchange reactions catalyzed by PDI ***a*** and ***a′***. Its high-resolution structure presented here will contribute to our understanding of PDIs role in various functions it has required other than the catalysis of thiol-sulfide exchange reactions alone.

Acknowledgements

We acknowledge T.E. Creighton for providing the facilities to make this work possible. This work was supported by E.U. grant FMRX-CT96-0013 (M.M.). We are indebted to Vibeke Westphal for many helpful discussions.

References

Archer, S.J., Ikura, M., Torchia, D.A. and Bax, A. (1991) *J. Magn. Reson.*, **95**, 636–641.

Bax, A. and Subramanian, S. (1986) *J. Magn. Reson.*, **67**, 565–569.

Bax, A., Clore, G.M., Driscoll, P.C., Gronenborn, A.M., Ikura, M. and Kay, L.E. (1990) *J. Magn. Reson.*, **87**, 620–627.

Bax, A. and Pochapsky, S.S. (1992) *J. Magn. Reson.*, **99**, 638–643.

Brünger, A.T. (1992) *X-PLOR version 3.1. A system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.

Cavanagh, J., Palmer, A.G., Wright, P.E. and Rance, M. (1991) *J. Magn. Reson.*, **91**, 429–436.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G. and Skelton, N.J. (1996) *Protein NMR Spectroscopy*, Academic Press, San Diego, CA.

Clore, G.M., Bax, A. and Gronenborn, A.M. (1991) *J. Biomol. NMR*, **1**, 13–22.

Creighton, T.E., Bagley, C.J., Cooper, L., Darby, N.J., Freedman, R.B., Kemmink, J. and Sheikh, A. (1993) *J. Mol. Biol.*, **232**, 1176–1196.

Darby, N.J. and Creighton, T.E. (1995a) *Biochemistry*, **34**, 16770–16780.

Darby, N.J. and Creighton, T.E. (1995b) *Biochemistry*, **34**, 11725–11735.

Darby, N.J., Kemmink, J. and Creighton, T.E. (1996) *Biochemistry*, **35**, 10517–10528.

Darby, N.J., Penka, E. and Vincentelli, R. (1998) *J. Mol. Biol.*, **276**, 239–247.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.
Edman, J.C., Ellis, L., Blacher, R.W., Roth, R.A. and Rutter, W.J. (1985) *Nature*, **317**, 267–270.
Freedman, R.B., Hirst, T.R. and Tuite, M.F. (1994) *Trends Biochem. Sci.*, **19**, 331–336.
Gane, P.J., Freedman, R.B. and Warwicker, J. (1994) *J. Mol. Biol.*, **249**, 376–387.
Goldberger, R.F., Epstein, C.J. and Anfinsen C.B. (1963) *J. Biol. Chem.*, **238**, 628–635.
Grzesiek, S. and Bax, A. (1992a) *J. Magn. Reson.*, **99**, 201–207.
Grzesiek, S. and Bax, A. (1992b) *J. Am. Chem. Soc.*, **114**, 6291–6293.
Grzesiek, S. and Bax, A. (1992c) *J. Magn. Reson.*, **96**, 432–440.
Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 185–204.
Hawkins, H.C. and Freedman, R.B. (1991) *Biochem. J.*, **275**, 335–339.
Holm, L. and Sander, C. (1993) *J. Mol. Biol.*, **233**, 123–138.
Holm, L. and Sander, C. (1994) *Nucleic Acids Res.*, **22**, 3600–3609.
Holmgren, A. (1995) *Structure*, **3**, 239–243.
Jeng, M.-F. and Dyson, H.J. (1996) *Biochemistry*, **35**, 1–6.
Johnson, B.A. and Blevins, R.A. (1994) *J. Biomol. NMR*, **4**, 603–614.
Kadkhodaei, M., Hwang, T.L., Tang, J. and Shaka, A.J. (1993) *J. Magn. Reson.*, **A105**, 104–107.
Kallis, G.-B. and Holmgren, A. (1980) *J. Mol. Biol.*, **255**, 10261–10265.
Kay, L.E., Torchia, D.A. and Bax, A. (1989) *Biochemistry*, **28**, 8972–8979.
Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10665.
Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R. and Forman-Kay, J.L. (1993) *J. Magn. Reson.*, **B101**, 333–337.
Kemmink, J., Darby, N.J., Dijkstra, K., Scheek, R.M. and Creighton, T.E. (1995) *Protein Sci.*, **4**, 2587–2593.
Kemmink, J., Darby, N.J., Dijkstra, K., Nilges, M. and Creighton, T.E. (1996) *Biochemistry*, **35**, 7684–7691.
Kemmink, J., Darby, N.J., Dijkstra, K., Nilges, M. and Creighton, T.E. (1997) *Curr. Biol.*, **7**, 239–245.
Klappa, P., Ruddock, L.W., Darby, N.J. and Freedman, R.B. (1998) *EMBO J.*, **17**, 927–935.
Kortemme, T. and Creighton, T.E. (1995) *J. Mol. Biol.*, **253**, 799–812.
Kraulis, P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
Kuboniwa, H., Grzesiek, S., Delaglio, F. and Bax, A. (1994) *J. Biomol. NMR*, **4**, 871–878.
Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.
Live, D.H., Davis, D.G., Agosta, W.C. and Cowburn, D. (1984) *J. Am. Chem. Soc.*, **106**, 1939–1941.
Lundström, J. and Holmgren, A. (1993) *Biochemistry*, **32**, 6649–6655.
Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989) *Biochemistry*, **28**, 6150–6156.
Martin, J.L. (1995) *Structure*, **3**, 245–250.
Muhandiram, D.R., Farrow, N.A., Xu, G.-Y., Smallcombe, S.H. and Kay, L.E. (1993) *J. Magn. Reson.*, **B102**, 317–321.
Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317–324.
Nilges, M., Gronenborn, A.M. and Clore, G.M. (1987) *FEBS Lett.*, **219**, 11–16.
Nilges, M. (1995) *J. Mol. Biol.*, **245**, 645–660.
Nilges, M., Macias, M.J., O'Donoghue, S.I. and Oschkinat, H. (1997) *J. Mol. Biol.*, **269**, 408–422.
Palmer, A.G., Cavanagh, J., Wright, P.E. and Rance, M. (1991) *J. Magn. Reson.*, **93**, 151–170.
Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllylä, R., Huhtala, M.-J., Koivu, J. and Kivirikko, K.I. (1987) *EMBO J.*, **6**, 643–649.
Qin, J., Clore, G.M., Kennedy, W.M.P., Huth, J.R. and Gronenborn, A.M. (1995) *Structure*, **3**, 289–297.
Qin, J., Clore, G.M. and Gronenborn, A.M. (1996) *Biochemistry*, **35**, 7–13.
Shaka, A.J., Lee, C.J. and Pines, A. (1988) *J. Magn. Reson.*, **77**, 274–293.
Tsibris, J.C.M., Hunt, L.T., Ballejo, G., Barker, W.C., Toney, L.J. and Spellacy, W.N. (1989) *J. Biol. Chem.*, **264**, 13967–13970.
Wang, S., Trumble, W.R., Liao, H., Wesson, C.R., Dunker, A.K. and Kang, C. (1998) *Nat. Struct. Biol.*, **5**, 476–483.
Wetterau, J.R., Combs, K.A., Spinner, S.N. and Joiner, B.J. (1990) *J. Biol. Chem.*, **265**, 9800–9807.
Yamazaki, T., Forman-Kay, J.D. and Kay, L.E. (1993) *J. Am. Chem. Soc.*, **115**, 11054–11055.
Zuiderweg, E.R.P. and Fesik, S.W. (1989) *Biochemistry*, **28**, 2387–2391.