# Structure Determination of the N-Terminal Thioredoxin-like Domain of Protein Disulfide Isomerase Using Multidimensional Heteronuclear $^{13}C/^{15}N$ NMR Spectroscopy†

Johan Kemmink,‡ Nigel J. Darby,‡ Klaas Dijkstra,§ Michael Nilges,‡ and Thomas E. Creighton*,‡

*European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany, and The Groningen Biomolecular Science and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*

*Received February 12, 1996; Revised Manuscript Received April 8, 1996*⊗

ABSTRACT: As a first step in dissecting the structure of human protein disulfide isomerase (PDI), the structure of a fragment corresponding to the first 120 residues of its sequence has been determined using heteronuclear multidimensional NMR techniques. As expected from its primary structure homology, the fragment has the thioredoxin fold. Similarities and differences in their structures help to explain why thioredoxins are reductants, whereas PDI is an oxidant of protein thiol groups. The results confirm that PDI has a modular, multidomain structure, which will facilitate its structural and functional characterization.

Protein disulfide isomerase (PDI)[1] is a 57 kDa protein present in the endoplasmic reticulum of eukaryotes. It is involved in disulfide bond formation, breakage, and rearrangement during the folding of proteins translocated into the endoplasmic reticulum [see Freedman (1992) and Freedman et al. (1994)]. PDI contains two active sites with the amino acid sequence -Cys-Gly-His-Cys-, each of which is cycled between the dithiol and the disulfide forms as in the related redox protein thioredoxin (Freedman, 1992; Vuori et al., 1992; Lyles & Gilbert, 1991, 1994; Darby & Creighton, 1995a,b). In addition, PDI is the $\beta$ subunit of prolyl-4-hydroxylase, which has an $\alpha_2\beta_2$ quaternary structure (Pihlajaniemi et al., 1987), and it is one subunit of the heterodimeric triglyceride transfer protein complex (Wetterau et al., 1990, 1991).

The 3D structure of PDI is not yet known; its amino acid sequence (Edman et al., 1985) has provided insight into its multidomain structure and reaction mechanism. Two segments of the sequence, designated ***a*** and ***a′***, are clearly homologous to the ubiquitous single-domain protein thioredoxin (TRX). Its active-site sequence is very similar to that of the PDI active sites, -Cys-Gly-Pro-Cys-, and among many other functions, this protein reduces various cellular constituents by cycling between the dithiol and the disulfide form (Kallis & Holmgren, 1980; Holmgren, 1985). In addition to the ***a*** and ***a′*** segments, two other segments of the sequence were discerned as being internal repeats and designated ***b*** and ***b′***. The order in the primary structure of these segments is ***a-b-b′-a′*** (Edman et al., 1985). Preliminary analysis of the ***a***, ***b***, and ***a′*** domains prepared by protein engineering methods has confirmed the multidomain nature of the protein (Darby & Creighton, 1995b,c; Kemmink et al., 1995), but the exact limits of the domains must be determined experimentally (Darby et al., 1996).

As a first step in elucidating the structural basis of the function of PDI, the global fold of the human PDI-***a*** domain is determined here, using heteronuclear ($^1H$,$^{13}C$,$^{15}N$) multidimensional NMR. As expected, the structure has the thioredoxin motif, which is present in a number of related proteins (Martin, 1995). The similarities and differences in their structures provide insight into their functional similarities and differences.

## MATERIALS AND METHODS

*Materials*. The cloned gene for human PDI was very kindly provided as cDNA clone S-138 by K. Kivirikko (University of Oulu, Finland). The PDI-***a*** domain was produced as described elsewhere (Darby & Creighton, 1995a).

*Sample Preparation*. The production and purification of uniformly $^{15}N$-labeled and $^{13}C/^{15}N$-labeled PDI-***a*** have been described previously (Kemmink et al., 1995). Both samples were dissolved to 1−2 mM in either 93% $^1H_2O$/7% $^2H_2O$ (v/v) or 100% $^2H_2O$. The pH was adjusted to 5.1 by adding small amounts of NaOH.

*NMR Spectroscopy*. *(1) General*. NMR spectra were recorded at 300 K on three different spectrometers: Bruker AMX500 (EMBL), Bruker AMX600 (EMBL), and Varian Unity 500 (GBB). In addition to the NMR experiments described previously (Kemmink et al., 1995), the following spectra were recorded on a $^{13}C/^{15}N$-labeled sample dissolved

† Coordinates of the N-terminal domain of protein disulfide isomerase have been deposited in the Brookhaven Protein Data Bank (filename 1MEK).

* Corresponding author.
‡ European Molecular Biology Laboratory.
§ University of Groningen.

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

[1] Abbreviations: 2D, two dimensional; 3D, three dimensional; NMR, nuclear magnetic resonance; HSQC, heteronuclear single-quantum coherence spectroscopy; NOE, nuclear Overhauser effect; NOESY, 2D NOE spectroscopy; TOCSY, 2D total correlation spectroscopy; $^{15}N$-NOESY-HSQC, 3D $^1H$ NOE $^1H^N-^{15}N$ HSQC; $^{13}C$-NOESY-HSQC, 3D $^1H$ NOE $^1H-^{13}C$ HSQC; $^{15}N$-TOCSY-HSQC, 3D $^1H$ total correlation spectroscopy $^1H^N-^{15}N$ HSQC; CBCANH, 3D $^{13}C^{\alpha/\beta}-^{15}N-^1H^N$ correlation spectroscopy; CBCA(CO)NH, 3D $^{13}C^{\alpha/\beta}-(^{13}CO)-^{15}N-^1H^N$ correlation spectroscopy; HBHA(CBCACO)NH, $^1H^{\alpha/\beta}-(^{13}C^{\alpha/\beta}-^{13}CO)-^{15}N-^1H^N$ correlation spectroscopy; COCAH, 3D $^{13}CO-^{13}C^{\alpha}-^1H^{\alpha}$ correlation spectroscopy; H(C)CH-COSY, 3D $^1H-(^{13}C)-^{13}C-^1H$ scalar correlated spectroscopy; HC(C)H-TOCSY, 3D $^1H-^{13}C-(^{13}C)-^1H$ total correlation spectroscopy; PDI-***a***, genetically engineered protein consisting of the 120 residues at the N-terminus of human protein disulfide isomerase and an initiating methionine residue; TRX, thioredoxin; PFG, pulsed-field gradient; ppm, parts per million; rms, root mean square; TSP, trimethylsilylpropionic acid.

S0006-2960(96)00335-2 CCC: $12.00 

in $^1H_2O$ using the Varian Unity 500 spectrometer: CBCANH (Grzesiek & Bax, 1992a), CBCA(CO)NH (Grzesiek & Bax, 1992b), HBHA(CBCACO)NH (Grzesiek & Bax, 1993), H(C)CH-COSY (Bax et al., 1990), HC(C)H-TOCSY (Kay et al., 1993), and $^{13}$C-NOESY-HSQC (Muhandiram et al., 1993). Furthermore, a $^{13}$C-NOESY-HSQC spectrum was recorded on a $^{13}C/^{15}N$-labeled sample dissolved in $^2H_2O$ using the Bruker AMX600 spectrometer.

Quadrature detection in the indirect dimensions was accomplished in all experiments by time-proportional phase incrementation (Marion & Wüthrich, 1983). The $^1$H carrier frequency was always positioned at the water resonance, unless indicated otherwise. Carrier frequencies of $^{15}$N and $^{13}$C are given in parts per million relative to liquid $NH_3$ (Live et al., 1984) and TSP (Bax & Subramanian, 1986), respectively. During acquisition, heteronuclear decoupling was accomplished by application of GARP (Shaka et al., 1985) or broad-band WALTZ-16 (Shaka et al., 1983) decoupling schemes. In all cases, PFGs were used in order to suppress unwanted signals. In most cases this was accomplished using PFGs as purge pulses (Bax & Pochapsky, 1992). In the case of the H(C)CH-COSY and $^{13}$C-NOESY-HSQC, PFGs were used for gradient selection of the desired coherence pathway. Losses in sensitivity were partially compensated (Cavanagh et al., 1991; Palmer et al.,1991; Kay et al., 1992). The spectral widths in the proton dimensions were, unless indicated otherwise, 6849.3 Hz for the experiments at 500 MHz and 8196.7 Hz for the experiments at 600 MHz.

*(2) Specific Assignment of $C^\beta$/$H^\beta$s.* For assignment of the frequencies belonging to $^{13}C^\beta$ and $^1H^\beta$ nuclei, CBCANH, CBCA(CO)NH, and HBHA(CBCACO)NH spectra were recorded. The $^{15}$N carrier was placed at 116.7 ppm and the $^{13}C^{\alpha/\beta}$ carrier at 45.8 ppm. Spectral widths were 10 000 Hz ($^{13}C^{\alpha/\beta}$) or 5000 Hz ($^1H^{\alpha/\beta}$) in $\omega_1$ and 2000 Hz in $\omega_2$ ($^{15}$N). Maximum values for $t_1$, $t_2$, and $t_3$ were 6.4, 22.3, and 74.8 ms [CBCANH; 32 scans per increment]; 6.4, 32.0, and 74.8 ms [CBCA(CO)NH; 8 scans per increment]; and 12.8, 22.3, and 74.8 ms [HBHA(CBCACO)NH; 16 scans per increment].

*(3) General Side-Chain Assignments.* In order to assign the more complicated spin systems, modified versions of the H(C)CH-COSY and HC(C)H-TOCSY spectra were recorded with four and eight scans per increment, respectively. In these experiments the magnetization of the detected protons is modulated with proton frequencies belonging to the same spin system and with the $^{13}$C frequencies of either the covalently bound carbons [H(C)CH-COSY] or the carbons covalently bound to the protons that were frequency modulated in $t_1$ [HC(C)H-TOCSY]. Spectral widths of 4000 Hz in $\omega_1$ ($^1$H), 10 000 Hz in $\omega_2$ ($^{13}$C), and 8000 Hz in $\omega_3$ ($^1$H) were used. Maximum values for $t_1$ ($^1$H), $t_2$ ($^{13}$C), and $t_3$ ($^1$H) acquisition were respectively, 16.0, 6.4, and 128.0 ms in the H(C)CH-COSY experiment and 16.0, 7.5, and 128.0 ms in the HC(C)H-TOCSY experiment. Coherence mixing in the HC(C)H-TOCSY experiment was accomplished by using a 24 ms DIPSI-3 (Shaka et al., 1988) mixing sequence. The $^1$H carrier was placed at 2.52 ppm and the $^{13}$C carrier at 44.5 ppm.

*(4) Collection of Structural Data*. A number of NOESY spectra were recorded, in addition to those described previously (Kemmink et al., 1995). A $^{13}$C-NOESY-HSQC spectrum was recorded with four scans per increment at 500 MHz on a sample dissolved in $^1H_2O$, using a spectral width in $\omega_2$ ($^{13}$C) of 10 000 Hz, maximum values of $t_1$, $t_2$, and $t_3$ of 29.2, 6.4, and 74.8 ms, respectively, and a mixing time of 100 ms. The $^{13}$C carrier was placed at 44.5 ppm.

A $^{13}$C-NOESY-HSQC spectrum was also recorded using eight scans per increment at 600 MHz on a sample dissolved in $^2H_2O$. A spectral width of 10 000 Hz was used in the $^{13}$C dimension, the maximum values of $t_1$, $t_2$, and $t_3$ were 18.1, 5.3, and 62.5 ms, respectively, and the mixing time was 150 ms. The $^{13}$C carrier was placed at 40.7 ppm.

*(5) Data Processing.* Data were processed on a Silicon Graphics Iris Indigo 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 $t_1$ and $t_2$ in the case of the 3D experiments. Before Fourier transformation, the resulting time domain data were multiplied by a suitable window function (Lorentz−Gauss or shifted sine bell). After Fourier transformation in three dimensions, the following data sets were obtained ($\omega_3 \times \omega_2 \times \omega_1$): $1024 \times 128 \times 128$ ($^1H^N$, $^{15}$N, $^{13}C^{\alpha/\beta}$) real points for the CBCANH and CBCA(CO)NH experiments; $1024 \times 128 \times 128$ ($^1H^N$, $^{15}$N, $^1H^{\alpha/\beta}$) real points for the HBHA(CBCACO)NH experiment; $1024 \times 256 \times 256$ ($^1$H, $^{13}$C, $^1$H) real points for the H(C)CH-COSY and HC(C)H-TOCSY experiments; $1024 \times 256 \times 512$ ($^1$H, $^{13}$C, $^1$H) real points for the $^{13}$C-NOESY-HSQC experiment in $^1H_2O$; $512 \times 128 \times 512$ ($^1$H, $^{13}$C, $^1$H) real points for the $^{13}$C-NOESY-HSQC experiment in $^2H_2O$. The HC(C)H-TOCSY data set was also processed by applying the 2D maximum entropy method algorithm available in the program Azara v1.0 (Wayne Boucher, University of Cambridge) to the indirectly detected dimensions $t_1$ and $t_2$. The acquisition domain $t_3$ was transformed with the Fourier transform algorithm of Snarf. The resulting data set had the same size as the fully Fourier transformed spectrum.

*(6) Collection of NOE Restraints*. Assignment of NOEs in a 150 ms $^{15}$N-NOESY-HSQC spectrum was accomplished by comparison of peak positions against a database containing all assignments implemented in the program Snarf. Unambiguously assigned peaks were classified as strong (0.18−0.25 nm), medium (0.18−0.35 nm), or weak (0.18−0.50 nm) by comparison with cross-peak intensities belonging to the fixed $H^{\epsilon 1}-H^{\delta 1}$ and $H^{\epsilon 1}-H^{\zeta 2}$ distances of W35 and W111. Further distance restraints were collected from NOESY and $^{13}$C-NOESY-HSQC spectra. Restraints derived from these spectra were treated in a more conservative fashion in that they were classified as weak (0.18−0.50 nm) in almost all cases.

*(7) Hydrogen Bond Restraints.* In addition to the NOE restraints, 34 hydrogen bond restraints were used in the structure calculations. These were derived from 17 NH protons, primarily located in the $\beta$-sheet, that exchange slowly with the solvent (Kemmink et al., 1995). Hydrogen bond acceptors were derived from the observed NOE patterns (Kemmink et al., 1995). For each hydrogen bond two distance restraints ($d_{NH-O}$ and $d_{N-O}$) were defined: $0.15 < d_{NH-O} < 0.23$ nm and $0.24 < d_{N-O} < 0.33$ nm.

*Structure Calculations.* A first set of 100 structures was calculated on the basis of 599 restraints (560 assigned NOEs, 34 hydrogen bond restraints, and 5 restraints imposing a disulfide bond between Cys36 and Cys39) with a hybrid distance geometry/simulated annealing approach (Nilges et al., 1988) using an extended version of X-PLOR (Brünger, 1992). Floating assignments for prochiral groups were used.

The initial data were augmented by an automated iterative assignment scheme, using a suite of X-PLOR additions (Kharrat et al., 1995; M. Nilges, unpublished results), as follows: The 20 best structures resulting from the initial X-PLOR run were selected and used to extract additional NOE restraints from the raw data. An automatically picked peak list of a 150 ms $^{15}$N-NOESY-HSQC spectrum was converted into ambiguous distance restraints (Nilges, 1995) based on the basis of the assigned proton chemical shifts. The tolerance was $\pm$0.02 ppm in $\omega_1$. The list was then searched with the seven lowest energy structures. A restraint was used in the next iteration of structure calculation if it was violated by a varying tolerance (e.g., 0.1 nm in the first iteration) in no more than three of the seven lowest energy structures. This list was then merged with the original list of restraints to avoid duplication of entries. All 20 structures were refined with the combined list. The procedure was repeated several times, and the criteria for acceptance of new restraints were tightened in the last iterations. In each cycle, the restraints that were accepted as ambiguous were partially assigned (i.e., unlikely assignments were discarded) on the basis of previous structures, as described in Kharrat et al. (1995). Also the assignment criteria were tightened in each iteration, starting with $p = 0.95$ and ending with $p = 0.90$ [see Kharrat et al. (1995)]. Nonetheless, a restraint that was initially assigned could be reassigned in a later iteration. The inclusion/exclusion scheme was applied only to restraints derived from the peak list, not to the manually assigned/selected peaks.

The final list contained a total of 1086 unambiguous restraints and 223 ambiguous restraints, where the ambiguous restraints were summed according to their partial assignments. The unambiguous restraint list consisted of 237 NOEs that were intraresidue, 344 between adjacent residues, 175 between residues 2−4 apart in the sequence, and 291 between residues more distant in the sequence (Table 1). In addition, this list contained 34 hydrogen bond restraints and 5 disulfide bond restraints (Table 1). Details of the function of the routines for restraint extraction, assigment, distance calibration, and merging will published elsewhere (M. Nilges et al., manuscript in preparation). The restraint list obtained after 10 iterations was visually inspected and then used to calculate a final ensemble of 400 structures with the hybrid distance geometry/simulated annealing approach, of which 40 were used for the final analysis. The ambiguous restraints were used only in the refinement part of the structure calculation. The final ensemble of 40 structures showed no NOE violations greater than 0.05 nm.

Table 1: Structural Statistics for the 40 Distance Geometry/Simulated Annealing Structures Calculated for PDI-***a***

| | |
|---|---|
| distance restraints | |
| NOE restraints | |
| unambiguous | 1047 |
| intraresidue | 237 |
| sequential | 344 |
| nonsequential ($\|i-j\| \geq 2$ and $\|i-j\| \leq 4$) | 175 |
| nonsequential ($\|i-j\| \geq 5$) | 291 |
| ambiguous | 223 |
| hydrogen bond restraints | 34 |
| disulfide bond restraints | 5 |
| rms deviation from experimental distance restraints(pm)[a] | |
| unambigious NOEs | 1.7 (0.2) |
| ambiguous NOEs | 1.7 (0.5) |
| hydrogen bonds | 1.1 (0.3) |
| all NOEs | 1.7 (0.2) |
| all data | 1.7 (0.2) |
| rms deviations from the average structure (nm) | |
| well-ordered residues[b] | |
| backbone (N, C$^\alpha$, C′) | 0.079 (0.012) |
| all heavy atoms | 0.132 (0.013) |
| all residues | |
| backbone (N, C$^\alpha$, C′) | 0.151 (0.043) |
| all heavy atoms | 0.179 (0.034) |
| deviations from idealized covalent geometry | |
| bonds (pm) | 0.204 (0.009) |
| angles (deg) | 0.529 (0.007) |
| impropers (deg) | 0.387 (0.006) |
| $E_{repel}$ (kJ/mol)[c] | 63 (9) |
| Lennard-Jones van der Waals energy $E_{L-J}$ (kJ/mol)[d] | −1412 (72) |

[a] The standard deviations are in parentheses. [b] The well-ordered residues comprise residues 4−54 and 58−116 of PDI-***a***. [c] $E_{repel}$ represents a repulsive van der Waals energy term, which is included in the target function for simulated annealing (Nilges et al., 1988). [d] Calculated with the CHARMM (Brooks et al., 1983) empirical energy function, which is not included in the target function for simulated annealing.

## RESULTS

The PDI-***a*** domain studied here consisted of residues 1−120 of human PDI, plus an initiating Met −1 residue. It was expressed in *Escherichia coli* as a stable, folded protein and could be purified by standard procedures (Darby & Creighton, 1995c). The form with the disulfide bond between the two active-site cysteine residues was studied here. It gave NMR spectra typical of a folded, native protein, and its structure could be determined at pH 5.1 and 300 K using standard techniques.

The NMR assignments of the backbone $^1$H, $^{13}$C, and $^{15}$N atoms of PDI-***a*** have been reported previously (Kemmink et al., 1995). However, the amide protons of residues Trp35, Gly 37, and His38 could not be found in 2D $^1$H−$^{15}$N HSQC spectra, while the signal of Cys36 was very weak. This has precluded accurate structural determination of these active-site residues.

*Side-Chain Assignments.* The $^1$H, $^{13}$C, and $^{15}$N resonances of the side chains were assigned using primarily the HC(C)H-TOCSY and H(C)CH-COSY spectra. Starting points for the assignment procedure were the $^1$H$^\alpha$ and $^{13}$C$^\alpha$ chemical shifts determined previously (Kemmink et al., 1995), plus the $^1$H$^\beta$ and $^{13}$C$^\beta$ chemical shifts determined using the HBHA(CBCACO)NH, CBCANH, and CBCA(CO)NH spectra. Information concerning the topology of the spin system was also used, as well as the expected $^{13}$C chemical shift ranges (Howarth & Lilley, 1978). First, delineation of the complete spin system was attempted with the HC(C)H-TOCSY spectrum ($^1$H and $^{13}$C chemical shifts). Then, this information was used to find the same spin system in the H(C)CH-COSY spectrum. In this experiment, magnetization is transferred only between *J*-coupled protons, and this can be used, when necessary, to distinguish between $^1$H$^\beta$, $^1$H$^\gamma$, $^1$H$^\delta$, etc. In many cases, this distinction could also be made on the basis of the observed $^{13}$C shifts (Howarth & Lilley, 1978). Finally, these assignments were confirmed by analysis of the $^{15}$N-TOCSY-HSQC and the $^{15}$N-NOESY-HSQC spectra.

The aromatic spin systems of the four phenylalanine, three histidine, two tryptophan, and six tyrosine residues were first characterized in a homonuclear TOCSY spectrum. Subsequently, sequence-specific assignments were made using 2D

FIGURE 1: Stereoscopic representation showing the $C^{\alpha}$ traces of the 10 best of the 40 final structures of the disulfide form of PDI-***a***. The structures were superimposed with respect to the backbone N, $C^{\alpha}$, and C′ atoms of the ordered regions in the complete ensemble (Nilges et al., 1987). The traces are oriented in a similar fashion as in Figure 4, with the active-site cysteine residues at the lower right corner.

and 3D NOESY experiments to find connectivities to the intraresidue aliphatic protons. The different types of aromatic protons were distinguished by the observed NOE patterns. All the tyrosine and phenylalanine aromatic protons could be assigned in this way.

Cross-peaks between the $^1H^{\delta 2}$ and $^1H^{\epsilon 1}$ protons of the three histidine residues present (H8, H24, and H38) were easily identified in a TOCSY spectrum. However, only in one case (H24) could NOEs be found between the aromatic and $\beta$ protons, so sequence specific assignments could be made for the aromatic protons of only this residue.

The aromatic side-chain protons of the two tryptophan residues (W35 and W111) could be characterized not only in the homonuclear TOCSY and NOESY experiments but also in the $^{15}N$-edited experiment via the side-chain $^{15}N^{\epsilon 1}$ atom. Cross-peaks in the $^{15}N$-TOCSY-HSQC experiment between the proton attached to this nitrogen and the $H^{\delta 1}$ proton, plus NOE connectivities in the $^{15}N$-NOESY-HSQC experiments, simplified the assignments.

Assignment of the side chains of asparagine and glutamine residues was also greatly simplified by the presence of the side-chain $^{15}N$ atoms ($N^{\delta}H_2$ and $N^{\epsilon}H_2$ groups). Slices taken from the $^{15}N$-NOESY-HSQC spectra at appropriate $^{15}N^{\delta/\epsilon}/^1H$ frequencies contained NOE cross-peaks to the $H^{\beta}$ or $H^{\gamma}$ side-chain protons, which were also present in slices taken at the backbone NH frequencies of these types of residues.

Assignments of the aspartic acid and glutamic acid residues were confirmed by inspection of the COCAH spectra. This experiment gives correlations not only between the backbone carbonyl, the $C^{\alpha}$, and the $H^{\alpha}$ but also between the side-chain carbonyl, the $C^{\beta}$ or $C^{\gamma}$, and the $H^{\beta}$ or $H^{\gamma}$ of aspartic acid, glutamic acid, asparagine, and glutamine residues. A table with the resonance assignments of PDI-***a*** is available as supporting information.

*Structure Calculations.* The three-dimensional structure of the PDI-***a*** domain was calculated from the NOE restraints and 34 hydrogen bond restraints, plus 5 restraints defining the disulfide bond. The structural statistics are summarized in Table 1. A superposition of the $C^{\alpha}$ traces of the 10 best of the ensemble of 40 structures is shown in Figure 1. The conformation of the backbone is well defined throughout the polypeptide chain, except for residues 1−3, 55−57, and 117−120 (Table 1). The best defined regions of the structure comprise the central $\beta$-sheet, which is apparent from the

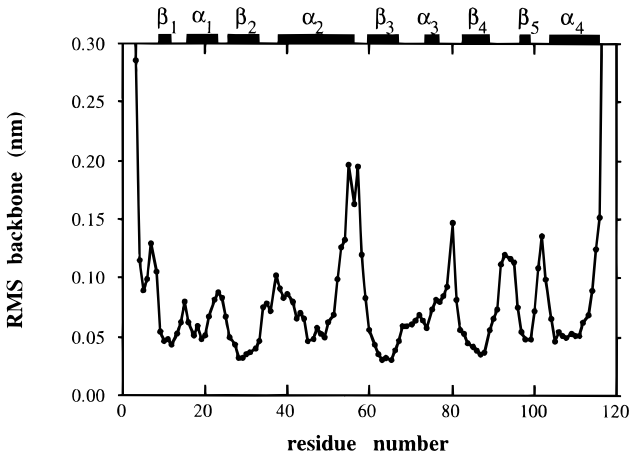


FIGURE 2: Differences in positions of the backbone N, $C^{\alpha}$, and C′ atoms for the ensemble of 40 structures of PDI-***a***. The positions of the α-helices and $\beta$-strands are indicated.

backbone rms deviations shown in Figure 2. Other well-defined regions comprise the four helices. The well-ordered portion of the backbone of the ensemble of 40 structures was determined using a procedure described by Nilges et al. (1987). It comprised 93% of the backbone N, $C^{\alpha}$, and C′ atoms, with an overall rms deviation from the mean structure of 0.08 nm.

The N- and C-termini of helix $\alpha_2$ are rather poorly defined. The disorder at the N-terminus is the result of an absence of information about the amide protons of residues Trp35, Gly37, and His38. The disorder at the C-terminus (residues 52−56) appears to be caused at least partly by spectral overlap hampering the interpretation of NOE restraints. Other ill-defined regions can be found, as usual, in loops on the surface connecting the major elements of secondary structure.

Figure 3 shows the Ramachandran plot for the 4800 backbone ($\phi$,$\psi$) pairs of the 40 structures in the ensemble. Apart from a very few outliers, the ($\phi$,$\psi$) coordinates of residues present in the well-ordered portions of the structures cluster in the allowed regions. The less-ordered residues also occupy forbidden regions and the left-handed α-helical region, which is not unusual for residues present in turn conformations.

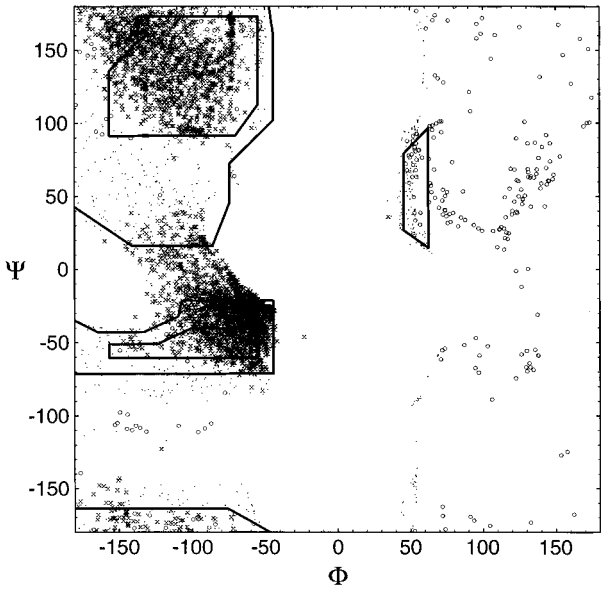


FIGURE 3: Ramachandran plot of the backbone ($\phi$,$\psi$) angles of the best 40 structures of PDI-***a*** generated. The ($\phi$,$\psi$) angles of residues in well-ordered regions of the structures are marked with crosses (×), while those in the less-ordered regions are marked with dots (·). The ($\phi$,$\psi$) angles of glycine residues are marked with circles (○). The regions of the Ramachandran plot normally allowed for alanine residues are indicated.

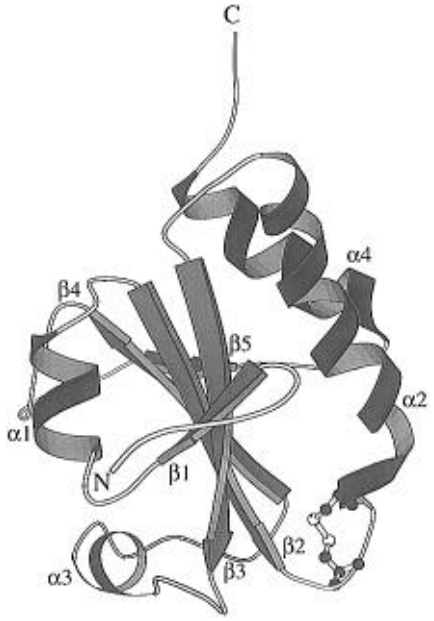


FIGURE 4: Molscript representation (Kraulis, 1991) of one of the 40 final structures of PDI-***a***. The $\beta$-strands are in red and the α-helices in blue. The individual atoms of the active-site cysteine residues (Cys36 and Cys39) are situated at the N-terminus of helix $\alpha_2$ and are indicated using a ball-and-stick representation.

## DISCUSSION

Residues 1−120 comprise only 24% of the 491 residues of the total polypeptide chain of mature human PDI, yet they are sufficient to adopt a compact folded structure (Figure 4) that is stable, soluble, and functional. This structure could be determined using NMR methods appropriate for fully folded proteins. It has all the characteristics of a typical folded, globular protein, with charged residues scattered over its entire surface. Only a few residues at each end of the chain appeared not to exist in relatively fixed conformations. The remainder of the residues comprise a folded structure that unfolds reversibly and cooperatively and has a stability comparable to that of the intact PDI molecule (Darby et al., 1996). Moreover, this protein fragment can account for some of the catalytic activities of the intact protein (Darby & Creighton, 1995a). By these very stringent criteria, PDI has a modular, multidomain structure, in which the ***a*** domain of residues 4−116 seems to be structurally independent of the remainder of the molecule. In agreement, there are no large hydrophobic patches on the surface of the PDI-***a*** domain, other than at its active site, of the type that would be expected if this domain interacted structurally with other parts of the PDI molecule.

This modular structure was predicted from the primary structure of PDI, which can be divided into at least four putative domains (Edman et al., 1985). The ***a*** domain is clearly homologous to thioredoxin and to a segment designated ***a′*** near the C-terminus; they are separated by two segments, designated ***b*** and ***b′***, that are homologous to each other. An ***a′*** fragment consisting of only residues 348−462 can also adopt a folded structure that is soluble and functional (Darby & Creighton, 1995a). The stability of its folded structure is, however, less than that of the intact protein, suggesting that it interacts structurally with other residues of intact PDI. The ***b*** and ***b′*** segments also appear to comprise individual domains. The precise residues of the human PDI primary structure comprising each of the ***b***, ***b′***, and ***a′*** domains are being determined experimentally (Darby et al., 1996).

It is unlikely that PDI contains any other folded structural domains, as the primary structure segments unaccounted for seem too short to encode independent folded domains. It had been suggested, on the basis of limited amino acid similarities to the estrogen receptor (Tsibris et al., 1989), that PDI also contains a further domain, designated ***e***, between the ***a*** and ***b*** domains and comprising residues 101−145 (Freedman et al., 1994). However, the first half of this segment is part of the fixed structure of the PDI-***a*** domain determined here and includes the C-terminal α-helix. The second half of the putative ***e*** domain appears to be an integral part of the ***b*** domain (Darby et al., 1996), making it very unlikely that the two halves could also interact to produce an autonomous folded structure. In confirmation, the polypeptide fragment of residues 100−157 corresponding to this domain was produced and found to be unfolded (Darby et al., 1996). The sequence similarity between PDI and the estrogen receptor is no greater than occurs between PDI and a significant number of other apparently unrelated proteins (Kemmink et al., 1995). Also, a second segment with comparable sequence similarity to the estrogen receptor is embedded within the ***b*** domain of PDI, and the two segments occur in the opposite order in the two proteins (Tsibris et al., 1989).

The homology of the amino acid sequences of the ***a*** and ***a′*** domains to thioredoxin (Edman et al., 1985) made it likely that they would adopt the TRX-like structural motif (Holmgren et al., 1975; Katti et al., 1990), and this has been confirmed here for PDI-***a*** (Figures 5 and 6). The structure of PDI-***a*** is like that of *E. coli* TRX in consisting of a central core made up of a five-stranded $\beta$-sheet surrounded by four α-helices. Similarly, it can be divided into two subdomains: ($\beta\alpha$)$\beta\alpha\beta$ at the N-terminus, residues 9−67, and $\beta\beta\alpha$ at the C-terminus, residues 83−116; the initial ($\beta\alpha$) is not considered part of

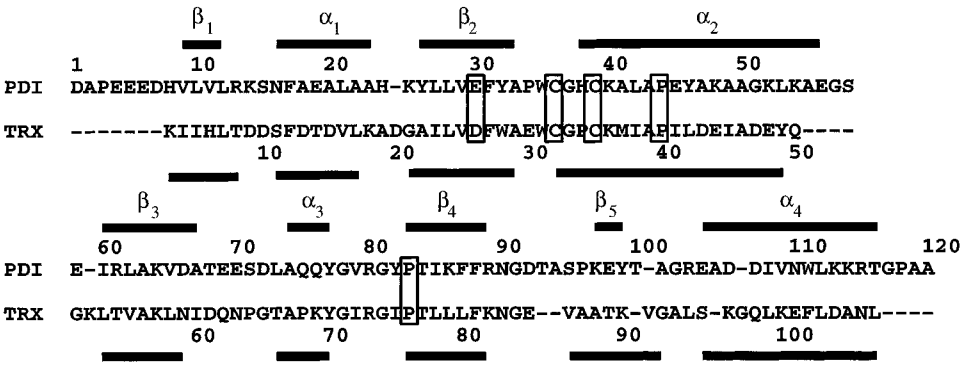


FIGURE 5: Sequence alignment on the basis of the three-dimensional structures of PDI-***a*** and *E. coli* TRX (Katti et al., 1990) using the program DALI (Holm & Sander, 1993). Specific residues addressed in the text are boxed. The elements of secondary structure are indicated. The limits of the $\alpha_2$-helix are not well defined by the data. The region indicated as $\alpha_3$ is a $3_{10}$ helix in the X-ray structure of *E. coli* TRX (Katti et al., 1990). The α-helix preceding the $\alpha_3$ $3_{10}$ helix, which is observed only in the X-ray structure of *E. coli* TRX (Katti et al., 1990; Dyson et al., 1989), was not observed in PDI-***a*** and is not indicated.

FIGURE 6: Stereoscopic representation of the $C^{\alpha}$ traces of PDI-***a*** (red) and *E. coli* TRX (blue; Katti et al., 1990) superimposed with respect to the $C^{\alpha}$ atoms on the basis of the sequence alignment shown in Figure 5.

the general thioredoxin motif, as it is not found in other members of this family (Martin, 1995). In contrast to proteins like DsbA and glutathione peroxidase, the thioredoxin motif has not been interrupted in PDI-***a*** by any substantial insertions. The longest insertion or deletion between PDI-***a*** and *E. coli* TRX is four residues (Figure 5).

The alignment of the sequences of PDI-***a*** and *E. coli* TRX based on their three-dimensional structures is shown in Figure 5. The amino acid identity is only 27% in this alignment. It differs slightly from alignments produced solely from the sequences by merely minimizing mutational differences between them, which tend to place Pro96 of PDI-*a* in the middle of strand $\beta_5$ primarily to match two alanine and two lysine residues in the sequences of PDI-***a*** and *E. coli* TRX. This is implausible structurally, and Pro 96 occurs at the start of strand $\beta_5$.

The PDI-***a*** domain shares other notable features with *E. coli* TRX: (1) The two active-site cysteine residues are in the same position at the N-terminus of helix $\alpha_2$. (2) Helix $\alpha_2$ is distorted by a proline residue at the same position, Pro44 in PDI-***a***. (3) Pro83 at the N-terminus of strand $\beta_4$ is preceded by a *cis* peptide bond. (4) The buried Asp26 of *E. coli* TRX, which has an elevated p$K_a$ value and has been postulated to play a role in the redox properties of the protein (Langsetmo et al., 1991), is Glu30 in PDI-***a*** and is also largely buried. (5) Asp26 is close to Lys57 in *E. coli* TRX, and the corresponding latter residue of PDI-***a*** is Lys64. All of these residues of PDI-***a*** are also conserved in other PDI molecules, making it likely that they have similar functional or structural roles.

The greatest difference between the structures of PDI-***a*** and *E. coli* TRX is that strand $\beta_5$ comprises only three residues in the former but seven in the latter. One turn of $3_{10}$ helix connecting the two subdomains of *E. coli* TRX, residues 66−70, is a regular α-helix in PDI-***a***, residues 74−77. This difference is probably caused by the replacement of Pro68 in TRX by Glu75 in PDI-***a***. A very similar situation is found in *Anabaena* TRX-2, where the corresponding residue is Lys (Saarinen et al., 1995).

The PDI-***a*** domain is also related functionally to thioredoxins, in that both function by transferring a disulfide bond to and from other protein molecules. Both PDI and thioredoxins react more rapidly with protein substrates than with comparable small thiol and disulfide molecules (Holmgren, 1995; Darby & Creighton, 1995a,b). The most likely explanation is that they have noncovalent and nonspecific binding interactions with protein substrates (Darby & Creighton, 1995c). The rate enhancements are only 2−3 orders of magnitude, which would require only weak interactions between the reactants, probably too weak to be detected in the absence of a disulfide bond linking them. A nonspecific hydrophobic patch involved in binding an unstructured peptide has been detected experimentally on human TRX (Qin et al., 1995), and the same region is involved in crystal contacts involving *Anabaena* TRX-2 (Saarinen et al., 1995). A very similar hydrophobic patch

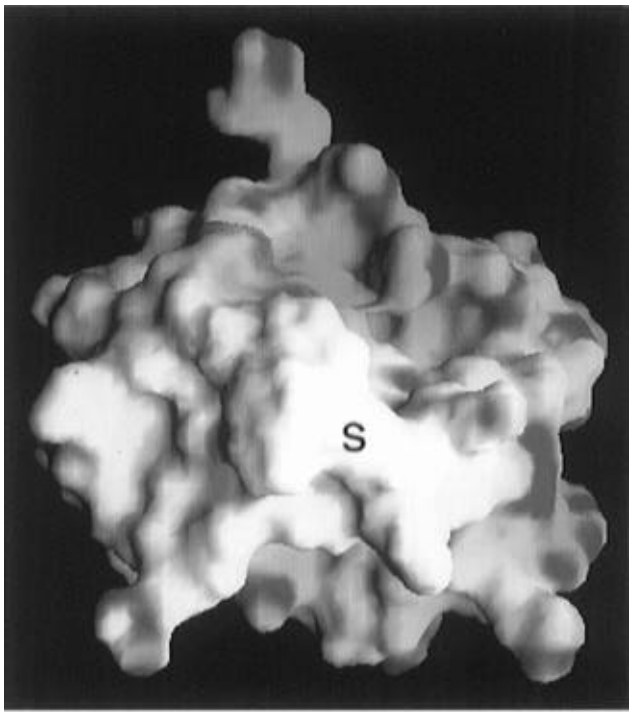


FIGURE 7: Molecular surface and charge distribution of PDI-***a***, generated using the program GRASP (Nicholls, 1993). The surface is colored according to the charge (white represents uncharged, red represents negatively charged, and blue represents positively charged surface). The position of the sulfur atom of the exposed cysteine (Cys36) is indicated by S.

occurs on PDI-***a*** (Figure 7). It is therefore likely that PDI-***a*** interacts noncovalently with unfolded proteins in a similar fashion, although PDI may also have other peptide binding sites on its surface (Noiva et al., 1993; Darby & Creighton, 1995a).

Both thioredoxins and PDI function by cycling between the dithiol and disulfide forms of the two cysteine residues at their homologous active sites, using the single accessible cysteine residue (Cys36 of PDI-***a***) to react with the cysteine residues of other proteins. PDI and TRX differ dramatically in their redox functions, however: PDI-***a*** is a potent oxidant, while thioredoxins are reductants (Holmgren, 1985, 1995). The basis for this difference is that the disulfide bond of PDI-***a*** is 5.7 kcal/mol less stable than that of *E. coli* TRX (Darby & Creighton, 1995c); one stabilizes the TRX fold and the other destabilizes it. In spite of this difference, the two disulfide bonds are structurally very similar to each other and to the extremely unstable disulfide bond of DsbA (Martin et al., 1993). This difference is now believed to arise not from differences in the disulfide bonds themselves but in differences in stabilization of the dithiol form, by electrostatic stabilization of the thiolate anion of the accessible cysteine residue (Nelson & Creighton 1994; Grauschopf et al., 1995). In that regard, it is probably significant that the backbone amide protons of residues at the active site, in particular, Trp35, Gly37, and His38, were largely absent from the NMR spectra of PDI-***a***. This is believed to arise from electrostatic interactions at the active site (Gane et al., 1995; Kortemme & Creighton, 1995) that cause these amide groups to exchange with solvent much more rapidly than is normal (Kemmink et al., 1995). Related but different effects have been observed with *E. coli* TRX (Kaminsky & Richards, 1992; Jeng & Dyson, 1995). Such effects are probably important for the thiol−disulfide chemistry at the active sites of these proteins and possibly also for the remarkable differences in stability of their disulfide bonds. The latter difference is probably not determined by the many aspects, described earlier, that the two proteins have in common. The structure of the PDI-***a*** domain should be valuable in elucidating this important functional property of these proteins.

## ACKNOWLEDGMENT

We thank K. Kivirikko for generously supplying the cloned gene for human PDI, R. M. Scheek for support and encouragement, and Elke Penka for excellent technical assistance.

## SUPPORTING INFORMATION AVAILABLE

A table of the $^{1}H$, $^{13}C$, and $^{15}N$ assignments of PDI-***a*** (8 pages). Ordering information is given on any current masthead page.

## REFERENCES

Bax, A., & Subramanian, S. (1986) *J. Magn. Reson. 67*, 565−569.
Bax, A., & Pochapsky, S. S. (1992) *J. Magn. Reson. 99*, 638−643.
Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., Ikura, M., & Kay, L. E. (1990) *J. Magn. Reson. 87*, 620−627.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
Brünger, A. (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., & Rance, M. (1991) *J. Magn. Reson. 91*, 429−436.
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., Kemmink, J., & Creighton, T. E. (1996) *Biochemistry* (submitted for publication).
Dyson, H. J., Holmgren, A., & Wright, P. E. (1989) *Biochemistry 28*, 7074−7087.
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, Freeman, New York.
Freedman, R. B., Hirst, T. R., & Tuite, M. F. (1994) *Trends Biochem. Sci. 19*, 331−336.
Gane, P. J., Freedman, R. B., & Warwicker, J. (1995) *J. Mol. Biol. 249*, 376−387.
Grauschopf, U., Winther, J. R., Korber, P., Zander, T., Dallinger, P., & Bardwell, J. C. A. (1995) *Cell 83*, 947−955.
Grzesiek, S., & Bax, A. (1992a) *J. Magn. Reson. 99*, 201−207.
Grzesiek, S., & Bax, A. (1992b) *J. Am. Chem. Soc. 114*, 6291−6293.
Grzesiek, S., & Bax, A. (1993) *J. Biomol. NMR 3*, 185−204.
Holm, L., & Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
Holmgren, A. (1985) *Annu. Rev. Biochem. 54*, 237−271.
Holmgren, A. (1995) *Structure 3*, 239−243.
Holmgren, A., Söderberg, B.-O., Eklund, H., & Brändén, C.-I. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2305−2309.
Howarth, O. W., & Lilley, D. M. J. (1978) *Prog. NMR Spectrosc. 12*, 1−40.
Jeng, M.-F., & Dyson, H. J. (1995) *Biochemistry 34*, 611−619.
Kallis, G.-B., & Holmgren, A. (1980) *J. Biol. Chem. 255*, 10261−10265.
Kaminsky, S. M., & Richards, F. M. (1992) *Protein Sci. 1*, 10−21.

Katti, S. K., LeMaster, D. M., & Eklund, H. (1990) *J. Mol. Biol. 212*, 167−184.
Kay, L. E., Keifer, P., & Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., & Forman-Kay, J. D. (1993) *J. Magn. Reson. B 101*, 333−337.
Kemmink, J., Darby, N. J., Dijkstra, K., Scheek, R. M., & Creighton, T. E. (1995) *Protein Sci. 4*, 2587−2593.
Kharrat, A., Macias, M. J., Gibson, T., Nilges, M., & Pastore, A. (1995) *EMBO J 14*, 3572−3584.
Kortemme, T., & Creighton, T. E. (1995) *J. Mol. Biol. 253*, 799−812.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Langsetmo, K., Fuchs, J. A., & Woodward, C. (1991) *Biochemistry 30*, 7603−7609.
Live, D. H., Davis, D. G., Agosta, W. C., & Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1939−1941.
Lyles, M. M., & Gilbert, H. F. (1991) *Biochemistry 30*, 619−625.
Lyles ,M. M., & Gilbert, H. F. (1994) *J. Biol. Chem. 269*, 30946−30952.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Martin, J. L. (1995) *Structure 3*, 245−250.
Martin, J. L., Bardwell, J. C. A., & Kuriyan, J. (1993) *Nature 365*, 464−468.
Muhandiram, D. R., Farrow, N. A., Xu, G.-Y., Smallcombe, S. H., & Kay, L. E. (1993) *J. Magn. Reson. B 102*, 317−321.
Nelson, J. W., & Creighton, T. E. (1994) *Biochemistry 33*, 5974−5983.
Nicholls, A. J. (1993) *GRASP Manual*, Columbia University, New York.
Nilges, M. (1995) *J. Mol. Biol. 245*, 645−660.
Nilges, M., Gronenborn, A. M., & Clore, G. M. (1987) *FEBS Lett. 219*, 17−21.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.
Noiva, R., Freedman, R. B., & Lennarz, W. J. (1993) *J. Biol. Chem. 268*, 19210−19217.
Palmer, A. G., Cavanagh, J., Wright, P. E., & Rance, M. (1991) *J. Magn. Reson. 93*, 151−170.
Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllylä, R., Huhtala, M.-L., Koivu, J., & Kivirikko, K. I. (1987) *EMBO J. 6*, 643−649.
Qin, J., Clore, G. M., Kennedy, W. M. P., Huth, J. R., & Gronenborn, A. M. (1995) *Structure 3*, 289−297.
Saarinen, M., Gleason, F. K., & Eklund, H. (1995) *Structure 3*, 1097−1108.
Shaka, A. J., Keeler, J., & Freeman, R. (1983) *J. Magn. Reson. 53*, 313−340.
Shaka, A. J., Barker, P. B., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547−552.
Shaka, A. J., Lee, C. J., & Pines, A. (1988) *J. Magn. Reson. 77*, 274−293.
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., Aggerbeck, L. P., Laplaud, P. M., & McLean, L. R. (1991) *Biochemistry 30*, 4406−4412.

BI960335M