Determinants of Protein Hyperthermostability: Purification and Amino Acid Sequence of Rubredoxin from the Hyperthermophilic Archaebacterium *Pyrococcus furiosus* and Secondary Structure of the Zinc Adduct by NMR†

Paul R. Blake,‡ Jae-Bum Park,§ Frank O. Bryant,§ Shigetoshi Aono,§ Jon K. Magnuson,∥ Eric Eccleston,∥ James B. Howard,∥ Michael F. Summers,*,‡ and Michael W. W. Adams*,§

Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, Maryland 21228, Department of Biochemistry and Center for Metalloenzyme Studies, University of Georgia, Athens, Georgia 30602, and Department of Biochemistry, Medical School, University of Minnesota, Minneapolis, Minnesota 55455

Received June 3, 1991; Revised Manuscript Received August 9, 1991

ABSTRACT: The purification, amino acid sequence, and two-dimensional ^{1}H NMR results are reported for the rubredoxin (Rd) from the hyperthermophilic archaebacterium *Pyrococcus furiosus*, an organism that grows optimally at 100 °C. The molecular mass (5397 Da), iron content (1.2 ± 0.2 g-atom of Fe/mol), UV–vis spectrophotometric properties, and amino acid sequence (60% sequence identity with *Clostridium pasteurianum* Rd) are found to be typical of this class of redox protein. However, *P. furiosus* Rd is remarkably thermostable, being unaffected after incubation for 24 h at 95 °C. One- and two-dimensional ^{1}H nuclear magnetic resonance spectra of the oxidized [Fe(III)Rd] and reduced [Fe(II)Rd] forms of *P. furiosus* Rd exhibited substantial paramagnetic line broadening, and this precluded detailed 3D structural studies. The apoprotein was not readily amenable to NMR studies due to apparent protein oxidation involving the free cysteine sulfhydryls. However, high-quality NMR spectra were obtained for the Zn-substituted protein, Zn(Rd), enabling detailed NMR signal assignment for all backbone amide and α and most side-chain protons. Secondary structural elements were determined from qualitative analysis of 2D Overhauser effect spectra. Residues A1–K6, Y10–E14, and F48–E51 form a three-strand antiparallel β-sheet, which comprises ca. 30% of the primary sequence. Residues C5–Y10 and C38–A43 form types I and II amide–sulfur tight turns common to iron–sulfur proteins. These structural elements are similar to those observed by X-ray crystallography for native Rd from the mesophile *C. pasteurianum*. However, the β-sheet domain in *P. furiosus* Rd is larger than that in *C. pasteurianum* Rd and appears to begin at the N-terminal residue. From analysis of the secondary structure, potentially stabilizing electrostatic interactions involving the charged groups of residues Ala(1), Glu(14), and Glu(52) are proposed. These interactions, which are not present in rubredoxins from mesophilic organisms, may prevent the β-sheet from "unzipping" at elevated temperatures.

In the past few years, several genera of novel hyperthermophilic bacteria have been isolated from sulfur-rich marine geothermal environments (Stetter, 1986; Stetter et al., 1990; Adams, 1990). These unusual organisms exhibit optimal growth at a remarkable 100 °C or greater. Although they have significant biotechnological potential as sources of a range of "hyperthermostable" enzymes, our present knowledge of these organisms is limited mainly to papers describing their isolation, and very little physiological and/or biochemical information is available. The one exception is *Pyrococcus furiosus*, which was isolated by Fiala and Stetter (1986). This bacterium is a strict anaerobe that metabolizes carbohydrates by a fermentative-type pathway in which H_2, CO_2, and organic acids are the products. *P. furiosus* also reduces S^0 to H_2S in a non-energy-yielding reaction believed to provide a means of removing H_2, since H_2 inhibits growth. To date, hydrogenase (Bryant & Adams, 1989), ferredoxin (Aono et al., 1989), aldehyde ferredoxin oxidoreductase (Mukund & Adams, 1990, 1991), a protease (Costantino et al., 1990), and α-glucosidase (Blumentals et al., 1990) have been purified from *P. furiosus*. All are extremely thermostable proteins with optimal temperatures for activity above 95 °C. However, an understanding at the molecular level of the factors that enable proteins to function under extreme conditions of temperature is primitive or lacking.

In order to elucidate the novel biochemistry that must be required for growth at extreme temperatures and, in particular, the redox chemistry of important electron transport metalloproteins, we have recently purified and initiated biophysical studies of the rubredoxin (Rd) from *P. furiosus*. As the simplest known class of redox proteins (Yasunobu & Tanaka, 1973; Lovenberg & Sobel, 1965), the study of rubredoxins isolated from divergent anaerobic species has the potential to provide insights regarding the structural requirements for this protein and the residues that influence the electron transfer and redox properties of the iron site. However, to date, the only structural information available on the low molecular weight rubredoxins, which now includes 11 amino acid sequences (Bachmeyer et al., 1968a,b; Watenpaugh et al., 1973; Bruschi, 1976a,b; Hormel et al., 1986; Woolley & Meyer, 1987; Voordouw, 1988; Saeki et al., 1989; Seki et al., 1989;

† Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society for support of this work. This work was also supported by the National Institutes of Health (Grant GM 3431 to J.B.H.), by a National Science Foundation Research Training Group Award to the Center for Metalloenzyme Studies of the University of Georgia (DIR 9014281), and by grants to M.W.W.A. from the Office of Naval Research (N00014-90-J-1894), the Department of Energy (FG09-88ER-13901), and the National Science Foundation (BCS-9011583). The 600-MHz NMR instrument was purchased with support from the NIH (Grants GM42561 and AI30917 to M.F.S.).

* To whom correspondence should be addressed.
‡ University of Maryland Baltimore County.
§ University of Georgia.
∥ University of Minnesota.

0006-2960/91/0430-10885$02.50/0

Shimizu et al., 1989; Meyer et al., 1990) and four crystal structures (Adman et al., 1977; Watenpaugh et al., 1979; Sieker et al., 1986; Frey et al., 1987), comes from those purified from eubacteria. Moreover, all but one of the rubredoxins that have been characterized so far are from mesophilic species, the exception being *Clostridium thermosaccharolyticum* (Meyer et al., 1990), which grows optimally at 55 °C.

Nuclear magnetic resonance (NMR) spectroscopy provides a potentially powerful approach for studying the temperature-dependent structural and dynamic properties of low-molecular mass proteins (< ca. 20 kDa). However, the method fails when nuclear magnetic relaxation rates becomes rapid relative to the internal time parameters of the essential NMR experiments. This problem is exacerbated in paramagnetic metalloproteins, where signals for protons close to the paramagnetic metals are sometimes difficult to detect due to severe line broadening (Krishnamoorthi et al., 1986; Werth et al., 1987). 2D NMR methods have been successfully applied to some paramagnetic metalloproteins (e.g., heme proteins such as metcyanomyoglobin with T_1 and half-height widths for proton signals of ca. 20 s^{-1} and 100 Hz, respectively) (Yu et al., 1990); however, protons near the metal centers of iron–sulfur proteins exhibit extreme broadening, with half-height widths as high as 4500 Hz (Werth et al., 1987), making high-resolution structure determination impossible using conventional NMR methods. Thus, no cross peaks were observed in 2D homonuclear NMR spectra of the ferrodoxin (Fd) from *Anabaena* for protons within ca. 7.8 Å of the paramagnetic iron atoms (Oh & Markley, 1990).

In this report, we describe the purification, amino acid sequence, and one- and two-dimensional (2D) nuclear magnetic resonance studies of *P. furiosus* Rd. As expected for the native protein, rapid nuclear magnetic relaxation resulted in extensive paramagnetic bleaching of critical 2D NMR spectra, and this precluded detailed structural studies. In addition, the apoprotein was not readily amenable to NMR studies due to apparent oxidation of the free cysteine sulfhydryls. To overcome these problems, the zinc-substituted protein, Zn(Rd), was prepared. Unlike the native protein, Zn(Rd) gave rise to 2D spectra rich with cross peaks, enabling complete assignment of the backbone amide and α and most side-chain proton signals and facilitating the amino acid sequence determination. Secondary structural elements were determined from qualitative analysis of 2D Overhauser effect data. Our findings indicate that *P. furiosus* Rd adopts a backbone conformation similar to that of Rd from the mesophile *Clostridium pasteurianum* (60% sequence identity). In addition, novel structural elements in *P. furiosus* Rd were identified that may represent determinants of protein hyperthermostability. This represents the first structural data obtained for a protein from a hyperthermophile and demonstrates the utility of Zn substitution for NMR-based 3D structural studies of paramagnetic metalloproteins.

Experimental Procedures

Protein Isolation and Purification. P. furiosus (DSM 3638) was grown in a 400-L fermentor as described previously (Bryant & Adams, 1989). The rubredoxin was routinely purified from 400 g of cells (wet weight). All procedures were carried out at 23 °C under strictly anaerobic conditions. All buffers were repeatedly degassed and flushed with Ar and maintained under a positive pressure of Ar. The rubredoxin was measured by its absorbance at 490 nm. The purification procedure was the same as for *P. furiosus* hydrogenase, up to and including the first Q-Sepharose column (Bryant & Adams, 1989). The rubredoxin was detected in the Q-Sepharose fractions at 0.30–0.34 M sodium chloride (pH 8.2), just prior to the hydrogenase. All subsequent columns were equilibrated with and eluted with 50 mM Tris-HCl buffer, pH 8.0, and were controlled by a Pharmacia FPLC system. The combined fractions from the Q-Sepharose column (1575 mg, 450 mL, A_{490}/A_{280} = 0.15) were diluted to 1 L with Tris-HCl buffer and applied to a column (5 × 20 cm) of DEAE Sepharose Fast Flow (Pharmacia LKB). An NaCl gradient of 0–150 mM (350 mL) and 150–400 mM (6300 mL) was applied to the column at 7 mL/min. The rubredoxin eluted at 238–247 mM NaCl. The pooled fractions (450 mg, 300 mL, A_{490}/A_{280} = 0.20) were loaded directly on to a column (2.5 × 25 cm) of hydroxyapatite (High Resolution, Behring Diagnostics). This was eluted with a 2100-mL gradient from 0 to 500 mM potassium phosphate (pH 8.0) at 3 mL/min. The rubredoxin eluted as 7–11 mM phosphate was applied. The combined fractions (60 mg, 100 mL, A_{490}/A_{280} = 0.33) were concentrated to approximately 5 mL using a Q-Sepharose column (1.0 × 5 cm) and were applied to a column (2.5 × 100 cm) of Sephacryl S-200 (Pharmacia LKB). This was equilibrated with buffer containing 0.2 M NaCl and was run at 1 mL/min. The rubredoxin-containing fractions that were judged pure by SDS–polyacrylamide electrophoresis were concentrated to 5 mL by Q-Sepharose chromatography (51 mg, A_{490}/A_{280} = 0.36) and were stored as pellets in liquid N_2.

Protein Characterization. Protein concentrations were routinely estimated by a colorimetric assay using bovine serum albumin as the standard (Lowry et al., 1951). The protein content of pure rubredoxin was also determined by the quantitative recovery of amino acids from compositional analyses (see below). The amounts of protein were 51 ± 5% of those measured by the Lowry colorimetric method. All analytical values for the pure rubredoxin that were based on the colorimetric assay have therefore been corrected by a factor of 1.96. Electrophoresis using polyacrylamide (18% w/v) in the presence of SDS was performed as described previously (Weber et al., 1972). Iron and acid-labile sulfide content was measured using *o*-phenanthroline (Lovenberg et al., 1963) and by methylene blue formation (Chen & Mortenson, 1977), respectively. The molecular weight of the rubredoxin was determined by gel filtration using two connected columns (HR 10/30) of Superose 12 operated by a Pharmacia FPLC system. This was calibrated with aldolase (158 000), bovine serume albumin (67 000), ovalbumin (45 000), carbonic anhydrase (29 000), myoglobin (16 900), and cytochrome *c* (12 300) using 50 mM Tris-HCl buffer, pH 8.0, containing either 0.2 or 1.0 M NaCl as the eluent.

Amino Acid Sequencing. The iron was removed from 0.4 mg of *P. furiosus* Rd by dissolving the protein in 100 μL of 88% formic acid. After 30 min at 25 °C, the protein was lyophilized. Residual acid was removed by dissolving the protein in water and lyophilizing a second time. The apoprotein was dissolved in a solution containing 6 M guanidine·HCl, 1 M Tris-HCl (pH 8.0), 10 mM EDTA, and 10 mM dithiothreitol. The solution was incubated at 80 °C for 4 h, and the free thiols modified by reaction with 20 mM iodoacetic acid for 30 min (Lundell & Howard, 1978). The protein solution was dialyzed exhaustively against 10% acetic acid, followed by lyophilization.

The S-carboxymethylated apo-Rd was digested with *N*-tosylphenylalanine chloromethyl ketone treated trypsin (2% w/w) (Worthington) for 2 h at 30 °C in 25 mM Tris-HCl, pH 8.0. The tryptic peptides were separated by reverse-phase chromatography on a 4.6 × 250 mm Vydac C-4 column using a gradient of 0.1% trifluoroacetic acid in water and 0.1%

trifluoroacetic acid in acetonitrile. The acetonitrile gradient was as follows: 0–20% in 5 min, 20–35% in 15 min, and 35–65% in 30 min followed by 100%. All peptides were eluted at acetonitrile concentrations in the range of 25–40%.

Amino acid sequences of the peptides and S-carboxymethylated apoprotein were determined by repetitive Edman degradation with an ABI 470 A gas-phase sequencer with on-line phenylthiohydantoin amino acid analysis. S-carboxymethylated rubredoxin was hydrolyzed in 6 N HCl, containing 1% (w/v) phenol, in vacuo, at 110 °C for 24, 48, and 72 h. Amino acid compositions of the resultant hydrolyzates were determined using a Beckman 6300 automated amino acid analyzer.

Zinc Substitution. Zinc was substituted for the iron atom of *P. furiosus* Rd as follows: Native protein (4.3 mM) was precipitated by the addition of trichloroacetic acid (TCA, 25% v/v) at 0 °C to a final concentration of 10% TCA. The white precipitate was washed twice with TCA (10% v/v), followed by dissolution into 2 mL of 6 M guanidine hydrochloride (gua·HCl) containing β-mercaptoethanol (β-metoh, 15 mM).

The sample was then dialyzed (Spectra-por 6; 2000 MW cut-off) against the following solutions: (A) 4 M gua·HCl, 10 mM β-metoh, 26 mM $ZnCl_2$, 250 mM NaCl, 200 mM acetate buffer, pH 6.1; 140-mL total volume; 4-h dialysis time. (B) 2 M gua·HCl, 2 mM β-metoh, 26 mM $ZnCl_2$, 200 mM acetate buffer, pH 6.2; 140-mL total volume; 4-h dialysis time. (C) 1 mM β-metoh, 1.3 mM $ZnCl_2$, 250 mM NaCl, 25 mM acetate buffer, pH 6.2; 250-mL total volume; 9-h dialysis time. (D) 0.1 mM $ZnCl_2$, 140 mM NaCl, 25 mM deuterated acetate buffer (acetate-d_3), pH* 6.3. At this pH, the ionic charge of the protein (estimated from the amino acid sequence) is −10. After the sample volume was reduced to 450 μL by centrifugation through a 3000 MW cut-off membrane (Centricon-3), 50 μL of D_2O (containing 200 mM NaCl) was added for NMR sample lock.

The zinc and iron contents were determined by plasma emission spectrophotometry (Jarrel Ash Plasma Co. instrument) for native and zinc-substituted Rd samples [treated by gel filtration (G-25)]: Zn(Rd), 0.97 and 0.002 g-atom/mol, respectively; Fe(Rd), 0.009 and 1.01 g-atom/mol, respectively.

NMR Data Collection and Processing. ^1H NMR data were collected at 45 °C with a GE GN-500 NMR spectrometer (500.1 MHz) and at 25 °C with a GE OMEGA-PSG 600-MHz spectrometer (599.7 MHz). Data were transferred via ethernet to Silicon Graphics computers (220 GTX-B and Personal Iris) and converted to a format suitable for processing with FTNMR (Hare Research Inc.). 2D 500-MHz data were collected with a spectral width of 7692.30 Hz and a digital resolution in t_2 of 3.756 Hz/point; 64 transients for each of 512 t_1 increments were obtained, and a second set of data was collected in an interleaved mode to achieve phase-sensitive quadrature detection in the t_1 dimension (States et al., 1982). 2D 600-MHz NMR data were collected with a spectral width of 9090.91 Hz and 4.44 Hz/point in t_2; 32 transients for each of 512 t_1 increments were obtained, with a second set of States-type data collected in a noninterleaved manner for each t_1 increment.

500-MHz Nuclear Overhauser effect (NOESY) (Jeener et al., 1979; Macura & Ernst, 1980) data were collected with a pulse sequencing employing a Dante pulse train (Hore, 1983) for saturation of the H_2O resonance, followed by a SCUBA sequence (Brown et al., 1988) to allow for dipolar relaxation involving resonances close to the water frequency. In addition, a composite 180° pulse was included in the middle of the mixing interval to invert H_2O magnetization which relaxed during the first half of the mixing time. Relevant parameters are as follows: 1.2-s Dante saturation interval, including 5° pulses separated by 100-μs intervals; 60-ms SCUBA delay period; 210-ms mixing time, which includes 5-ms homospoil (HS) pulses at the beginning of the mixing period and immediately following the composite −180° pulse. 600-MHz NOESY data were collected with the same pulse sequence and parameters, except that solvent suppression was achieved by application of a 50-Hz continuous-wave presaturation pulse.

Homonuclear Hartmann–Hahn (HOHAHA) (Braunschweiler & Ernst, 1983; Davis & Bax, 1985) and double-quantum filtered homonuclear correlated (DQF-COSY) (Pinantini et al., 1982) spectra were also collected with Dante and SCUBA preparation intervals. A $90^{\phi+1}$–HS–$90^{\phi-1}$ (ϕ = phase of trim pulse) pulse train was included in the HOHAHA sequence prior to the acquisition period to suppress the dispersive base line due to residual H_2O signal. A spin lock field of 8.7 kHz was applied for 94 ms for the collection of HOHAHA data which includes two 2-ms trim pulses. Other relevant parameters were as described above for NOESY data collection.

500-MHz NOESY and HOHAHA data were processed with 3 Hz of Gaussian line broadening in t_2, and a 60°-shifted sine squared window function in the t_1 dimension and were zero filled to a final matrix size of 2K × 2K real data points. DQF-COSY data were processed similarly except that trapezoidal apodization was employed in the t_1 dimension. The 600-MHz NOESY data were processed with a 30°- and 65°-shifted squared sine bell filters in the t_2 and t_1 dimensions, respectively, with zero-filling in both dimensions to a final matrix size of 4096 × 4096 real data points.

Results

Amino Acid Sequence. The amino acid composition of S-carboxymethylated rubredoxin is given in Table SI (Supplementary Material) and indicates that the protein contains 52 residues. The protein was subjected to repetitive Edman degradation, resulting in the identification of the first 49 residues (see Figure S1 in the Supplementary Material). Five peptides from the tryptic digestion of Rd were separated by reverse-phase chromatography (Figure S2) and were sequenced by repetitive Edman degradation (Table SII). The sequences of these tryptic peptides fully corroborate the sequence of the first 49 residues obtained for the intact protein. Fortuitously, no cleavage occured after Lys(50), and T-5 provided the overlap to the protein carboxyl terminus. However, both the Edman degradation and amino acid composition of T-5 when added to the other peptides indicated that there was one additional aspartyl residue compared to that expected from the composition of the whole protein. Although the carboxyl terminus is the most difficult residue to establish, any question about the existence or identity of the residue was resolved by the 2D NMR spectra, which clearly indicate a residue following Glu(52). As described below, the scalar connectivities and chemical shifts (Figure 2) are consistent with the terminal residue as asparatic acid. Therefore, we conclude that the protein contains 53 residues with a carboxyl terminus of aspartic acid. The complete amino acid sequence is given in Figure S1 and is compared to other rubredoxins in Figure 1.

Physical Properties. The molecular weight of the rubredoxin estimated by gel filtration was 6800 ± 1000 in the presence of 1 M NaCl and 14000 ± 2000 in the presence 0.2 M NaCl. These results are in reasonable agreement with the value (5397, see above) calculated from the amino acid sequence and indicate that at 23 °C and high ionic strength the

```
                5         10        15        20        25        30        35        40        45        50
                ↓ ↓       ↓         ↓ ↓       ↓         ↓         ↓                             ↓
Pf        A K W V C K I C G Y I Y D E D A G D P D N G I S P G T K F E E L P D D W V C P I C G A P K S E F E K L E D
Cpt     M K K Y T C T V C G Y I Y D P E D G D P D D G V N P G T D F K D I P D D W V C P L C G V G K D E F E E V E E
Cpf     M K K F I C D V C G Y I Y D P A V G D P D N G V E P G T E F K D I P D D W V C P L C G V D K S Q F S E T E E
Cts     M E K W Q C T V C G Y I Y D P E V G D P T Q N I P P G T K F E D L P D D W V C P D C G V G K D Q F E K I
Cbt     M Q K Y V C S V C G Y V Y D P A D G E P D D P I D P G T G F E D L P D E W V C P V C G V D K D L F E P E S
DvH     M K K Y V C T V C G Y E Y D P Q E G D P D N G V K P G T S F D D L P A D W V C P V C G A P K S E F E A A
FvM     M K K Y V C T V C G Y E Y D P A E G D P D N G V K P G T A F E D V P A D W V C P I C G A P K S E F E P A
Dg      M D I Y V C T V C G Y E Y D P A K G D P D S G I K P G T K F E D L P D D W A C P V C G A S K D A F E K Q
Dd      M Q K Y V C N V C G Y E Y D P A E H D - - - - - - - - - N V P F D Q L P S S W C C P V C G V S K D Q F S P A
Me      M D K Y E C S I C G Y I Y D E A E G D - - D G N V A A G T K F A D L P A D W V C P T C G A D K D A F V K M D
Pa      M Q K F E C T L C G Y I Y D P A L V G P D T P D Q D G - - A F E D V S E N W V C P L C G A G K E D F E V Y E D
Bm      M Q K Y V C D I C G Y V Y D P A V G D P D N G V A P G T A F A D L P E D W V C P E C G V S K D E F S P E A
                  ↑     ↑ ↑ ↑   ↑ ↑                       ↑               ↑   ↑ ↑   ↑ ↑   ↑         ↑
```

FIGURE 1: Comparison of rubredoxin amino acid sequences from anaerobic bacteria. The abbreviations and sources of the data are: Pf, *P. furiosus* (this work); Cpt, *Clostridium pasteurianum* (Yasunobu et al., 1973, Watenpaugh et al., 1973); Cpf: *C. perfringens* (Seki et al., 1989); Cts, *C. thermosaccharolyticum* (Meyer et al., 1990); Cbt, *Chlorobium thiosulfatophilum* (Woolley et al., 1987); DvH, *Desulfovibrio vulgaris* strain Hildenborough (Bruschi et al., 1976a, Voordouw et al., 1988); DvM, *Desulfovibrio vulgaris* strain Miyazaki (Shimizu, et al., 1989); Dg, *D. gigas* (Bruschi et al., 1976b); Dd, *D. desulfuricans* (Hormel et al., 1986); Me, *Megasphaera elsdenii* (Bachmeyer et al., 1968a); Pa, *Peptococcus aerogenes* (Bachmeyer et al., 1968b); Bm, *Butyribacterium methylotrophicum* (Saeki et al., 1989). Conserved cysteins are given in bold. ↓ indicates residues (or lack of) unique to Pf rubredoxin and ↑ indicates residues conserved in all known rubredoxins. In this figure only, the N-terminal residue of *P. furiosus* Rd is labeled Ala(2).

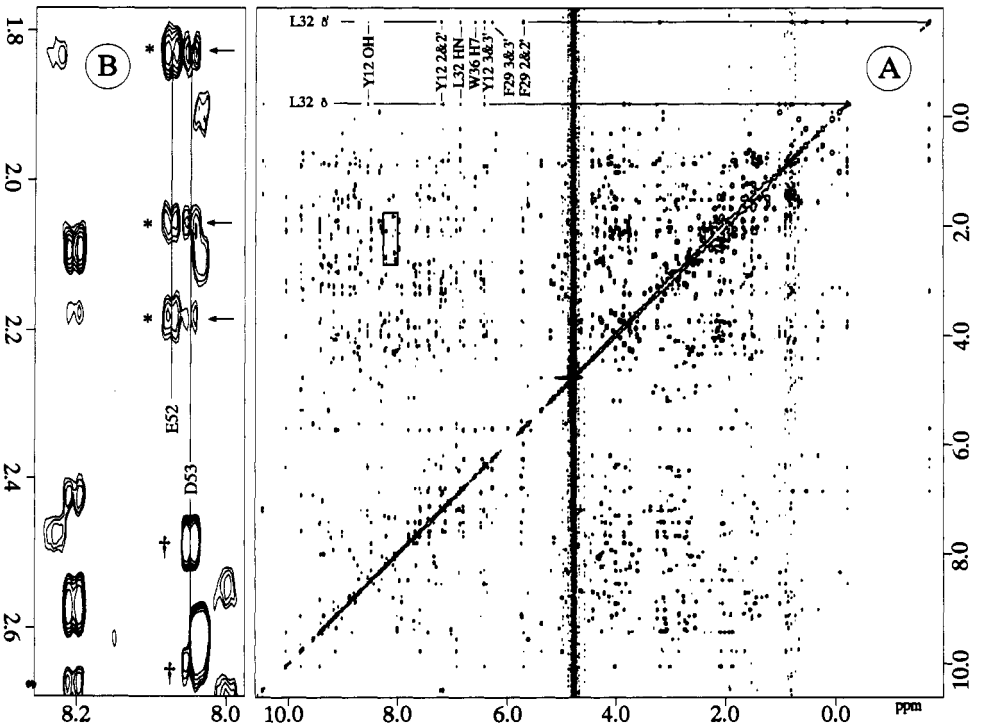

FIGURE 2: Resolution-enhanced 600-MHz NOESY data obtained for *P. furiosus* Zn(Rd) at 25 °C. (A) Full spectrum with cross peaks for the pro-*R* (δ') and pro-*S* (δ) methyl protons of Leu(32) labeled. These methyl groups constitute a portion of hydrophobic core of the protein. Similar chemical shifts and cross-peak patterns were observed for the Leu(32) methyl protons of the native protein. (B) Expansion of the boxed region of Figure 2A showing NOESY cross peaks utilized to sequence the C-terminal Asp; * and † symbols refer to intraresidue cross peaks of Glu(52) and Asp(53), respectively, as determined from a 600-MHz 2D HOHAHA spectrum. The arrows denote sequential NOE connectivities involving the Glu(53)NH and Asp(52) side-chain protons.

protein exists in a monomeric form. The purified protein gave rise to a single protein band after SDS gel electrophoresis, which migrated with the dye front even when 20% (w/v) acrylamide was used. The UV–visible spectrum of the oxidized protein (Figure S3) exhibited adsorption coefficients at the maxima of 25.6 (280 nm), 10.7 (390 nm), and 9.22 (494 nm) mM^{-1} cm^{-1}, and the A_{494}/A_{280} ratio was 0.36. The dithionite-reduced protein showed no absorption in the visible region. The protein contained 1.2 ± 0.2 g-atom of Fe/5397 g of protein, and acid-labile sulfide could not be detected. The molecular weight, Fe content, and UV–visible properties of the Rd are therefore typical for this class of redox protein. However, it was remarkably thermostable: the visible absorption properties of the protein (1 mg/mL in 50 mM EPPS buffer, pH 8.4) were unaffected after a 24 h incubation at 95 °C.

NMR Data for Native Rd. Preliminary 1D and 2D NMR data were obtained for the native, dithionite-reduced form of Rd [7 mM Fe(II)Rd, pH 7.8, 20 mM phosphate buffer, 200 mM NaCl; T = 45 °C], and a portion of the 1D ^1H NMR spectrum of Fe(II)Rd is shown at the top of Figure 3. All signals in the 1D spectrum exhibited half-height width values greater than ca. 25 Hz. As expected, severe paramagnetic bleaching of the 2D NOESY and HOHAHA spectra (Figure S4) precluded detailed signal assignments and structural studies. In fact, only 26 of the 45 expected backbone amide proton signals were observed in the 2D HOHAHA spectrum of Fe(II)Rd (Figure S4).

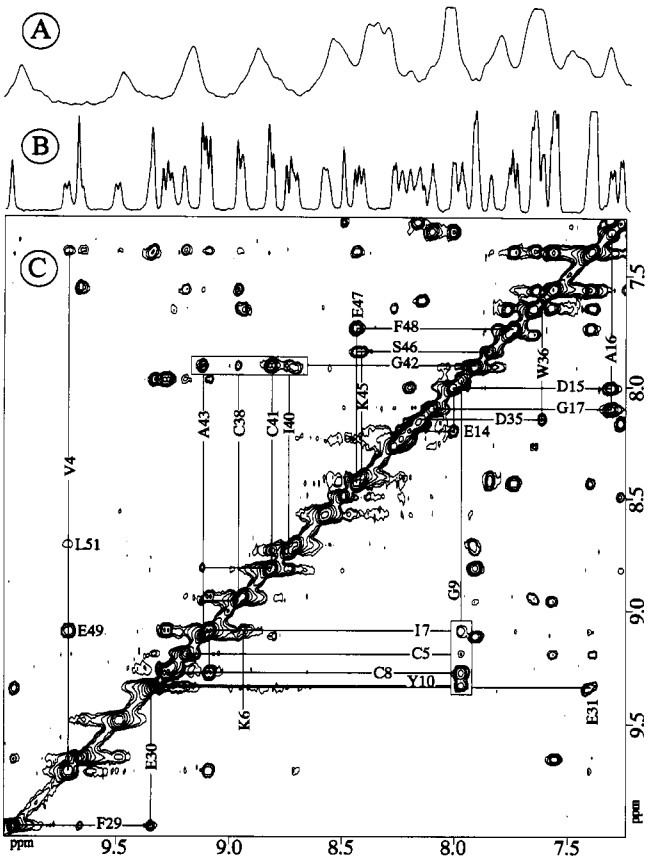

FIGURE 3: Downfield portions of the 1D ^{1}H NMR spectra obtained at 45 °C for Fe(II)Rd (A) and Zn(Rd) (B) showing signals due to backbone amide and aromatic side-chain protons. Substantial signal loss in the Fe(II)Rd spectrum (due to extreme paramagnetic line broadening) precluded 2D NMR studies of the metal-binding residues. (C) Downfield region of the 2D NOESY spectrum obtained for Zn(Rd) (7.2 mM, τ_m = 210 ms). Amide-to-amide proton dipolar cross peaks used to make sequential signal assignments and to determine secondary structure are indicated. Signals due to NH protons of residues in the metal-binding site are outlined.

NMR Data for Zn(Rd). In contrast to results obtained for Fe(II)Rd, high-quality 1D and 2D NMR data were obtained for the zinc-substituted protein. The downfield region of the 1D ^{1}H NMR spectrum is compared to the spectrum obtained for Fe(II)Rd at the top of Figure 3. Line widths of ca. 6 Hz were measured for several backbone amide and methyl group proton signals [45 °C, 7 mM Zn(Rd), 200 mM NaCl, 25 mM acetate-d_3, pH 6.3]; these signals are ca. 4 times narrower than corresponding resonance lines observed in the 1D spectrum of Fe(II)Rd (Figure 3). Cross peaks for 47 backbone amide proton signals were readily identified in the 2D HOHAHA spectrum of Zn(Rd) (Figure 4).

Comparison of the NOESY and HOHAHA data revealed that most of the cross-peak patterns (including relative intensities and chemical shifts) observable in spectra of Fe(II)Rd could also be seen in the spectra of the zinc adduct (Figure S4). Indeed, both spectra contain an upfield-shifted leucine methyl signal (δ = −1.6 ppm), which exhibits NOE cross relaxation to aromatic proton signals of Phe, Trp, and Tyr residues (Figures 2 and S4). The upfield shift of the Leu methyl signal can not be due to a paramagnetic contact effect. Contact effects were proposed to contribute to the shielding of an upfield methyl signal in the 1D spectrum of *C. pasteurianum* Rd (Krisnamoorthi et al., 1986).

Signal Assignments for Zn(Rd). Sequence-specific ^{1}H NMR signal assignments were made via standard procedures (Wüthrich, 1986). Spin systems were identified and categorized based on scalar connectivities observed in 2D DQF-COSY and HOHAHA spectra. Dipolar connectivities observed in 2D NOESY spectra were then used to make sequential assignments.

(*i*) *Spin System Identification.* As established by chemical methods (above) and NMR methods (below), Rd contains the following residues that generally exhibit unique scalar spin systems: five Gly, three Ala, one Thr, two Val, four Ile, two Leu, and five Pro. In addition, *P. furiosus* Rd contains 20 residues with AMX spin systems (corresponding to Hα–Hβ,β' of seven Asp, one Asn, four Cys, two Ser, two Phe, two Tyr, and two Trp) and 11 residues with "longer" side chains (including six Glu and five Lys).

Several unique spin systems are visible in the "fingerprint" regions of the 2D HOHAHA and DQF-COSY data (see Figures 4 and 5, respectively). Expected NH to CHα,α' cross

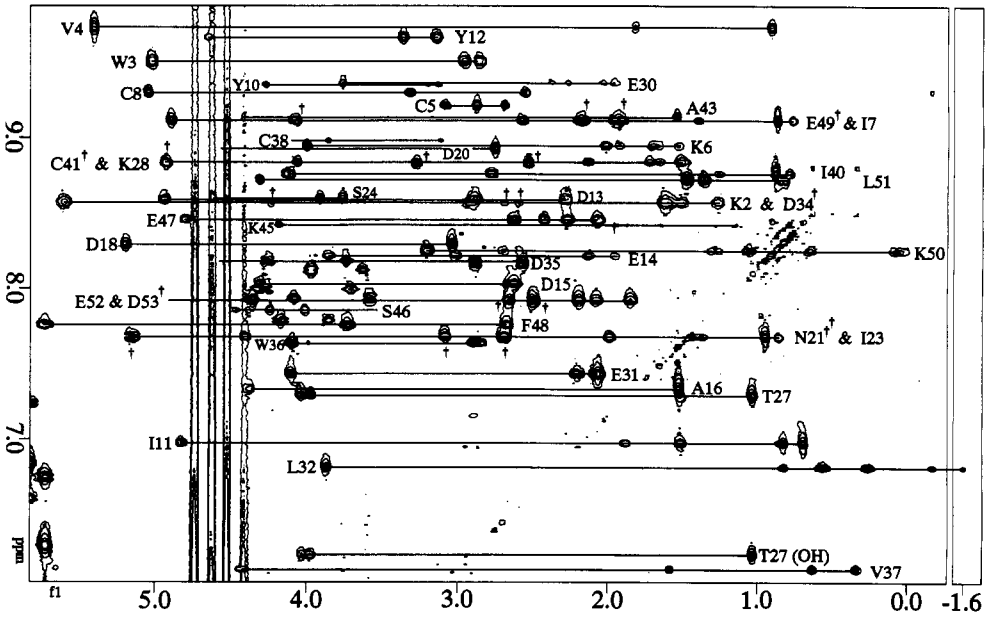

FIGURE 4: Fingerprint region of the 500-MHz HOHAHA spectrum obtained for Zn(Rd) (90%H_2O/10%D_2O, 45 °C). Spin systems for 39 of the 45 expected NH to αH and side chains resonances are labeled; NH to α,α'H connectivities of Gly residues have not been labeled (see Figure 5). Daggers identify signals of unique spin systems associated with overlapping NH signals.

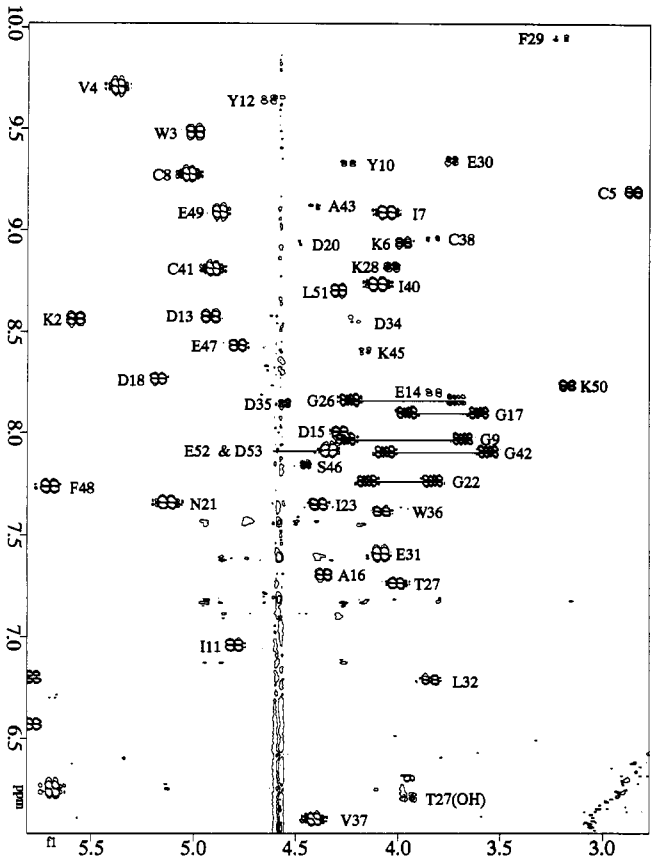

FIGURE 5: Portion of the 500-MHz DQF-COSY spectrum of *P. furiosus* Zn(Rd) showing HN to αH and T27 OH to βH connectivities. A SCUBA sequence was incorporated to allow for the detection of cross peaks associated with α protons which had been saturated. Note the large chemical shift dispersion of HN and Hα resonances: 6.11 (V37) to 9.95 ppm (F29) and 2.85 (C5) to 5.70 ppm (F48), respectively.

peaks for all five of the Gly residues are clearly visible in the DQF-COSY spectrum (Figure 5). In addition, cross peaks for two Ala residues (not including the N-terminal Ala) and the Thr residue were readily identified in the 2D HOHAHA spectrum (Figure 4). Although Leu and Ile residues generally exhibit unique scalar connectivity patterns, overlap in the upfield region of the DQF-COSY spectra precluded differentiation of several of these spin systems. One of the Leu residues exhibited a large upfield shift for a methyl group (δ = −1.6 ppm, see Figures 2 and 4), facilitating the spin system assignment for this residue. Most of the AMX networks were identified, but additional unambiguous assignments of the spin systems were not made due to overlap in the upfield region of the DQF-COSY spectrum (data not shown).

Spin systems for aromatic side-chain protons were also assigned from the 2D HOHAHA and DQF-COSY spectra. Unique spin systems for the two Trp, two Tyr, and two Phe residues are labeled in the downfield region of the HOHAHA spectrum given in Figure 6.

(ii) Sequential Signal Assignments. Sequential assignments were made using 2D NOESY data collected in 90% H_2O/10% D_2O solution with a mixing time of 210 ms. In general, either NH–NH or NH–αH sequential connectivities were identified, and, in several cases, both NH–NH and NH–αH connectivities were observed. For prolines, which lack a backbone amide proton, sequential connectivities were established via δ,δ'H(i) to αH(i−1) cross relaxation.

The downfield region of the NOESY spectrum showing NH–NH connectivities is shown in Figure 3. In general,

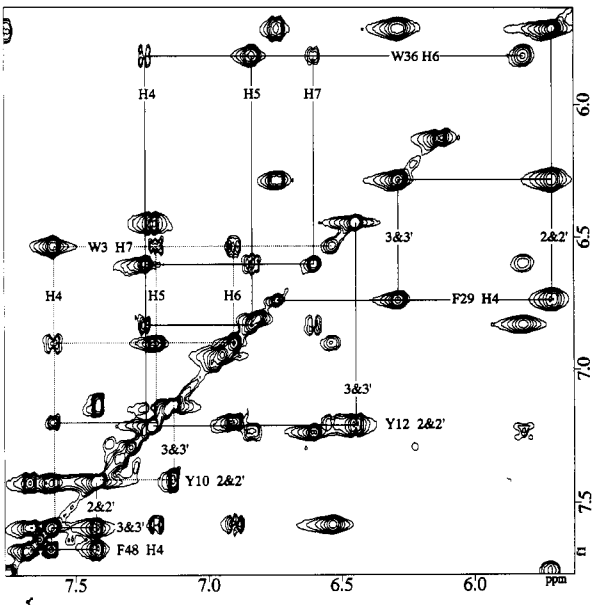

FIGURE 6: Downfield region of the 500-MHz HOHAHA spectrum of Zn(Rd). Spin systems for all six aromatic side chains are labeled.

sequential NH–NH connectivities were observed only for stretches of four sequential amino acids or less. For example, beginning with the Ala NH signal at 9.11 ppm, sequential connectivities were observed to the amide proton signals at 7.90 ppm (Gly), 8.81 ppm (AMX spin system), and 8.73 ppm (Ile or Leu spin system) (Figure 3). On the basis of the amino acid sequence, these residues could be uniquely assigned (and were later confirmed, based on long-range dipolar connectivity patterns) as Ala(43)-Gly(42)-Cys(41)-Ile(40). These residues contribute to the metal-binding site of Rd, and signals for these residues were not detected in 2D spectra obtained for the native protein. Additional NH–NH sequential connectivities were observed for amide signals at 8.40 ppm (long side chain spin system), 7.84 ppm (Ser), 8.43 ppm (long side chain spin system), and 7.74 ppm (aromatic containing side chain). On the basis of the primary sequence, these spins systems were assigned as Lys(45)-Ser(46)-Glu(47)-Phe(48).

Several stretches of sequential NH–Hα connectivities were observed in the fingerprint region of the NOESY spectrum (see Figure 7). Beginning with the Ala α proton signal at 4.21 ppm [which did not correlate with a detectable backbone amide proton and was later identified as Ala(1)], a sequential αH–NH connectivity involving the NH signal at 8.56 ppm was observed (Figure 7). The 8.56 ppm amide comprises a "long side chain" residue with an α proton chemical shift of 5.58 ppm, based on the HOHAHA and DQF-COSY data. The 5.58 ppm Cα proton exhibited a strong Hα–NH connectivity with the backbone amide proton signal at 9.48 ppm; this latter signal is associated with an AMX spin system which was identified as a Trp on the basis of dipolar connectivities between the Trp aromatic and nonaromatic protons. These three residues were thus assigned as Ala(1)-Lys(2)-Trp(3), and the sequential connectivities extended to include residues Val(4), Cys(5), and Lys(6). Hα–Hδ,δ'(Pro) connectivities involving Asp(18) and Pro(19) are also visible in Figure 7. Sequential HN–HN, Hα–NH, and Hα–Hδ,δ'(Pro) connectivities are summarized in Figure 8.

As indicated above, the amino acid sequence analysis of the C-terminus of the protein was somewhat ambiguous, providing limited evidence for the presence of a C-terminal Asp. Sequential signal assignments described above suggested the presence of 53 residues. Unfortunately, the chemical shifts

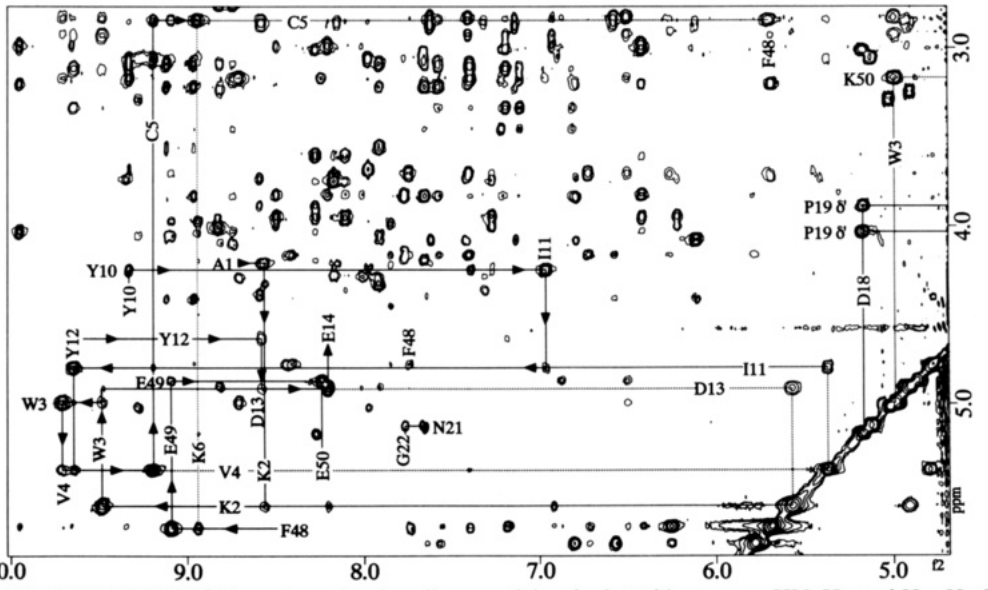

FIGURE 7: Portion of the 500-MHz NOESY spectrum showing all sequential and selected long-range HN–Hα and Hα–Hα NOE cross peaks for protons involved in the β-sheet region of Zn(Rd). The solid arrows denote sequential HN–αH cross peaks observed for residues A1–K6, Y10–E14, and F48–E51. Broken lines denote long-range connectivities that also define the β-sheet.

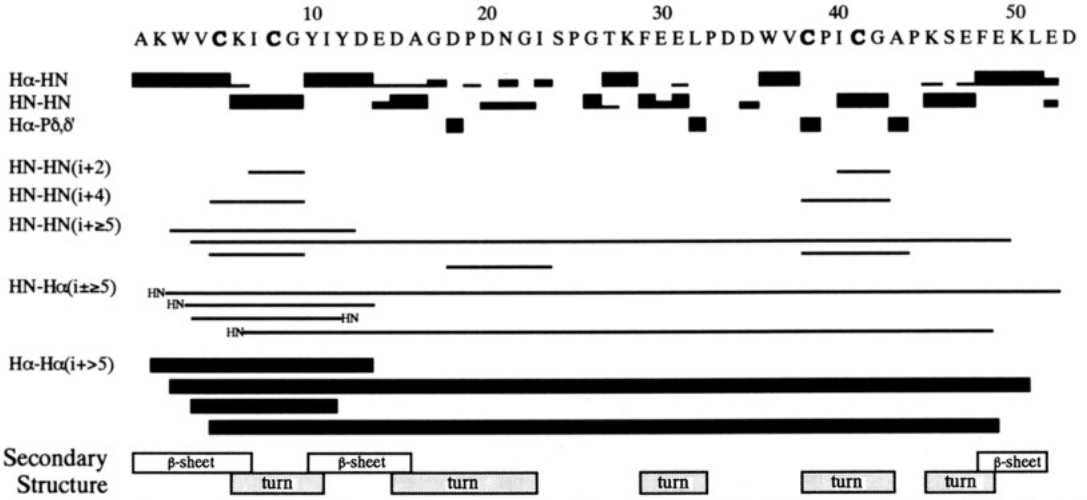

FIGURE 8: Summary of long (i to $>i+5$), medium (i to $<i+5$) and sequential (i to $i+1$) HN–HN, HN–Hα, Hα–Hα, and Hα–Proδ,δ′ connectivities observed in 2D NOESY spectra of Zn(Rd). The thickness of the bars reflects the relative intensities of the cross peaks observed. Secondary structural elements are given at the bottom of the figure.

of the HN and Hα proton signals of the 53rd residue overlapped with the HN and Hα proton signals, respectively, of Glu(52), precluding the observation of sequential connectivities involving these protons. Fortunately, the chemical shift degeneracies were not present in 2D NOESY and HOHAHA data collected at 25 °C and 600-MHz field strength. As shown in the expansion of Figure 2, sequential NOE connectivities involving the amide proton of the additional residue and the side-chain protons of Glu(52) were observed, providing strong evidence that the protein indeed contains an Asp at position 53. ^{1}H NMR signal assignments are summarized in Table I.

(*iii*) *Secondary Structure*. Regular secondary structural elements in *P. furiosus* Zn(Rd) can be identified from the data summarized in Figure 8. Moderate to strong NH–NH sequential connectivities observed for residues Lys(6)–Tyr(10), Glu(14)–Gly(17), Asp(20)–Ile(23), Gly(26)–Thr(27), Phe(29)–Leu(32), Asp(35)–Trp(36), Ile(40)–Ala(43), and Lys(45)–Phe(48) indicate that these stretches comprise tight turns and/or helical corners. Residues Cys(5) and Cys(38), which are located in the metal-binding site, exhibit NH–NH(i+4) connectivities to residues Gly(9) and Gly(42), respectively. Residues Ile(7) and Ile(40) exhibit NH(i)–NH(i+2) connectivities to residues Gly(9) and Gly(42), respectively (see signals in boxes, Figure 3 and also Figure 8). Similar nonsequential NH(i)–NH(i+2,4) connectivity patterns were observed by NMR spectroscopy for analogous Cys-X-X-Cys-Gly-X residues (X = variable amino acid) in the Zn domains of the HIV-1 nucleocapsid protein (NCP) (Summers et al., 1990; South et al., 1990, 1991). Since the secondary structures of these residues in the HIV-1 NCP and in *C. pasteurianum* Rd (determined by X-ray crystallography) match (Summers et al., 1990; South et al., 1991), the secondary folding of these residues in Zn-substituted *P. furiosus* Rd and in native *C. pasteurianum* Rd must also match. No stretches of NH(i)–NH(i+2) or αH(i)–NH(i+4) connectivities were observed in the 210-ms NOESY spectrum of Zn(Rd), indicating that the protein does not contain long α-helical stretches.

Strong αH–NH connectivities observed for residues Ala(1)–Lys(6), Tyr(10)–Glu(14), and Phe(48)–Glu(52) indicate that these regions exist in a regular extended conformation. Long-range αH–αH, αH–NH, and NH–NH connectivities

Table I: ^{1}H NMR Chemical Shifts for Zn(Rd), pH 6.3, T = 45 °C

residue		chemical shift δ (ppm)[a]			
		HN	Hα	Hβ	others
1	Ala		4.21	1.63	
2	Lys	8.56	5.58		1.62, 1.51, 1.27
3	Trp	9.48	5.00	2.96, 2.85	H2 6.92; NH 9.66; H4 7.56; H5 7.19; H6 6.83; H7 6.50
4	Val	9.71	5.37	1.81	γCH_3 0.91
5	Cys	9.19	2.85	3.09, 2.69	
6	Lys	8.94	3.98		2.01, 1.92, 1.68, 1.53
7	Ile	9.09	4.06		γCH_2 1.40; γCH_3 0.87; δCH_3 0.77
8	Cys	9.27	5.02	3.31, 2.55	
9	Gly	7.97	4.26, 3.68		
10	Tyr	9.33	4.25	3.20, 3.13	H2, 2′ 7.38, H3, 3′ 7.11
11	Ile	6.96	4.80	1.51	γCH_2 1.88; γCH_3 0.70; δCH_3 0.83
12	Tyr	9.65	4.64	3.36, 3.14	H2, 2′ 7.18; H3, 3′ 6.42; OH 8.49
13	Asp	8.57	4.92	2.89, 2.28	
14	Glu	8.20	3.83	2.13, 1.92	γCH_2 3.02
15	Asp	8.00	4.28	2.64	
16	Ala	7.30	4.36	1.52	
17	Gly	8.10	3.96, 3.61		
18	Asp	8.27	5.17	3.01	
19	Pro		4.19		δCH_2 4.03, 3.89
20	Asp	8.93	4.58[b]	2.73	
21	Asn	7.65	5.13	3.08, 2.70	
22	Gly	7.76	4.15, 3.83		
23	Ile	7.65	4.39	1.99	γCH_2 1.44, 1.37; γCH_3 0.95; δCH_3 0.86
24	Ser	8.60	4.58[b]	3.91, 3.76	
25	Pro		3.75		
26	Gly	8.16	4.24, 3.71		
27	Thr	7.26	4.01	3.95	γCH_3 1.10; OH 6.22
28	Lys	8.82	4.03	1.51, 2.13	1.72, 1.66, 1.51
29	Phe	9.95	3.21		H2, 2′ 5.71; H3, 3′ 6.28; H4 6.74
30	Glu	9.34	3.74	2.03, 1.96	γCH_2 2.37 and 2.26
31	Glu	7.41	4.08	2.06	γCH_2 2.20
32	Leu	6.79	3.84	0.56, 0.26	γCH 0.83; δCH_3 −0.17, −1.61
33	Pro		4.32		γCH_2 3.23, 3.71
34	Asp	8.56	4.22	2.60, 2.69	
35	Asp	8.14	4.56	2.89, 2.57	
36	Trp	7.61	4.08	2.88	H2 7.13; HN 10.24; H4 7.23; H5 6.83; H6 5.81; H7 6.60
37	Val	6.11	4.41	1.58	γCH_3 0.61, 0.32
38	Cys	8.96	3.83	3.09, 3.22	
39	Pro		4.03		δCH_2 3.45, 3.18
40	Ile	8.73	4.10	2.78	γCH_2 1.26; γCH_3 0.88; δCH_3 0.79
41	Cys	8.81	4.91	3.27, 2.53	
42	Gly	7.90	4.06, 3.56		
43	Ala	9.11	4.41	1.54	
44	Pro		4.77		δCH_2 3.86, 3.60
45	Lys	8.40	4.16		1.92, 1.85, 1.64, 1.15
46	Ser	7.84	4.45	4.24, 4.01	
47	Glu	8.43	4.78	2.06, 2.26	γCH_2 2.63, 2.42
48	Phe	7.74	5.70	3.72, 2.68	H2, 2′ 7.42; H3, 3′ 7.59; H4 7.68
49	Glu	9.09	4.87		
50	Lys	8.23	3.17	1.29, 1.06	γCH_2 0.65, 0.07; δCH_2 2.70
51	Leu	8.70	4.29	1.47, 1.36	δCH 0.85, 0.82
52	Glu	7.92	4.32		2.19, 2.07, 1.85
53	Asp	7.91	4.32	2.48, 2.64	

[a] Chemical shifts were referenced to H_2O, 4.58 ppm at 45 °C. [b] Resonances which reside under H_2O.

further demonstrate that these segments comprise a three-strand antiparallel β-sheet. For example, as shown in Figure 7, residues Asp(13) and Ile(11) exhibit strong αH–αH connectivities to residues Lys(2) and Val(4), respectively. Long-range and sequential dipolar connectivities that define the β-sheet domain are summarized in Figure 9.

Discussion

Physical Comparisons of Rubredoxins. The data presented herein show that *P. furiosus* Rd is very similar in molecular size, Fe content, and UV–visible absorption properties to the rubredoxins that have been previously purified from mesophilic eubacteria. As might be expected, however, as the first rubredoxin to be characterized from a hyperthermophilic organism, the *P. furiosus* protein is extremely thermostable, being unaffected by incubation for 24 h at 95 °C. For comparison, the rubredoxin from *C. pasteurianum* is rapidly denatured at 80 °C (Lovenberg & Sobel, 1965), while the protein from *Desulfovibrio gigas* loses 50% of its visible absorption ($t_{50\%}$) after approximately 2 h at 80 °C (Papavassiliou & Hatchikian, 1985). The rubredoxin from the moderately thermophilic sulfate-reducing bacterium *Thermodesulfobacterium commune*, which grows optimally at 70 °C, is somewhat more stable, having a $t_{50\%}$ value of about 6 h at 80 °C (Papavassiliou & Hatchikian, 1985).

Sequence Homologies with Other Rubredoxins. The sequence of *P. furiosus* Rd is aligned with those of the rubredoxins from 10 mesophilic eubacteria and one moderately thermophilic eubacterium (*C. thermosaccharolyticum*) in Figure 1. All comprise between 52 and 54 residues, with the exception of the *Desulfovibrio desulfuricans* protein, which has 45 residues. Meyer et al. (1990) noted 15 conserved

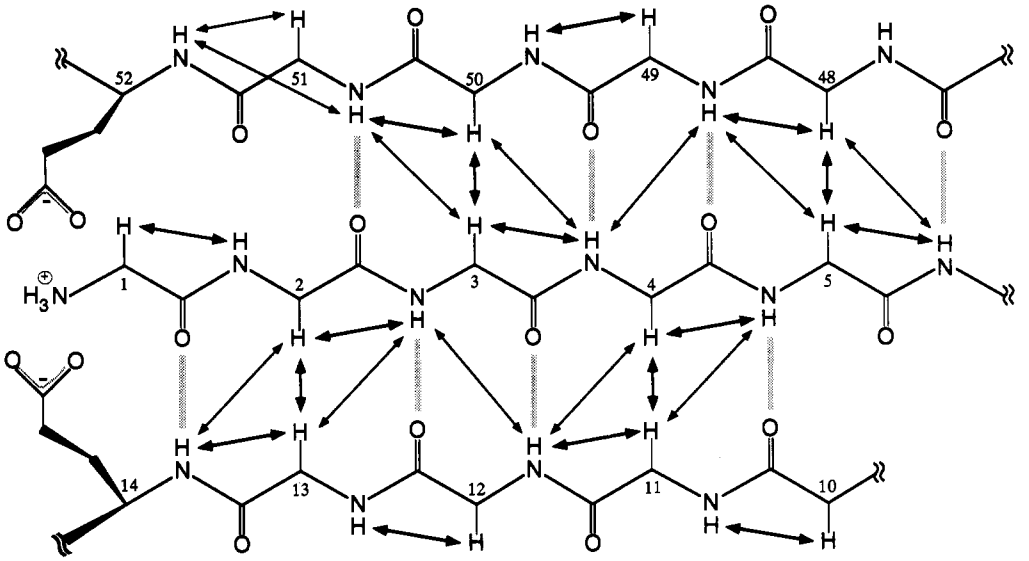

FIGURE 9: Summary of the strong (heavy arrows) and medium or weak (light arrows) NOE connectivities that define the β-sheet domain of *P. furiosus* Zn(Rd). This pleated sheet includes residues A1–K6, Y10–E14, and F48–E51. The arrows correspond to labeled cross peaks shown in Figure 6. Carboxyl groups of Glu(14) and Glu(52) and the N-terminal amine of Ala(1), which are proposed to contribute to hyperthermostability via electrostatic interactions, are also shown.

residues in their comparison of 10 rubredoxin sequences, and these were also present in the subsequently published sequence of rubredoxin from *Desulfovibrio vulgaris* strain Miyazaki (Shimizu et al., 1989). As indicated in Figure 1, 14 of these residues are also conserved in *P. furiosus* rubredoxin, the exception being the N-terminal methionine residue. The consensus residues are the four cysteinyl residues that bind the Fe atom together with an adjacent prolyl residue, five aromatic residues, which from crystal studies constitute a hydrophobic core, two glycyl residues, and one lysyl and one aspartyl residue. Figure 1 shows that residues of similar property are found in all the rubredoxins at positions 4 (aromatic), 8, and 34 (Ile, Leu, or Val), and 44 (Ala or Val). Other than these, there are several positions that are occupied by only one residue in most of the sequences, but the exceptions show that a particular type of residue cannot be essential for the structure or function of rubredoxins in general. For example, all of the rubredoxins contain a lysyl residue at position 3, except the *D. gigas* protein, which has an isoleucyl group. Similarly, all contain a prolyl residue at position 15, except the *P. furiosus* and *M. elsdenii* proteins, which contain a glutamyl group.

The overall sequence homology (identity) of *P. furiosus* rubredoxin with the other rubredoxins (Figure 1) ranges from 42% (with *D. desulfuricans*) to 67% (with *D. vulgaris* strain Miyazaki). This is similar to the homologies found between the other rubredoxins, which range between 41% (*P. aerogenes* and *D. desulfuricans*) and 90% (*D. vulgaris* strains Miyazaki and Hildenborough). Also shown in Figure 1 are the seven residues that are found in *P. furiosus* rubredoxin but in none of the others (excluding the absence of the N-terminal methionine). Of significance to the enhanced thermostability of the *P. furiosus* protein may be the fact that only in this rubredoxin is there (a) not a charged or amide group at position 2 (Ala), (b) a charged residue at position 7 (Lys), (c) a hydroxy group (Ser) at position 25, and (d) a glutamyl rather than aspartyl group at position 32 (note that in this numbering system, which refers to all of the rubredoxin sequences shown in Figure 1, *P. furiosus* rubredoxin lacks residue 1). A previous study of the rubredoxin from the moderate thermophile *C. thermosaccharolyticum* (Meyer et al., 1990) identified four unique residues that might render this protein more stable compared with those of mesophilic species. These included a tryptophan residue at position 4, which is also present in the *P. furiosus* protein. However, the other three unique residues (Thr, Pro, and Asp at positions 21, 25, and 41, respectively) are not found in the hyperthermophilic protein (Asp, Ser, and Ile at positions 21, 25, and 41, respectively).

The complete amino acid sequence of only one other protein from a hyperthermophilic organism, i.e., one that grows optimally ≥100 °C, has been reported: that of glyceraldehyde-3-phosphate dehydrogenase (GAPD, M = 36 500) from the related bacterium *P. woesii* (Zwickl et al., 1990). Its amino acid sequene was deduced from the gene sequence. Comparisons with homologous mesophilic and thermophilic enzymes from eubacteria and eucaryotes were consistent with previous studies (Pauptit et al., 1988; Zuber, 1988; Menendez-Arias & Argos, 1989; Fontana, 1990) which showed a preference for alanine and discrimination against serine and glycine in the "thermophilic" proteins (Zwickl et al., 1990). Contrary to previous findings with non-archaebacterial proteins, GAPD from thermophilic and hyperthermophilic archaebacteria did not show the characteristic lysine to arginine and serine to threonine exchanges, while the hyperthermophilic GAPD showed a striking preference for phenylalanine and discrimination against aspartate, methionine, and cysteine (Zwickl et al., 1990). However, none of these preferences appear to be a general feature of "hyperthermostability" since none are evident when one compares the hyperthermophilic rubredoxin with the others listed in Figure 1. The exception may be the lack of methionine in *P. furiosus* rubredoxin, whereas the complete preference of lysine for arginine in this protein is not significant since none of the rubredoxins contain arginine, and the *D. gigas* protein has the same lysine content (five residues) as *P. furiosus* rubredoxin. Indeed, systematic comparisons show no obvious preference for any amino acid in the latter protein, with the exception of isoleucine (four in *P. furiosus* Rd, three in the *C. thermosaccharolyticum* Rd, and two or less in the others).

NMR and Structural Findings. As expected, zinc-substitution provides an effective means for overcoming paramagnetic-induced bleaching in the NMR spectra of *P. furiosus* Rd. The fact that the intensities and chemical shifts of observable cross peaks in the 2D NOESY spectrum of Fe(II)Rd

match relevant cross peaks in the spectrum of Zn(Rd) provides good evidence that the global folding is the same. Both proteins give rise to upfield-shifted Leu methyl signals, and the Leu methyl signals of both samples exhibit dipolar cross relaxation to aromatic protons of Trp, Phe, and Tyr residues (Figures 2 and S4).

In addition, many secondary structural elements observed in *P. furiosus* Zn(Rd) match those of *C. pasteurianum* Rd. Thus, both proteins contain a three-stranded antiparallel β-sheet and several tight turns and a hydrophobic core. Also, cross-peak patterns for residues that comprise the metal-binding site of *P. furiosus* Rd are consistent with types I and II NH–S tight turns common to Fe–S proteins. The presence of strong interresidue cross relaxation involving the aromatic side chains of residues Trp(3), Tyr(10), Tyr(12), Phe(29), Trp(36), and Phe(48) indicates that these residues comprise a hydrophobic core in *P. furiosus* Rd. Thus, the combined data provide good evidence that zinc substitution has not led to significant changes in protein tertiary structure and that the tertiary structure of *P. furiosus* Zn(Rd) is similar to that of native *C. pasteurianum* Rd. Indeed, we expected Zn substitution to produce minimal structural changes at the metal-binding site since, as demonstrated previously, the backbone conformation of Cys-X-X-Cys-Gly-X residues in the zinc domain of retroviral-type zinc finger proteins is essentially identical to the folding of relevant residues in the Fe domain of Rd (see above). Structural studies of the native (by X-ray diffraction) and zinc-substituted (by NMR) proteins, now underway, will enable a detailed assessment of the effect of zinc substitution on protein structure, and a general assessment of the utility of zinc substitution for NMR-based structural studies of paramagnetic metalloproteins.

It is noteworthy that both the Fe and Zn forms give a very upfield-shifted Leu methyl signal ($\delta = -1.6$ ppm). Although the 3D structure determination of *P. furiosus* Rd is not yet complete, numerous interresidue cross peaks involving the side-chain protons of Tyr(12), Phe(29), Leu(32), and Trp(36) provide good evidence that these hydrophobic groups pack in a manner similar to that observed for the related Tyr(13), Phe(30), Ile(33), and Trp(37) residues of *C. pasteurianum* Rd. The 1D ^{1}H NMR spectrum of *C. pasteurianum* Rd exhibits several upfield-shifted signals ($\delta < 0$ ppm), all of which have been attributed (to some extent) to paramagnetic contact effects. Since *C. pasteurianum* Rd contains an Ile substituted for the Leu(32) of *P. furiosus* Rd, it is possible that the upfield shift of one of the signals observed for *C. pasteurianum* Rd is due predominantly to aromatic shielding.

Perhaps the most significant finding thus far is that the N-terminal residue of *P. furiosus* Rd is incorporated into the β-sheet via apparent hydrogen bonding between the Ala(1) carbonyl and the Glu(14) amide. This secondary structure implicitly places the Glu(14) side chain carboxyl group adjacent to the N-terminal amine of Ala(1) (Figure 9). The carboxyl group of Glu(52) may also be in close proximity to the N-terminal amine (Figure 9). The resulting electrostatic interactions, possible only for *P. furiosus* Rd, may contribute significantly to protein hyperthermostability by stabilizing one end of the three-stranded β-sheet, thus preventing the protein from "unzipping" at elevated temperatures. By comparison, rubredoxins from mesophilic organisms contain, with one exception, a Pro at position 14 [where the first Cys is labeled Cys(5)], and they are longer by one residue at the N-terminus. Thus residues at relevant positions in rubredoxins from mesophilic organisms are not capable of conferring the potentially stabilizing electrostatic interactions. Temperature-dependent NMR-based structural studies now underway will allow us to evaluate these and other local interactions that may be determinants of protein hyperthermostability.

Supplementary Material Available

Two tables showing the amino acid composition and Edman degradation of the tryptic peptides of *P. furiosus* rubredoxin and four figures showing the complete amino acid sequence, reverse-phase HPLC analysis, UV–visible spectrum, and 2D NOESY and HOHAHA spectra of *P. furiosus* rubredoxin (7 pages). Ordering information is given on any current masthead page.

Registry No. RdPf, 136475-78-0.

References

Adams, M. W. W. (1990) *FEMS Microbiol. Rev. 75*, 219.

Adman, E. T., Sieker, L. C., Jensen, L. H., Bruschi, M., & LeGall, J. (1977) *J. Mol. Biol. 112*, 113.

Adman, E.; Watenpaugh, K. D., & Jensen, L. H. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 4854.

Aono, S., Bryant, F. O., & Adams, M. W. W. (1989) *J. Bacteriol. 171*, 3433–3439.

Bachmeyer, H., Yasunobu, K. T., Peel, J. L., & Mayhew, S. (1968a) *J. Biol. Chem. 243*, 1022.

Bachmeyer, H., Benson, A. M., Yasunobu, K. T., Garrard, W. T., & Whiteley, H. R. (1968b) *Biochemistry 7*, 986.

Blumentals, I. I., Robinson, A. S., & Kelly, R. M. (1990) *Appl. Environ. Microbiol. 56*, 1992–1998.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521.

Brown, S. C., Weber, P. L., & Mueller, L. (1988) *J. Magn. Reson. 71*, 166.

Bruschi, M. (1976a) *Biochim. Biophys. Acta 434*, 4.

Bruschi, M. (1976b) *Biochem. Biophys. Res. Commun. 70*, 615.

Bryant, F. O., & Adams, M. W. W. (1989) *J. Biol. Chem. 264*, 5070–5079.

Carter, C. W., Jr. (1977) *J. Biol. Chem. 252*, 7802.

Chen, J.-S., & Mortenson, L. E. (1977) *Anal. Biochem. 79*, 157.

Costantino, H. R., Brown, S. H., & Kelly, R. M. (1990) *J. Bacteriol. 172*, 3654–3660.

Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820.

Fiala, G., & Stetter, K. O. (1986) *Arch. Microbiol. 145*, 56.

Fontana, A. (1990) in *Life Under Extreme Conditions: Biochemical Adaptation* (di Prisco, G., Ed.) pp 89–113, Springer-Verlag, Heidelberg.

Frey, M., Sieker, L. C., Payan, F., Haser, R., Bruschi, M., Pepe, G., & LeGall, J. (1987) *J. Mol. Biol. 197*, 525.

Hore, P. J. (1983) *J. Magn. Reson. 55*, 283–300.

Hormel, S., Walsh, K. A., Prickril, B. C., Titani, K., LeGall, J., & Sieker, L. C. (1986) *FEBS Lett. 201*, 147.

Jeener, B. H., Meier, P., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546.

Krishnamoorthi, R., Cusanovick, M. A., Przysiecki, C. T., & Markely, J. L. (1986) *Biochemistry 25*, 2187.

Lovenberg, W., & Sobel, B. E. (1965) *Proc. Natl. Acad. Sci. U.S.A. 54*, 193–199.

Lovenberg, W., Buchanan, B. B., & Rabinowitz, J. C. (1963) *J. Biol. Chem. 238*, 3899.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265.

Lundell, D., & Howard, J. B. (1978) *J. Biol. Chem. 253*, 3422–3426.

Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95.

Menendz-Arias, L., & Argos, P. (1989) *J. Mol. Biol. 206*, 397.

Meyer, J., Gagnon, J., Sieker, L. C., van Dorsselaer, A., & Moulis, J.-M. (1990) *Biochem. J. 271*, 839.
Mukund, S., & Adams, M. W. W. (1990) *J. Biol. Chem. 265*, 11508–11516.
Mukund, S., & Adams, M. W. W. (1991) *J. Biol. Chem. 266*, 14208.
Oh, B.-H., & Markley, J. L. (1990) *Biochemistry 29*, 3993.
Papavassiliou, P., & Hatchikian, E. C. (1985) *Biochim. Biophys. Acta 810*, 1.
Paulptit, R. A., Karlsson, R., Picot, D., Jenkins, J. A., Niklaus-Reimer, A., & Jansonius, J. N. (1988) *J. Mol. Biol. 199*, 525.
Pinantini, U., Sorensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800.
Saeki, K., Yao, Y., Wakabayashi, S., Shen, G.-J., Zeikus, J. G., & Matsubara, H. (1989) *J. Bacteriol. 171*, 4736–4741.
Seki, Y., Seki, S., Satoh, M., Ikeda, A., & Ishimoto, M. (1989) *J. Biochem. (Tokyo) 106*, 336.
Shimizu, F., Ogata, M., Yagi, T., Wakabayashi, S., & Matsubara, H. (1989) *Biochimie 71*, 1171.
Sieker, L. C., Stenkamp, R. E., Jensen, L. H., Prickril, B., & LeGall, J. (1986) *FEBS Lett. 208*, 73.
South, T. L., Blake, P. R., Sowder, R. C., III, Arthur, L. O., Henderson, L. E., & Summers, M. F. (1990) *Biochemistry 29*, 7786.
South, T. L., Blake, P. R., Hare, D. R., & Summers, M. F. (1991) *Biochemistry 30*, 6342.
States, D. J., Haberkorn, R. H., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Stetter, K. O. (1986) in *The Thermophiles: General, Molecular and Applied Microbiology* (Brock, T. D., Ed.) pp 39–74, John Wiley, New York.
Stetter, K. O., Fiala, G., Huber, G., Huber, R., & Segerer, G. (1990) *FEMS Microbiol. Rev. 75*, 117–124.
Summers, M. F., South, T. L., Kim, B., & Hare, D. R. (1990) *Biochemistry 29*, 329.
Voordouw, G. (1988) *Gene 69*, 75.
Watenpaugh, K. D., Sieker, L. C., Herriott, J. R., & Jensen, L. H. (1973) *Acta Crystallogr. B29*, 943.
Watenpaugh, K. D., Sieker, L. C., & Jensen, L. H. (1979) *J. Mol. Biol. 131*, 509.
Weber, K., Pringle, J. R., & Osborn, M. (1972) *Methods Enzymol. 26*, 2–27.
Werth, M. T., Kurtz, D. M., Jr., Moura, I., & LeGall, J. (1987) *J. Am. Chem. Soc. 109*, 273.
Woolley, K. J., & Meyer, T. E. (1987) *Eur. J. Biochem. 163*, 161–166.
Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Yasunobu, K. T., & Tanaka, M. (1973) in *Iron–Sulfur Proteins* (Lovenberg, W., Ed.) Vol. II, pp 27–130, Academic Press, New York.
Yu, L. P., LaMar, G. N., & Rajarathan, K. (1990) *J. Am. Chem. Soc. 112*, 9527.
Zuber, H. (1988) *Biophys. Chem. 29*, 171.
Zwickl, P., Fabry, S., Bogedain, C., Haas, A., & Hensel, R. (1990) *J. Bacteriol. 172*, 4329.

Construction of a Synthetic Gene for an R-Plasmid-Encoded Dihydrofolate Reductase and Studies on the Role of the N-Terminus in the Protein†

Lisa J. Reece,‡ Robert Nichols,‡ Richard C. Ogden,§ and Elizabeth E. Howell*,‡
Department of Biochemistry, University of Tennessee, Knoxville, Tennessee 37996-0840, and The Agouron Institute, 505 Coast Boulevard South, La Jolla, California 92037

Received May 24, 1991; Revised Manuscript Received August 1, 1991

ABSTRACT: R67 dihydrofolate reductase (DHFR) is a novel protein that provides clinical resistance to the antibacterial drug trimethoprim. The crystal structure of a dimeric form of R67 DHFR indicates the first 16 amino acids are disordered [Matthews et al. (1986) *Biochemistry 25*, 4194–4204]. To investigate whether these amino acids are necessary for protein function, the first 16 N-terminal residues have been cleaved off by chymotrypsin. The truncated protein is fully active with k_{cat} = 1.3 s^{-1}, $K_{m(NADPH)}$ = 3.0 μM, and $K_{m(dihydrofolate)}$ = 5.8 μM. This result suggests the functional core of the protein resides in the β-barrel structure defined by residues 27–78. To study this protein further, synthetic genes coding for full-length and truncated R67 DHFRs were constructed. Surprisingly, the gene coding for truncated R67 DHFR does not produce protein in vivo or confer trimethoprim resistance upon *Escherichia coli*. Therefore, the relative stabilities of native and truncated R67 DHFR were investigated by equilibrium unfolding studies. Unfolding of dimeric native R67 DHFR is protein concentration dependent and can be described by a two-state model involving native dimer and unfolded monomer. Using absorbance, fluorescence, and circular dichroism techniques, an average ΔG_{H_2O} of 13.9 kcal mol^{-1} is found for native R67 DHFR. In contrast, an average ΔG_{H_2O} of 11.3 kcal mol^{-1} is observed for truncated R67 DHFR. These results indicate native R67 DHFR is 2.6 kcal mol^{-1} more stable than truncated protein. This stability difference may be part of the reason why protein from the truncated gene is not found in vivo in *E. coli*.

Dihydrofolate reductase (DHFR;[1] EC 1.5.1.3) catalyzes the NADPH-dependent reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate. Since tetrahydrofolate is required for the synthesis of thymidylate, purine nucleosides, methionine, and other metabolic intermediates (Kraut & Matthews,

† This research was supported by NIH Grant GM35308 (to E.E.H.) and NSF Grant DMB8318244 (to R.C.O.).
‡ University of Tennessee.
§ The Agouron Institute.

[1] Abbreviations: DHFR, dihydrofolate reductase; TMP, trimethoprim; DHF, dihydrofolate; GdnHCl, guanidine hydrochloride.

0006-2960/91/0430-10895$02.50/0