Modeling the structure of *Pyrococcus furiosus* rubredoxin by homology to other X-ray structures

JOHN E. WAMPLER,¹ ELIZABETH A. BRADLEY,¹
DAVID E. STEWART,² AND MICHAEL W.W. ADAMS¹

¹ Department of Biochemistry and the Center for Metalloenzyme Studies, University of Georgia, Athens, Georgia 30602

² University Computing and Networking Services, University of Georgia, Athens, Georgia 30602

(RECEIVED July 17, 1992; REVISED MANUSCRIPT RECEIVED October 30, 1992)

Abstract

The three-dimensional structure of rubredoxin from the hyperthermophilic archaeobacterium, *Pyrococcus furiosus*, has been modeled from the X-ray crystal structures of three homologous proteins from *Clostridium pasteurianum*, *Desulfovibrio gigas*, and *Desulfovibrio vulgaris*. All three homology models are similar. When comparing the positions of all heavy atoms and essential hydrogen atoms to the recently solved crystal structure (Day, M.W., et al., 1992, *Protein Sci.* 1, 1494–1507) of the same protein, the homology models differ from the X-ray structure by 2.09 Å root mean square (RMS). The X-ray and the zinc-substituted NMR structures (Blake, P.R., et al., 1992b, *Protein Sci.* 1, 1508–1521) show a similar level of difference (2.05 Å RMS). On average, the homology models are closer to the X-ray structure than to the NMR structures (2.09 vs. 2.42 Å RMS).

Keywords: extreme thermostability; molecular modeling; protein modeling; rubredoxin

Although there is a growing body of data on the differences between analogous proteins isolated from thermophilic and mesophilic organisms (Bull & Breese, 1973; Stellwagen & Wilgus, 1978; Zuber, 1978; Fontana, 1991), little structural information is available on proteins that have been purified from the recently discovered hyperthermophilic archaeobacteria. All of these novel organisms are able to grow near to and even above 100 °C (see Adams & Kelly, 1992). Several proteins from hyperthermophiles have now been sequenced that are homologous to proteins with known structure (Costantino et al., 1990; Zwickl et al., 1990; Blake et al., 1991; Zulli et al., 1991; Adams and coworkers, unpubl.), and work is in progress to determine some of their structures. Along with accurate measurements of physical and chemical differences between these proteins and their mesophilic analogs, structural comparisons should greatly enhance our understanding of how extreme temperature stability can be engineered into proteins.

It is obvious that the keys to enhanced thermostability lie in the amino acid sequences of proteins, which deter-

mine atomic interactions within the folded structures. Because there does not appear to be a consistent composition or sequence difference between thermophilic and mesophilic proteins (Zuber, 1978; Costantino et al., 1990; Zwickl et al., 1990; Blake et al., 1991; Zulli et al., 1991), it is not clear at this time what aspects of protein structure contribute to thermostability. However, it is evident that striking changes in temperature stability can be engineered with very simple changes in sequence. For example, in the case of the bacilli lactate dehydrogenases (Zulli et al., 1991), sequence comparisons have led directly to a genetic modification of a mesophilic protein to enhance its thermostability by 20 °C with only two amino acid changes. Thermal stability studies of mutant *Escherichia coli* tryptophan synthase α subunits (Lim et al., 1992) have also shown that a single amino acid substitution substantially increases the subunit T_m .

Protein-stabilizing forces introduced by sequence and structure differences can be investigated using the tools of computational chemistry (for review see Wampler et al., 1992). Of course, such studies require a set of three-dimensional structures for a homologous series of proteins covering a range of thermostabilities. Although the best computational approach would make use of high-resolution X-ray or NMR structures, it is not always possible to obtain such structures. With an alternative

Reprint requests to: John E. Wampler, Department of Biochemistry, University of Georgia, Athens, Georgia 30602.

Abbreviations: Cp, *Clostridium pasteurianum*; Dd, *Desulfovibrio desulfuricans*; Dg, *Desulfovibrio gigas*; Dv, *Desulfovibrio vulgaris*; Pf, *Pyrococcus furiosus*; RMS, root mean square.

computational approach called homology modeling, however, it may be possible to develop a set of homologous structures for comparative studies by basing the structures of some of the members of the series on the known structures of others. One of the goals of this work was to investigate this approach in conjunction with separate and independent structure determinations using X-ray crystallographic (Day et al., 1992) and NMR techniques (Blake et al., 1992b).

Rubredoxin is a small electron transfer metalloprotein ($M_r \sim 5,500$). It is nearly ideal for homology modeling and for computational thermostability studies because many rubredoxin amino acid sequences have been reported (see Blake et al., 1991) and four high-resolution crystal structures have been solved and refined (Watnough et al., 1979, 1980; Frey et al., 1987; Stenkamp et al., 1990; Dauter et al., 1992). Sequence homology is high between the rubredoxin sequences, although there is considerable variation at some residue positions (Blake et al., 1991). The four known rubredoxin structures are very homologous in their three-dimensional conformation. All are comprised of a single peptide chain folded into one triple-stranded β -sheet region with several additional β bridges and tight turns. Despite this homology, thermostability data, though sketchy and incomplete, suggest that several members of the rubredoxin family exhibit a considerable range in temperature sensitivity (Lovenberg & Sobel, 1965; Papavassiliou & Hatchikian, 1985; Blake et al., 1991). At the extreme is the rubredoxin from the hyperthermophile, *Pyrococcus furiosus*, an organism that grows optimally at 100 °C. This rubredoxin is one of the most thermostable proteins yet studied (stable at 95 °C for 24 h [Blake et al., 1991]).

We have previously tested a method of homology modeling of proteins and have shown that it gives fairly accurate predictions when the sequence homology between the protein being modeled and the related known structure is greater than 60% identity (Stewart et al., 1987;

Stewart, 1989). The approach has been particularly successful with rubredoxins from mesophilic bacteria where the structural homology is very high. In one case, predicted structures showed less than 1.0 Å RMS deviation in the positions of all atoms when comparing a minimized X-ray structure and a predicted model. When an all-atom molecular mechanics force field was used, the difference between the main-chain atom positions for predicted and minimized X-ray structures approached 0.5 Å RMS.

This report presents the structures of homology modeled Pf rubredoxin as a component of a tripartite effort to define the structure of this protein using all three major structural approaches (X-ray crystallography, NMR spectroscopy, and molecular modeling). The results are compared with both the X-ray and NMR results and with our previous homology modeling efforts. This study is also the prelude to the first effort toward understanding hyperthermostability of proteins through molecular dynamics simulations (Bradley et al., 1993). The data indicate that the procedure of homology modeling leads to modeled thermostable rubredoxin structures that are significantly more stable than the mesophilic rubredoxin structures (as evaluated by the molecular mechanics energy function) and that the homology models are indeed reasonable approximations to experimental structures.

Results

Alignment and sequence comparisons

The aligned X-ray structures of the four known rubredoxins (Fig. 1; main-chain atoms only) reveal their close structural homology and the β -sheet structure. As shown in Table 1, the sequence of Pf rubredoxin exhibits strong homology to all four sequences of rubredoxins with known structure (Table 1). The degree of homology, quantified using the PAM 250 mutation probability index (Schwartz

Fig. 1. Main-chain atoms (N, C1, and Cα) for four rubredoxin crystal structures as indicated in the figure: Cp, *Clostridium pasteurianum* rubredoxin; Dd, *Desulfovibrio desulfuricans* rubredoxin; Dg, *D. gigas* rubredoxin; Dv, *D. vulgaris* rubredoxin. See text for source and authors for these coordinate sets.

Table 3. *Distribution of amino acid side-chain types in rubredoxins*

	Cp	Dg	Dv	Pf
Polar, + charged				
k	4	5	4	5
r	0	0	0	0
Totals	4	5	4	5
Polar, - charged				
d	10	8	5	6
e	6	3	4	6
Totals	16	11	9	12
Polar, uncharged				
c	4	4	4	4
m	1	1	1	0
n	1	0	1	1
q	0	1	0	0
s	0	2	2	2
t	3	2	3	1
Totals	9	10	11	8
Nonpolar				
a	0	4	5	3
g	6	5	5	5
i	2	2	0	4
l	1	1	1	2
p	5	5	6	5
v	5	3	5	2
Totals	19	20	22	21
Aromatic				
f	2	2	2	2
h	0	0	0	0
w	1	1	1	2
y	3	3	3	2
Totals	6	6	6	6

^a Key to symbols: ---- = gap (i.e., no residue in this position); ... = identical residues to *Clostridium pasteurianum* rubredoxin. Cp—*C. pasteurianum*; Dd—*Desulfovibrio desulfuricans*, strain 27774; Dg—*Desulfovibrio gigas*; Dv—*Desulfovibrio vulgaris* (Miyazaki); Pf—*Pyrococcus furiosus*.

& Dayhoff, 1978), is shown in Table 2. All scores for intercomparison of these rubredoxins are high. The average of scores for 500 comparisons to random sequences of the same composition was -10.06 with a standard deviation of 2.19 . Thus, the lowest rubredoxin vs. rubredoxin scoring (29.8 for Dd vs. Pf) is over 18 standard deviations from the mean of the random comparisons. The scores for "perfect" alignments are shown by the shaded diagonal of the table. Pf rubredoxin is clearly more closely related to the Cp, Dg, and Dv rubredoxins (in that order) than to the Dd protein.

In spite of the close homology between rubredoxins, the protein from Pf has 7 residues that are found uniquely in its aligned sequence compared to the other 11 known sequences (Blake et al., 1991). These are the Ala 1, Lys 6, Asp 15, Ala 16, Ser 24, Glu 31, and Leu 51 (using the Pf rubredoxin numbering). Overall, we can detect no pattern of substitution that changes the composition in a way that correlates with thermostability (Table 3).

Table 2. Alignment scoring for rubredoxin crystal structure sequences and the amino acid sequence for *Pf* rubredoxin^a

	Cp	Dg	Dv	Dd	Pf	Random ^b
Cp	56.5	41.6	45.4	34.3	42.3	-10.4
Dg	41.6	53.1	44.7	35.0	40.2	-9.9
Dv	45.4	44.7	52.9	37.5	41.6	-10.1
Dd	34.3	35.0	37.5	48.8	29.8	-9.2
Pf	42.3	40.2	41.6	29.8	55.4	-10.7

^a Average score per residue using the Schwartz and Dayhoff (1978) PAM 250 matrix. The PAM 250 matrix scores are log probabilities multiplied by 100 based on observed amino acid mutation. Positive values indicate substitution events with greater probability than chance, whereas negative values indicate less probable substitution events. Boxed values indicate self comparison of each sequence.

^b Average score (100 iterations) of random sequences based on the amino acid compositions of the five known rubredoxin sequences.

Homology modeling

The structure of Pf rubredoxin was modeled from each of its three close homologs, Cp, Dg, and Dv rubredoxins. The three resulting structures are similar (Fig. 2). This similarity can be shown quantitatively from the analysis of the minimized structures (Table 4). In all three cases, the AMBER energy and the radii of gyration of the models of Pf rubredoxin are nearly identical. Note that the values of the parent structures differ markedly from each other and from the modeled Pf proteins.

Intercomparison of the atomic positions in the modeled structures (Table 5) shows that the main-chain atomic positions are about equally similar, as are the all-atom comparisons. There is considerable variation in the positions of the C-terminal residues Glu 52 and Asp 53 as can be seen in Figure 2.

The structures from two separate experimental efforts for Pf rubredoxin have been published by Blake et al. (1992b) and Day et al. (1992) together with a comparison of the two data sets (Blake et al., 1992a). A major com-

Fig. 2. Three homology models of Pf rubredoxin as indicated in the figure: Pf/Cp, modeled from Cp; Pf/Dg, modeled from Dg; Pf/Dv, modeled from Dv. Upper stereo view is the backbone (N, C1, and C α); lower stereo view is of all the atoms.

ponent of this tripartite effort was the comparison of the structures from the three different approaches. The AMBER energy, radius of gyration, and molar volume of the crystal structure were all very similar to the values from the homology models (Table 4). Using the SYBYL fitting

function, several comparisons have been made (Tables 6, 7). From the crystallography study, both oxidized and reduced structures have been reported; the oxidized structure was used in the comparisons and is identified here as Pf-Xray. While the X-ray structure was refined, the struc-

Table 4. Comparison of the minimized homology modeled rubredoxins and their minimized parent proteins

Protein	Number of atoms (residues)	AMBER energy	Radius of gyration	Molar volume
Cp	504 (54)	-1,057	9.6	3,712
Pf/Cp	495 (53)	-1,181	9.5	3,584
Dg	480 (52)	-935	9.4	3,470
Pf/Dg	495 (53)	-1,130	9.5	3,529
Dv	469 (52)	-973	9.1	3,200
Pf/Dv	495 (53)	-1,072	9.5	3,593
Pf-Xray ^a	495 (53)	-1,103	9.5	3,578

^a Pf-Xray = Pf rubredoxin X-ray crystal structure.

Table 5. Comparison of the homology models with each other^a

		RMS position differences in Å		
	Atoms	Pf/Cp	Pf/Dg	Pf/Dv
Pf/Cp	ALL		1.89	1.82
Pf/Dg	ALL	1.89		2.27
Pf/Cp	MC		1.25	1.26
Pf/Dg	MC	1.25		1.89

^a Root mean square (RMS) values were calculated after the selected atoms of the two structures were superimposed by fitting (see Methods and materials). Pf rubredoxin structures are identified as in Table 1. ALL = all heavy atoms plus the "essential" hydrogen atoms, i.e., hydrogens on noncarbon atoms. MC = main-chain atoms include N, C1, C α , HN, and O.

Table 6. Comparison of Pf-Xray and minimized Pf-Xray structures versus the homology modeled structures^a

	Atoms	RMS differences in Å	
		X-ray	X-ray minimized
X-ray	ALL		1.32
X-ray minimized	ALL	1.32	
Pf/Cp	ALL	2.01	2.10
Pf/Dg	ALL	1.79	1.91
Pf/Dv	ALL	2.46	2.41
Avg. of models	ALL	2.09	2.14
X-ray	MC		0.99
X-ray minimized	MC	0.99	
Pf/Cp	MC	1.41	1.36
Pf/Dg	MC	1.16	1.15
Pf/Dv	MC	1.95	1.96
Avg. of models	MC	1.51	1.49

^a RMS values were calculated after the selected atoms of the two structures were superimposed by fitting (see Materials and methods). Pf rubredoxin structures are identified as in Table 1. ALL = all heavy atoms plus the "essential" hydrogen atoms, i.e., hydrogens on noncarbon atoms. MC = main-chain atoms include N, C1, C α , HN, and O. The Pf-Xray structure was provided by Day et al. (1992).

ture was not at the energy minimum according to the AMBER force field as indicated by the RMS comparison between X-ray and minimized X-ray structures (Table 6). In our previous studies, AMBER minimization of X-ray structures has produced structures that are more comparable to homology models than to the original X-ray structures (Stewart, 1989), which might be expected because the homology models are, of course, minimized. However, in the case of Pf rubredoxin there is no significant difference in the comparison of the models to the original X-ray structure or the minimized X-ray structure.

Forty solution-state zinc-substituted NMR structures were solved using a simulated annealing approach based on hybrid distance geometry (Blake et al., 1992b). A random sample of 21 of these NMR structures was used for comparison. Table 7 shows the statistics of the comparisons of each of these with the X-ray, minimized X-ray, and homology modeled structures. Figure 3 shows the global RMS backbone difference (C α , C1, and N) between each of the modeled structures and the Pf X-ray structures (original and minimized) with all 40 NMR structures. In order to see where the folding of the chains differs, the data of Figure 3 are presented for each residue.

Discussion

The results from homology modeling indicate that similar models of Pf rubredoxin may be built using different starting structures. Our previous studies (Stewart et al., 1987; Stewart, 1989) showed that mesophilic rubredoxin structures can be accurately predicted from each other

Table 7. Average comparison of 21 selected NMR structures with the X-ray and homology modeled structures^a

	Atoms	RMS position differences in Å			
		Average	SD	Lowest	Highest
X-ray	ALL	2.05	0.38	1.62	2.79
X-ray minimized	ALL	2.24	0.42	1.71	2.70
Pf/Cp model	ALL	2.44	0.25	2.09	2.96
Pf/Dg model	ALL	2.36	0.29	2.03	2.96
Pf/Dv model	ALL	2.47	0.21	2.02	2.88
X-ray	MC	1.32	0.39	0.83	2.13
X-ray minimized	MC	1.53	0.40	1.11	2.26
Pf/Cp model	MC	1.55	0.23	1.23	2.18
Pf/Dg model	MC	1.63	0.27	1.28	2.26
Pf/Dv model	MC	1.75	0.27	1.23	2.31

^a RMS values were calculated after the selected atoms of the two structures were superimposed by fitting (see Methods and materials). Pf rubredoxin structures are identified as in Table 1. Twenty-one NMR structures were selected from those provided by Blake et al. (1992b). For each comparison the RMS difference was calculated between each structure and all 21 NMR structures. The statistics reported (mean, standard deviation, and range) are for each group of 21 comparisons. ALL = all heavy atoms plus the "essential" hydrogen atoms, i.e., hydrogens on noncarbon atoms. For the NMR structures, hydrogen atoms were striped from the structure and then the essential hydrogens were added back. This perturbation represented about 0.1 Å RMS difference due to slightly out-of-register placement of hydrogens on the ϵ amino group of lysine. MC = main-chain atoms include N, C1, C α , HN, and O.

Fig. 3. Comparison of backbone root mean square (RMS) differences (C α , C1, and N) between the models and crystal structures and the 40 NMR structures of Blake et al. (1992b). The overall RMS difference for the three backbone atoms of each model were calculated after fitting each to all 40 NMR structures and are presented for each residue position. The heavy solid line is for the comparison with the Pf X-ray structure from Day et al. (1992), and the lighter solid line is for the minimized X-ray structure; short dashed line, Pf/Cp model; long dashed line, Pf/Dg model; dotted line, Pf/Dv model. For model designations see Figure 2.

Table 8. Comparison of homology modeled rubredoxins with known X-ray structures^a

Predicted structure	AMBER version used	Force field	Parent structure	Compared to	All atoms RMS	Main-chain ^b RMS	Reference
Dv	2.0	United	Cp	X-ray	0.93	0.51	Stewart et al., 1987
Cp	2.0	United	Dv	X-ray	1.15	0.51	Stewart et al., 1987
Dd	2.0	United	Dv	X-ray	1.84	0.59	Stewart et al., 1987
Dg	2.0	United	Cp	X-ray	3.02	0.74	Stewart et al., 1987
Dv	3.0	United	Cp	X-ray		1.15	Stewart, 1989
Dv	3.0	United	Cp	X-ray min		0.92	Stewart, 1989
Cp	3.0	United	Dv	X-ray		1.24	Stewart, 1989
Cp	3.0	United	Dv	X-ray min		0.85	Stewart, 1989
Dv	3.0	ALL	Cp	X-ray		0.94	Stewart, 1989
Dv	3.0	ALL	Cp	X-ray min		0.54	Stewart, 1989
Cp	3.0	ALL	Dv	X-ray		0.54	Stewart, 1989
Cp	3.0	ALL	Dv	X-ray min		0.24	Stewart, 1989
Pf	3.0	United	Cp	X-ray	2.01	1.41	This work
Pf	3.0	United	Cp	X-ray min	2.10	1.36	This work
Pf	3.0	United	Dg	X-ray	1.79	1.16	This work
Pf	3.0	United	Dg	X-ray min	1.91	1.15	This work
Pf	3.0	United	Dv	X-ray	2.46	1.95	This work
Pf	3.0	United	Dv	X-ray min	2.41	1.96	This work

^a X-ray min = minimized X-ray crystal structure.^b In previous work using AMBER to calculate main-chain RMS values only C1, C α , and N atoms were used. In this work using SYBYL, both the carbonyl oxygen, O, and the hydrogen atom on the nitrogen, HN, are included. This difference would lower the MC RMS values on average about 0.1 Å for the Pf comparisons.

using the same techniques employed in this study (see Table 8). The current results also support this conclusion and suggest that the approach used is a reasonable way to model rubredoxins. When comparing the modeled structures with the Pf X-ray structure from Day et al. (1992), it is evident that the RMS differences are slightly greater than those found in the previous test, but comparable to those obtained using AMBER 3.0 and the united-atom parameters (Table 8). However, this study did not use the all-atom force field, which, as shown in Table 8, gave better results in the previous study (Stewart, 1989).

Comparing the Pf rubredoxin structures obtained from the three separate efforts shows that there is a similar level of difference in atomic positions between the X-ray crystal structure, the NMR solution-state structures, and the homology models. The NMR vs. X-ray comparison (2.05 ± 0.38 Å RMS for all atoms) and the average of the comparisons of the homology models with the X-ray structure (2.09 ± 0.34 Å RMS) indicate a similar level of difference. Correspondingly, the average of the RMS differences between the three homology models (1.99 ± 0.24 Å RMS) is comparable. Note, the RMS differences reported by Blake et al. (1992a) are smaller than those given in Table 7; their RMS values include only the main-chain N, C1, and C α atoms for the first 51 residues. The homology models are closer to the X-ray structure (average 2.09 Å RMS) than to the NMR structures (average 2.42 Å RMS), which might be reasonable because they were themselves derived from X-ray crystal structures.

What should be the scale of the comparison? Previ-

ously, we argued (Stewart et al., 1987; Wampler et al., 1990) that by several criteria a 1-Å RMS difference in structures (experimental vs. model) might be an acceptable criterion for a successful modeling result. Recent results (Chandrasekhar et al., 1992) comparing three separately determined X-ray structures of interleukin-1 β , a 153-residue β -barrel protein, seem to support this criterion. These workers have compared the three separately determined X-ray structures and the mean solution structure from the NMR study. Their results (Table 9) give some feeling for the scale of differences that might be expected in the current study.

Table 9. Comparison of experimental interleukin-1 β structures^a

	Atoms	RMS difference in Å		
		X-ray #2	X-ray #3	NMR
X-ray #1	MC	0.355	0.317	0.894
	ALL	1.088	0.844	1.501
X-ray #2	MC		0.366	0.878
	ALL		1.164	1.619
X-ray #3	MC			0.874
	ALL			1.548

^a Data from Chandrasekhar et al. (1992). Structures from Finzel et al. (1989) (X-ray #1), Priestle et al. (1989) (X-ray #2), Veerapandian et al. (1991) (X-ray #3), and Clore et al. (1991) (NMR).

Obviously, minimization itself changes the X-ray structure. The degree of such change has, in fact, been used as a criterion of the quality of the molecular mechanics function and its parameters. Because the function and parameters are empirically determined based on structural measurements, this criterion seems reasonable. The caveat is that the "perfect" function would be equivalent to "no operation," i.e., writing the input structure to the output without change. Hall and Pavitt (1984) tested a number of molecular mechanics energy functions and parameter sets in minimizing three cyclic hexapeptides. In this test the AMBER energy function and the Kollman parameters set (ca. 1984) gave the least difference relative to the crystal structures. Whitlow and Teeter (1986) investigated variations of the selections of the AMBER program in minimizing the very high-resolution (0.945 Å) crystal structure of the protein crambin. Their best results (lowest RMS change) without including any environmental interactions were with a reduced distance-dependent dielectric model (dielectric constant of 4.0 multiplied by the separation distance) and a cutoff of 6–10 Å in nonbonded interaction calculations (RMS ~ 0.25 Å for all heavy atoms). Inclusion of interactions with neighboring molecules in a crystal environment or solvent molecules gave somewhat larger changes. With a wide range of program selections and environment representations, the RMS change in crambin structure on minimization varied from around 0.2 Å to 1.2 Å. It should be pointed out, however, that reducing electrostatic effects by raising the dielectric constant might well be expected to give less change because the largest single component in the typical AMBER energy evaluation is due to the electrostatic terms. Reducing the effect of this term then considerably alters the calculation in favor of bonded interactions and deep, narrow potential wells that favor standard bond distances, angles, and dihedrals. The selections used in this work (dielectric constant 1.0, distance dependence, 9 Å nonbonded cutoff) gave changes in the crambin structure of 0.4 Å comparable to the changes seen with explicit solvation (0.3 Å with a dielectric constant of 1 and a 10-Å cutoff). Minimizations of the mesophilic rubredoxin X-ray structures in AMBER 3.0 using these selections previously gave main-chain RMS changes on the order of 0.9 Å with the united atom parameters and 0.5–0.7 Å with all-atom parameters (Stewart, 1989). In this work, minimization of the Pf structure gave a main-chain RMS difference of 0.99 Å (Table 6).

The results reported here comparing the X-ray structure and the models, the X-ray and NMR structures, and the models and NMR structures are all larger than 1 Å RMS. Figure 3 shows that the largest contributions to these larger than expected RMS values are due to positional variations at the ends of the structures and at the extremities of the loops. The reasons for these differences may be attributed to inaccuracies in the AMBER energy function or the parameters used, lack of explicit solvent,

actual flexibility in the Pf rubredoxin structure, a structural difference induced by substitution of zinc for the NMR studies, or some combination of these factors. The largest potential error in application of the energy function is in the parameters for the iron center and the representation of the ligand-iron interactions (see Merz et al. [1991] for discussion). The extensive study of Whitlow and Teeter (1986) varying minimization selections and environment does not indicate that explicit solvation has much effect on minimization results. The molecular dynamics study reported in the accompanying paper (Bradley et al., 1993) does not indicate more flexibility in the Pf rubredoxin models. Of course, the denaturation/renaturation procedure used to substitute the metal (Blake et al., 1991) could also result in changes affecting comparisons to the NMR structures.

The homology modeling method used in this work does not change the core of the structure (i.e., the common atoms) very much. This is because the procedure (see Materials and methods) is controlled with constraints so that conflicts arising from the residue substitutions are relieved by local movements, not global changes. Thus, the common atoms in the homology models remain close to their parent structures (average 1.19 Å RMS). This difference is comparable to the difference in core, common atoms when comparing the different X-ray structures (average 1.01 Å RMS), or their minimized structures (average 1.20 Å RMS). Because most of the substitutions involve periphery atoms, extensions of the side chains, and the C-termini (in models built from Dg and Dv), a considerable portion of the total difference resides in the positioning of these periphery atoms during the process.

Detailed examination of the structures looking at residue-by-residue changes and AMBER energy components does not reveal any simple explanation for the difference in thermostability or the similarity in global parameters of the models (Table 4). The position differences are larger at the extremities of both individual residues (not shown) and the folded chains as indicated by the data of Figure 3. Yet, it is clear that the changes in sequence generate models that have very similar dynamic behavior as shown in the accompanying paper (Bradley et al., 1993).

Alignment of the rubredoxin sequences, scoring using the PAM 250 matrix, and composition differences all point to the Cp structure as most related to the Pf structure (this work and Blake et al., 1991). Indeed, this conclusion is reinforced by the results in Table 4, which indicate that the molecular size and AMBER energy of the Pf models are nearer to the minimized Cp X-ray structure than to the other parent proteins. If the AMBER energy is an indication of stability, the predicted order (greatest to least) of rubredoxin stability is Pf, Cp, Dg, and, finally, Dv. The consistently lower AMBER energy values found for the Pf models as compared to the minimized X-ray structures seems to be a global effect. Detailed examination of the data on a per-residue basis indicates that

differences are due to many small changes. For more discussion of the detailed interactions see the accompanying paper (Bradley et al., 1993).

Materials and methods

Computational resources

Molecular modeling and some computations were carried out using a Silicon Graphics 4D/20 Iris graphics workstation along with the computational resources of the Biological Sciences Sequence/Structure Computational Facility and those provided by University Computing and Network Services, both at the University of Georgia. The software utilized has been previously described (Stewart et al., 1987; Stewart, 1989; Wampler et al., 1990). Molecular mechanics calculations were performed using the AMBER 3.0 suite of programs (Singh et al., 1986). Molecular modeling, fitting, and general manipulations of the structures were carried out using the program SYBYL (Tripos Associates, Inc., St. Louis, Missouri). Some data manipulation and statistical analyses were performed using the program SPECOS SA (Wampler, 1991) and an IBM PC compatible computer system.

Alignment and scoring

The four known X-ray crystal structures (see below) were displayed on the molecular graphics workstation and aligned in three dimensions until the main-chain atoms were overlaid. The sequence alignment was then read from the structures. The sequence of Pf rubredoxin was visually aligned with the sequences from structural superposition so that a single leading gap placed the residues in homologous register as shown in Table 1. The resulting alignment of these five sequences is identical to that previously reported (Blake et al., 1991). Two-way sequence alignments using the Needleman-Wunsch algorithm with the Schwartz and Dayhoff (1978) PAM 250 scoring matrix produced the same sequence alignment (E.A. Bradley, unpubl.). Additionally, the alignments were analyzed to obtain a per-residue score based on the PAM 250 matrix. This matrix of amino acid substitution scores was developed based on natural mutation probabilities. Because the alignments came from known structures, gaps and nonaligned end residues were scored as if they were purely random occurrences. To gain insight into the scale of homology per residue, alignment of each sequence with itself was calculated, and 500 randomized sequences of the same composition were also scored.

Homology modeling

The procedures of homology modeling previously described (Stewart et al., 1987) involved two simple steps:

(1) the homolog with known structure was "mutated" into the unknown by editing the molecular structure to substitute each differing amino acid residue, and (2) the resulting "mutant" structure, consisting primarily of atoms positioned as in the parent molecule, was subjected to energy minimization to relieve strain introduced by the editing process. In our previous homology modeling efforts (Stewart et al., 1987; Stewart, 1989), step 1 was carried out by editing the structure file to remove all atoms not common to the two structures. The names of the substituted residues were changed accordingly, and then the AMBER EDIT module was used to add missing atoms. In this study, step 1 was carried out using the "mutate" and "excise" functions of the molecular modeling program SYBYL. The three X-ray crystal structures used as "parent structures" were obtained from the Brookhaven Protein Data Bank (Bernstein et al., 1977; Abola et al., 1987). These structures were *Clostridium pasteurianum* (Cp: entry 5rxn, version of April 1985; Watenpaugh et al., 1980) *Desulfovibrio gigas* (Dg: entry 1rdg, version of April 1989; Frey et al., 1987), and *Desulfovibrio vulgaris* (Dv: entry 3rxn, version of September 1983; Adman et al., 1991).

The second step was carried out using the AMBER 3.0 minimization module in three phases. The initial phase was an all-atom constrained minimization that allowed bad contacts due to substitutions of residues to be relieved with minimum change in the overall molecular conformation. For this study, the initial constraint weights of 100, 50, and 10 were applied in descending order as convergence was reached. The second phase involved constraint of atoms comprising the backbone with constraint weights of 100, 50, and 10 applied in descending order as convergence was reached. The final phase was a full minimization of the structure with no constraints. For both types of minimizations, the following selections and options of the AMBER program were used: united atom force field, distance-dependent dielectric constant, 9.5-Å cutoff for nonbonded calculations, and conjugate gradient minimization until the RMS gradient in the energy was less than 0.001 kcal/mol·Å.

Standard AMBER 3.0 United Atom parameters were used for all calculations. The nonstandard parameters added for the rubredoxin iron center are given in Table 10. These parameters are from the work of Stewart (1989). The iron charge was empirically adjusted in cycles of minimization of known X-ray structures (Stewart, 1989).

Static analysis

Static analysis of the minimized, X-ray, and NMR structures was performed using the Static Analysis module of AMBER 3.0 and the fitting functions of SYBYL 5.5. Energy decomposition and calculation of the radius of gyration were performed with AMBER. RMS position differences were calculated using SYBYL to fit the struc-

Table 10. Nonstandard AMBER parameters used in this study^a

Atom	Mass (g/mol)	Description	
C2	14.03	United-atom sp3 carbon with 2 hydrogens	
LP	12.00	Sulfur lone pair	
S	32.06	Sulfur in ligating cysteine	
FE	55.85	Iron cation	
Bond	K_r (kcal/Å ²)	r_{eq} (Å)	
S-FE	200.00	2.280	
Angle	K_θ (kcal/radian ²)	θ_{eq} (degrees)	
C2-S-FE	50.00	116.9	
LP-S-FE	0.00	110.0	
S-FE-S	5.00	110.0	
VDW	R^* (Å)	ϵ^* (kcal/mol)	
FE	0.85	0.10	
Torsions	$V_n/2$ (kcal/mol)	γ (degrees)	n (periodicity)
X-FE-S-X	0.50	0	-6
X-FE-S-X	0.50	180	6
Electrostatics	q (kcal/mol)		
FE	1.5		

^a AMBER parameters as defined by Weiner et al. (1984).

tures after selecting the atoms of interest, thus the values reported are weighted RMS values. For comparisons with the NMR structures the individual X-ray and modeled structures were compared after fitting, calculating either the statistics of the distribution of the individual RMS values (Table 7) or the combined, global RMS differences from the sum of squares of differences of all 40 comparisons (Fig. 3).

Acknowledgments

This work was funded in part by grants from the following: National Institutes of Health (GM41482), the North Atlantic Treaty Organization (0404/88), the National Science Foundation (DIR-9014281 and BCS-9011583), and the Office of Naval Research (N00014-90-J-1894). Special thanks are due to Dr. Walter McRae, Assistant Vice President for Computing, for support funds and computer time on the high-performance machines, to Doug C. Rees for kindly providing the Pf rubredoxin X-ray crystal coordinates, and to Michael F. Summers for providing the NMR solution-state coordinates.

References

Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F., & Weng, J. (1987). Protein Data Bank. In *Crystallographic Databases—Information Content, Software Systems, Scientific Applications* (Allen,

- F.H., Bergerhoff, G., & Sievers, R., Eds.), pp. 107–132. Commission of the International Union of Crystallography, Bonn/Cambridge/Chester.
- Adams, M.W.W. & Kelly, R.M., Eds. (1992). *Biocatalysis at Extreme Temperatures: Enzyme Systems Near and Above 100 °C*. American Chemical Society, Washington, D.C.
- Adman, E.T., Sieker, L.C., & Jensen, L.H. (1991). Structure of rubredoxin from *Desulfovibrio vulgaris* at 1.5 Å resolution. *J. Mol. Biol.* 217, 337–352.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535–542.
- Blake, P.R., Day, M.W., Hsu, B.T., Joshua-Tor, L., Park, J.-B., Zhou, Z.H., Hare, D.R., Adams, M.W.W., Rees, D.C., & Summers, M.F. (1992a). Comparison of the X-ray structure of native rubredoxin from *Pyrococcus furiosus* with the NMR structure of the zinc-substituted protein. *Protein Sci.* 1, 1522–1525.
- Blake, P.R., Park, J.-B., Bryant, F.O., Shigetoshi, A., Magnuson, J.K., Eculston, E., Howard, J.B., Summers, M.F., & Adams, M.W.W. (1991). Determinants of protein hyperthermostability: 1. Purification, amino acid sequence and secondary structure by NMR of the rubredoxin from the hyperthermophilic archaeobacterium, *Pyrococcus furiosus*. *Biochemistry* 30, 10885–10895.
- Blake, P.R., Park, J.-B., Zhou, Z.H., Hare, D.R., Adams, M.W.W., & Summers, M.F. (1992b). Solution-state structure by NMR of zinc-substituted rubredoxin from the marine hyperthermophilic archaeobacterium *Pyrococcus furiosus*. *Protein Sci.* 1, 1508–1521.
- Bradley, E.A., Stewart, D.E., Adams, M.W.W., & Wampler, J.E. (1993). Investigations of the thermostability of rubredoxin models using molecular dynamics simulations. *Protein Sci.* 2, 650–665.
- Bull, H.B. & Breese, K. (1973). Thermal stability of proteins. *Arch. Biochem. Biophys.* 158, 681–686.
- Chandrasekhar, I., Clore, G.M., Szabo, A., Gronenborn, A.M., & Brooks, B.R. (1992). A 500 ps molecular dynamics simulation study of interleukin-1β in water, correlation with nuclear magnetic resonance spectroscopy and crystallography. *J. Mol. Biol.* 226, 239–250.
- Clare, G.M., Wingfield, P.T., & Gronenborn, A.M. (1991). High-resolution three-dimensional structure of interleukin 1β in solution by three- and four-dimensional nuclear magnetic resonance spectroscopy. *Biochemistry* 30, 2315–2323.
- Costantino, H.R., Brown, S.H., & Kelly, R.M. (1990). Purification and characterization of an α-glucosidase from a hyperthermophilic archaeobacterium *Pyrococcus furiosus* exhibits a temperature optimum of 105 to 115 °C. *J. Bacteriol.* 172, 3654–3660.
- Dauter, Z., Sieker, L.C., & Wilson, K.S. (1992). Refinement of rubredoxin from *Desulfovibrio vulgaris* at 1.0 Å with and without restraints. *Acta Crystallogr.* B48, 42–59.
- Day, M.W., Hsu, B.T., Joshua-Tor, L., Park, J.-B., Zhou, Z.H., Adams, M.W.W., & Rees, D.C. (1992). X-ray crystal structures of the oxidized and reduced forms of the rubredoxin from the marine hyperthermophilic archaeobacterium *Pyrococcus furiosus*. *Protein Sci.* 1, 1494–1507.
- Finzel, B.C., Clancy, L.L., Holland, D.R., Muchmore, S.W., Watenpugh, K.D., & Einspahr, H.M. (1989). Crystal structure of recombinant human interleukin-1β at 2.0 Å resolution. *J. Mol. Biol.* 209, 779–791.
- Frey, M., Sieker, L., Payan, F., Haser, R., Bruschi, M., Pepe, G., & LeGall, J. (1987). Rubredoxin from *Desulfovibrio gigas*. A molecular model of the oxidized form at 1.4 angstroms resolution. *J. Mol. Biol.* 197, 525–541.
- Fontana, A. (1991). How nature engineers protein (thermo)stability. In *Life Under Extreme Conditions* (di Prisco, G., Ed.), pp. 89–113. Springer-Verlag, Berlin.
- Hall, D. & Pavitt, N. (1984). An appraisal of molecular force fields for the representation of polypeptides. *J. Comp. Chem.* 5, 441–450.
- Lim, W.K.L., Brouillette, C., & Hardman, J.K. (1992). Thermal stabilities of mutant *Escherichia coli* tryptophan synthase α subunits. *Arch. Biochem. Biophys.* 292, 34–41.
- Lovenberg, W. & Sobel, B.E. (1965). Rubredoxin: A new electron transfer protein from *Clostridium pasteurianum*. *Proc. Natl. Acad. Sci. USA* 54, 193–199.
- Merz, K.M., Murcko, M.A., & Kollman, P.A. (1991). Inhibition of carbonic anhydrase. *J. Am. Chem. Soc.* 113, 4484–4490.

- Papavassiliou, P. & Hatchikian, E.C. (1985). Isolation and characterization of a rubredoxin and a two (4Fe-4S) ferredoxin from *Thermodesulfobacterium commune*. *Biochim. Biophys. Acta* **810**, 1–11.
- Priestle, J.P., Schar, H.-P., & Grutter, M.G. (1989). Crystallographic refinement of interleukin-1 β at 2.0 Å resolution. *Proc. Natl. Acad. Sci. USA* **86**, 9667–9671.
- Schwartz, R.M. & Dayhoff, M.O. (1978). Matrices for detecting distant relationships. In *Atlas of Protein Sequence and Structure Supplement 3* (Dayhoff, M.O., Ed.), pp. 353–358. National Biomedical Research Foundation, Washington, D.C.
- Singh, U.C., Wiener, P.K., Caldwell, J.W., & Kollman, P.A. (1986). AMBER (USCF), Version 3.0. Department of Pharmaceutical Chemistry, University of California, California.
- Stellwagen, E. & Wilgus, H. (1978). Relationship of protein thermostability to accessible surface area. *Nature* **275**, 342–343.
- Stenkamp, R.E., Sieker, L.C., & Jensen, L.H. (1990). The structure of rubredoxin from *Desulfovibrio desulfuricans* strain 27774 at 1.5 Å resolution. *Proteins Struct. Funct. Genet.* **8**, 352–364.
- Stewart, D.E. (1989). The structure, interactions, and dynamics of electron transport proteins from *Desulfovibrio*: A molecular modeling and computational study. Ph.D. Dissertation, University of Georgia, Athens, Georgia.
- Stewart, D.E. & Wampler, J.E. (1991). Molecular dynamics simulations of the cytochrome c_3 -rubredoxin complex from *Desulfovibrio vulgaris*. *Proteins Struct. Funct. Genet.* **11**, 142–152.
- Stewart, D.E., Weiner, P.K., & Wampler, J.E. (1987). Prediction of the structure of proteins using related structures, energy minimization and computer graphics. *J. Mol. Graph.* **5**, 137–144.
- Veerapandian, B., Gilliland, G.L., Raag, R., Svensson, A.L., Masui, Y., Hirai, Y., & Poulos, T.L. (1991). Functional implications of interleukin-1 β based on the three-dimensional structure. *Proteins Struct. Funct. Genet.* **12**, 10–23.
- Wampler, J.E. (1991). Specos SA: A computer program for managing, graphing and manipulating laboratory data. *Anal. Instrum.* **19**, 203–230.
- Wampler, J.E., Bradley, E.A., Adams, M.W.W., & Stewart, D.E. (1992). Computational approaches to modeling and analyzing thermostability in proteins. In *Biocatalysis at Extreme Temperatures: Enzyme Systems Near and Above 100°C* (Adams, M.W.W. & Kelly, R.M., Eds.), pp. 153–173. American Chemical Society, Washington, D.C.
- Wampler, J.E., Stewart, D.E., & Gallion, S.L. (1990). Molecular dynamics simulations of proteins and protein-protein complexes. In *Computer Simulation Studies in Condensed Matter Physics II* (Landau, D.P., Mon, K.K., & Schüttler, H.-B., Eds.), pp. 68–84. Springer-Verlag, New York.
- Watenpaugh, K.D., Sieker, L.C., & Jensen, L.H. (1979). The structure of rubredoxin at 1.2 Å resolution. *J. Mol. Biol.* **131**, 509–522.
- Watenpaugh, K.D., Sieker, L.C., & Jensen, L.H. (1980). Crystallographic refinement of rubredoxin at 1.2 Å resolution. *J. Mol. Biol.* **138**, 615–633.
- Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 165–174.
- Whitlow, M. & Teeter, M.M. (1986). An empirical examination of potential energy minimization using the well-determined structure of the protein crambin. *J. Am. Chem. Soc.* **108**, 7163–7172.
- Zuber, H. (1978). Comparative studies of thermophilic and mesophilic enzymes: Objectives, problems, results. In *Biochemistry of Thermophily* (Friedman, S.M., Ed.), pp. 267–285. Academic Press, New York.
- Zulli, F., Schneider, R., Urfer, R., & Zuber, H. (1991). Engineering thermostability and activity of lactate dehydrogenases from bacilli. *Biol. Chem. Hoppe-Seyler* **372**, 363–372.
- Zwickl, P., Fabry, S., Bogedain, C., Haas, A., & Hensel, R. (1990). Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeobacterium *Pyrococcus woesei*: Characterization of the enzyme, cloning and sequencing of the gene, and expression in *Escherichia coli*. *J. Bacteriol.* **172**, 4329–4338.