J|A|C|S
ARTICLES
Published on Web 03/26/2002

O_2 Penetration and Proton Burial Depth in Proteins: Applicability to Fold Family Recognition

Griselda Hernández,[†] Ching-Ling Teng,[‡] Robert G. Bryant,[‡] and David M. LeMaster*,[†]

Contribution from the Wadsworth Center and Department of Biomedical Sciences, University at Albany—SUNY, Empire State Plaza, Albany, New York 12201-0509, and Department of Chemistry, University of Virginia, Charlottesville, Virginia 22901

Received October 20, 2001

Abstract: Paramagnetically induced relaxation effects of O_2 and the nitroxide 4-hydroxy TEMPO were measured for the amide protons of perdeuterated rubredoxin from the hyperthermophilic archaeon *Pyrococcus furiosus* and the mesophilic bacterium *Clostridium pasteurianum*. For both O_2 and the impermeant nitroxide, the induced relaxation at the static solvent inaccessible amide sites is dominated by long-range interactions with the paramagnetic species in the bulk aqueous phase. The upper bound of O_2 solubility in the internal matrix of the rubredoxins is one-tenth that of the bulk aqueous phase. Furthermore, the difference between the oxygen solubilities inside the two rubredoxins is at most 1% that of bulk water O_2 solubility, suggesting that there are only modest differences in this measure of fluidity for the mesophile vs hyperthermophile protein interiors. Calculations based on the assumption of a paramagnet uniformly distributed on the protein exterior yield accurate predictions at nearly all amide sites for the minimum relaxation value observed from either the O_2 or nitroxide data. Model calculations indicate that the readily obtained paramagnetically induced relaxation effects should prove effective in recognition of structural homology for proteins that are too widely diverged for sequence-based recognition.

Introduction

The myoglobin X-ray structure reveals no unhindered path for O_2 diffusion to the heme iron, thus indicating the necessity of conformational dynamics to achieve biological function. Molecular simulations[1,2] have predicted structural fluctuations of a scale suitable for O_2 diffusion to and from the heme site. Largely drawing upon the kinetics of CO geminate recombination to the myoglobin iron atom,[3,4] extensive efforts have been made to experimentally characterize the process of O_2 binding to the heme. In another approach to characterizing protein$-O_2$ interactions, Weber and co-workers[1] utilized the fluorescence quenching ability of the paramagnetic O_2 triplet to probe the apparent tryptophan accessibility for a series of proteins. These experiments suggested the presence of rapid diffusion of O_2 into the protein interior, supporting the view of a more fluidlike structure than that inferred from the crystallographic models. However, the detailed interpretation of such dynamic quenching experiments has remained controversial.[6,7]

Most proteins have stable internal structural cavities large enough to accommodate a water or an O_2 molecule. The majority of these cavities do not contain well-ordered water molecules.[8,9] For both wild type[10] and genetically engineered variants,[11] stable cavities have been shown to efficiently bind nonpolar organic molecules and noble gases. NMR exchange measurements have indicated that the residence times for water[12] as well as various nonpolar small molecules[13−15] bound to internal protein cavities lie in the range between 10^{-8} and 10^{-3} s. In the case of myoglobin, four connected internal cavities have been shown to bind xenon.[16] Subsequent demonstration[17,18] that photodissociated CO occupies at least some of these xenon

* To whom correspondence should be addressed. E-mail: lemaster@wadsworth.org.
[†] University at Albany−SUNY.
[‡] University of Virginia.

(1) Case, D. A.; Karplus, M. *J. Mol. Biol.* **1979**, *132*, 343−368.
(2) Karplus, M.; McCammon, J. A. *Sci. Am.* **1986**, *254 (Apr)*, 42−51.
(3) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H. *Biochemistry* **1975**, *14*, 5355−5373.
(4) Beece, D.; Eisenstein, L.; Frauenfelder, H.; Good, D.; Marden, M. C.; Reinisch, L.; Reynolds, A. H.; Sorensen, L. B.; Yue, K. T. *Biochemistry* **1980**, *19*, 5147−5157.
(5) Lakowicz, J. R.; Weber, G. *Biochemistry* **1973**, *12*, 4171−4179.
(6) Calhoun, D. B.; Englander, S. W.; Wright, W. W.; Vanderkooi, J. M. *Biochemistry* **1988**, *27*, 8466−8474.
(7) Cioni, P.; Strambini, G. B. *J. Mol. Biol.* **1999**, *291*, 955−964.
(8) Hubbard, S. J.; Gross, K. H.; Argos, P. *Protein Eng.* **1994**, *7*, 613−626.
(9) Williams, M. A.; Goodfellow, J. M.; Thorton, J. M. *Protein Sci.* **1994**, *3*, 1224−1235.
(10) Schoenborn, B. P.; Watson, H. C.; Kendrew, J. C. *Nature* **1965**, *207*, 28−30.
(11) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. *Nature* **1992**, *355*, 371−373.
(12) Otting, G.; Liepinsh, E.; Wuthrich, K. *Science* **1991**, *254*, 974−980.
(13) Feher, V. A.; Baldwin, E. P.; Dahlquist, F. W. *Nat. Struct. Biol.* **1996**, *3*, 516−521.
(14) Otting, G.; Liepinsh, E.; Halle, B.; Frey, U. *Nat. Struct. Biol.* **1997**, *4*, 396−404.
(15) Mulder, F. A. A.; Hon, B.; Muhandiram, D. R.; Dahlquist, F. W.; Kay, L. E. *Biochemistry* **2000**, *39*, 12614−12622.
(16) Tilton, R. F.; Kuntz, I. D.; Petsko, G. A. *Biochemistry* **1984**, *23*, 2849−2857.
(17) Chu, K.; Vojtchovsky, J.; McMahon, B. H.; Sweet, R. M.; Berendzen, J.; Schlichting, I. *Nature* **2000**, *403*, 921−923.
(18) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. *Nature* **2000**, *404*, 205−208.

binding cavities supports the argument that transfer between these connected cavities explains the multiexponential kinetics observed in CO geminate recombination experiments. In a similar vein, a cavity mutant of T4 lysozyme, which is able to internally bind various benzene derivatives, appears to open that cavity to the bulk solvent phase in a single concerted transition.[19]

Is the migration of O_2 and other small nonpolar molecules through the protein interior restricted to transient conformational fluctuations which open a stable cavity to either bulk solvent or to another connected internal cavity? To what degree can dioxygen diffuse through the protein interior in a more fluidlike process as envisioned in the mobile defect model.[20,21] A directly related question involves the determination of the average solubility of O_2 in the protein interior in the absence of specific binding to a preformed cavity, as the efficiency of diffusion is proportionate to the local concentration.

Recently, we[22] demonstrated that for the most thermostable protein characterized to date, the rubredoxin from the hyperthermophilic archaeon *Pyrococcus furiosus*,[23] the amide protons throughout its structure become accessible to water and hydroxide catalyst in less than a second under native conditions at room temperature. At the same time, others of us[24] showed that O_2 at 10 atm yielded paramagnetically induced ^1H T_1 relaxation not only for protons on the external surface of ribonuclease but for many of the static solvent inaccessible protons as well. The paramagnetically induced relaxation of the buried protons arise from either O_2 penetration, longer range relaxation effects from O_2 in the bulk phase, or by a spin diffusion process in which magnetization from efficiently relaxed protons on the protein surface is passively transferred to the buried nuclei.

Reasons for considering the indirect spin diffusion relaxation mechanism to be of modest significance have been offered previously.[24] In the present study such spin diffusion pathways are directly suppressed by the use of perdeuterated protein samples which bear ^1H nuclei only in the amide positions. Distinguishing O_2 penetration effects from bulk-phase paramagnetic contributions is achieved by comparison to the relaxation effects of a comparatively large water soluble nitroxide. If selective binding interactions can be adequately recognized and the O_2 and nitroxide relaxation effects suitably normalized, O_2 penetration will yield enhanced paramagnetically induced relaxation relative to the impermeant nitroxide. On the other hand, deconvolution of the bulk-phase paramagnetic contribution can potentially lead to predictable relaxation effects when compared to a known tertiary protein structure.

Paramagnetically induced relaxation measurements were carried out on the aforementioned rubredoxin from *Pyrococcus furiosus* (*Pf*). The 1.1 Å X-ray structure determination of this protein[25,26] indicates no internal cavities of sufficient size to bind dioxygen. Hence, any O_2 penetration will reflect transient cavity expansion. Relaxation measurements were carried out in parallel on the structurally homologous rubredoxin from the mesophilic bacterium *Clostridium pasteurianum* (*Cp)* for which there is an equally high-resolution X-ray structure determination.[27] These rubredoxins differ at 23 of 54 residues, yet the root mean square deviation for their backbone C^α positions is only 0.6 Å. This comparative study serves two distinct roles. By considering two similar proteins, the sensitivity of these paramagnetic relaxation measurements to the detailed aspects of small structural differences can be monitored. In addition, a dominant paradigm for explaining the extreme thermostability of the proteins from the hyperthermophilic organisms invokes a more rigid, more crystalline-like interior for the hyperthermophile protein when compared under similar conditions to the homologous protein from a mesophile.[28,29] If the interior of the rubredoxin from the room temperature *Clostridium pasteurianum* is in fact more liquidlike, then this property could be expected to be manifest in a higher apparent solubility of dioxygen.

Materials and Methods

Preparation of Isotopically Labeled Rubredoxins. Synthetic codon usage optimized genes for both *Clostridium pasteuranium*[30,31] and *Pyrococcus furiosus* rubredoxin,[31] cloned into the pT7−7 vector and transformed in BL21(DE3), were used for protein expression (both plasmids kindly provided by M. K. Eidsness, University of Georgia). The expression strains were grown in M9 glucose ^2H_2O minimal medium in which 1 g/L $^{15}NH_4Cl$ was substituted and the RPMI 1640 vitamins mixture (Sigma-Aldrich) added. Purification of both rubredoxin samples followed published procedures.[30,32,33] The N-terminal methionine containing form of *Pf* rubredoxin was utilized for these studies. This form offers increased sequence homology with the *Cp* protein. Furthermore, during expression in *Escherichia coli* (*E. coli*) the terminal methionine form was produced in higher yields than the mature wild type form due to inefficient processing of the N-terminus in the bacterial system.[31] Mass spectral analysis indicated an average deuteration level of >95% for both proteins.

NMR Spectroscopy. Samples of 3 mM ^2H,^{15}N *Cp* rubredoxin and 3 mM ^2H,^{15}N *Pf* rubredoxin were both dialyzed against two 30 mL aliquots of the same 7% ^2H_2O buffer solutions composed of 0.1 M sodium chloride, 0.02 M sodium phosphate (pH 6.3), similar to the conditions previously used for the resonance assignment of these proteins.[33−35] The samples were placed in pressure/vacuum valve NMR tubes (Wilmad Glass, No. 524-PV-7). Air was purged from the samples by multiple cycles of freeze−thaw evacuations. The samples were then equilibrated against 10 atm of either O_2 or N_2. For the subsequent nitroxide experiments, 4-hydroxy TEMPO was added to each O_2 sample to a nominal final concentration of 5 mM and then purged and equilibrated under an N_2 atmosphere. Data collection was carried out at 25 °C.

As discussed below, the observed paramagnetically induced relaxation values can be expected to be approximately independent of the

(19) Skrynnikov, N. R.; Mulder, F. A. A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 4556−4566.
(20) Lumry, R.; Rosenberg, A. *Colloq. Int. CNRS* **1975**, *246*, 53−61.
(21) Pain, R. H. *Nature* **1987**, *326*, 247.
(22) Hernandez, G.; Francis E. Jenney, J.; Adams, M. W. W.; LeMaster, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3166−3170.
(23) Hiller, R.; Zhou, Z. H.; Adams, M. W. W.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11329−11332.
(24) Teng, C. L.; Bryant, R. G. *J. Am. Chem. Soc.* **2000**, *122*, 2667−2668.
(25) Day, M. W.; Hsu, B. T.; Joshua-Tor, L.; Park, J. B.; Zhou, Z. H.; Adams, M. W. W.; Rees, D. C. *Protein Sci.* **1992**, *1*, 1494−1507.
(26) Bau, R.; Rees, D. C.; Kurtz, D. M.; Scott, R. A.; Huang, H. S.; Adams, M. W. W.; Eidsness, M. K. *J. Biol. Inorg. Chem.* **1998**, *3*, 484−493.
(27) Watenpaugh, K. D.; Sieker, L. C.; Jensen, L. H. *J. Mol. Biol.* **1980**, *138*, 615−633.
(28) Jaenicke, R.; Zavodszky, P. *FEBS Lett.* **1990**, *268*, 344−349.
(29) Tang, K. E. S.; Dill, K. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 397−411.
(30) Eidsness, M. K.; O'Dell, S. E.; D. M. Kurtz, J.; Robson, R. L.; Scott, R. A. *Protein Eng.* **1992**, *5*, 367−371.
(31) Eidsness, M. K.; Richie, K. A.; Burden, A. E.; Kurtz, D. M.; Scott, R. A. *Biochemistry* **1997**, *36*, 10406−10413.
(32) Lovenberg, W.; Sobel, B. E. *Proc. Natl. Acad. Sci. U.S.A.* **1965**, *54*, 193−199.
(33) Hernandez, G.; LeMaster, D. M. *Biochemistry* **2001**, *40*.
(34) Blake, P. R.; Park, J. B.; Bryant, F. O.; Aono, S.; Magnuson, J. K.; Eccleston, E.; Howard, J. B.; Summers, M. F.; Adams, M. W. W. *Biochemistry* **1991**, *30*, 10885−10895.
(35) Richie, K. A.; Teng, Q.; Elkin, C. J.; D. M. Kurtz, J. *Protein Sci.* **1996**, *5*, 883−894.

protein molecular tumbling rate. Nevertheless, to directly verify that any differential relaxation effect was not a result of variations in the protein global tumbling rate, the mean of the 1.5 σ filtered[36] ^{15}N T_2 relaxation times[37] were determined to vary less than 1% among the samples. Note that the paramagnetic relaxation effect scales as γ_N^2 and thus is at least 100-fold smaller for the ^{15}N nuclei than for the ^1H nuclei. The average T_2 value of 0.226 s is consistent with the expected global correlation time of the monomeric rubredoxin. To ensure consistent ^1H T_1 relaxation measurements, a ^1H purge pulse[38] was placed at the beginning of the recycle delay of the $^1H-^{15}N$ TROSY experiments.[39,40] Relaxation delays of 0.4, 0.8, 1.2, 1.6, 2.0, and 2.4 s were used for the protein samples under N_2 atmosphere. To reflect the wider range of relaxation values, additional T_1 relaxation delays of 0.1, 0.2, 0.3, and 0.6 s were used for the O_2 and 4-hydroxy TEMPO samples. The protein relaxation values were normalized to the difference in the ^1H_2O T_1 values of O_2 vs N_2 (2.33 and 2.20 s^{-1} for *Cp* and *Pf*, respectively) and TEMPO vs N_2 (2.78 and 3.08 s^{-1} for *Cp* and *Pf*, respectively). The NMR data were transformed and peak integrations carried out using FELIX 2000 software (Molecular Simulations, Inc.). The ^1H T_1 relaxation data for the amides of both rubredoxins fit exponential decays with a mean root mean square deviation of 1.2%.

Modeling of Paramagnetic Relaxation Effects. Hydrogens were added to the X-ray coordinates of both *Cp* (5RXN) and Met-*Pf* (1BQ8) rubredoxins using the Builder function of InsightII (Molecular Simulations, Inc.). The coordinates of each rubredoxin structure were embedded in a cubic lattice with 0.1 Å grid spacing that extended at least 8.4 Å beyond all protein atoms. To determine the protein exterior, all lattice points within the van der Waals radius[41] + 1.4 Å of any protein atom were turned off. For each of the remaining lattice points, it was determined whether a face-centered path of "solvent" lattice points extended to the boundary of the cubic lattice. The lattice points which lacked such a connectivity were considered internal cavity sites and were turned off for the subsequent calculation. For each backbone amide hydrogen of the protein structure, $1/r^6$ was determined for each exterior lattice point and the result was summed over the entire lattice. This simple summation corresponds to an extended trapezoidal numerical integration. Since the integration interval extends to the boundary of the excluded volume, the values for the lattice points at the boundary need to be scaled by a factor of 0.5.[42] The points that are within half lattice spacing of the probe boundary were so scaled.

One estimate of the robustness of this numerical integration is the comparison of the lattice sums for different spacings. All of the solvent accessible amide hydrogens of the rubredoxins (and several other proteins that were examined) gave lattice sums for a 0.1 Å spacing that were within 3% of being 8-fold larger than the sum for a 0.2 Å spacing. Furthermore, in no case did the normalized lattice sums for the solvent inaccessible amide hydrogens with the 0.1 and 0.2 Å spacings differ by more than 0.01. A different test of numerical stability involves the shifting of the lattice by a half-spacing (0.05 Å) along each axis which resulted in variations of less than 0.5% in the individual summations. The lattice box boundary extended at least six times the probe radius beyond all protein atoms. To test the suitability of this cutoff distance, for each amide hydrogen the lattice sum was binned according to the distance between each lattice point and the nearest protein atom. The fraction of the total lattice sum arising from each 1.4 Å thick shell of lattice points was determined. Truncation of the summation at the lattice boundaries gave rise to an error of less than 0.5% in all cases as compared to a calculation extending the boundary an additional 5.6 Å in each direction. Solvent accessible amide hydrogens were defined as those possessing an exterior lattice point within 2.587 Å (the sum of the 1.4 Å probe, the 1.1 Å proton van der Waals radius, and half the lattice diagonal).

Results

Four matched NMR samples of 3 mM U-^2H,^{15}N labeled *Cp* and *Pf* rubredoxin were prepared under 10 atm of O_2 and 10 atm of N_2. The >95% average deuteration at the carbon bound positions combined with the short molecular rotational correlation time of these 6.5 kDa monomers can be expected to strongly suppress cross-relaxation effects between the residual ^1H nuclei. ^1H T_1 values were obtained by variation of the recycle delay period in a series of $^1H-^{15}N$ TROSY experiments.[39,40] Saturation of the ^1H_2O resonance during the recycle delay had no observable effect on the buried amide resonances considered below, thus indicating that transfer via chemical exchange is negligible. The difference between the ^1H amide relaxation rates, $1/T_1$, in the presence and absence of oxygen were normalized to the spin−lattice relaxation rate of the ^1H_2O resonance which was measured by inversion recovery with the probe detuned to suppress radiation dampening. These normalized relaxation values are independent of O_2 concentration at the level used here (~12 mM).

In no case did the O_2-induced relaxation of a protein amide exceed that of the ^1H_2O resonance, indicating an absence of significant selective binding of dioxygen in the vicinity of the nuclei monitored in this study. In Figure 1a is illustrated the normalized O_2-induced relaxation of the *Cp* (●) and *Pf* (◇) rubredoxins for all but one of the 28 main chain amide protons which are inaccessible to a 1.4 Å probe in both X-ray structures. The buried residue 4 was excluded due to resonance overlap with Asp 35 in the *Cp* rubredoxin spectra. A substantial paramagnetic relaxation effect is observed throughout the interior of both proteins. The data of Figure 1a yield an average value of 0.30 and 0.29 for *Cp* and *Pf* rubredoxins, respectively, which is fully 60% of the average O_2-induced relaxation observed for the static solvent accessible amides. With a few exceptions, the individual relaxation values are quite similar for both proteins.

On average, the solvent inaccessible amide hydrogens are only modestly less susceptible to O_2-induced paramagnetic relaxation than are the surface exposed sites which would appear suggestive of dioxygen permeation. To test whether these O_2-induced relaxation values do reflect penetration of O_2 into the protein interior, analogous measurements were carried out on samples containing the paramagnetic nitroxide 4-hydroxy TEMPO. Although the oxygen of the nitroxide group presents a radius of closest approach similar to that of dioxygen, the molecular volume of 4-hydroxy TEMPO is approximately 10 times larger. In contrast to the O_2 experiment, both rubredoxins have five solvent-exposed amides which exhibit TEMPO-induced relaxation values that are substantially higher than that of the ^1H_2O resonance, with normalized rates ranging from 1.2 to 2.0 indicative of a very weak binding interaction. Four of these residues (12, 25, 51, and 53) are elevated in both proteins. Due to spectral overlap, as noted above, the elevated relaxation value for Asp 35 of *Pf* rubredoxin could not be accurately

(36) Clore, G. M.; Driscoll, P. C.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 7387−7401.
(37) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984−6003.
(38) Muhandiram, D. R.; Xu, G. Y.; Kay, L. E. *J. Biomol. NMR* **1993**, *3*, 463−470.
(39) Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12366−12371.
(40) Weigelt, J. *J. Am. Chem. Soc.* **1998**, *120*, 10778−10779.
(41) Kuhn, L. A.; Swanson, C. A.; Pique, M. E.; Tainer, J. A.; Getzoff, E. D. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 536−547.
(42) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes*, 1st ed.; Cambridge University Press: Cambridge, U.K., 1989.

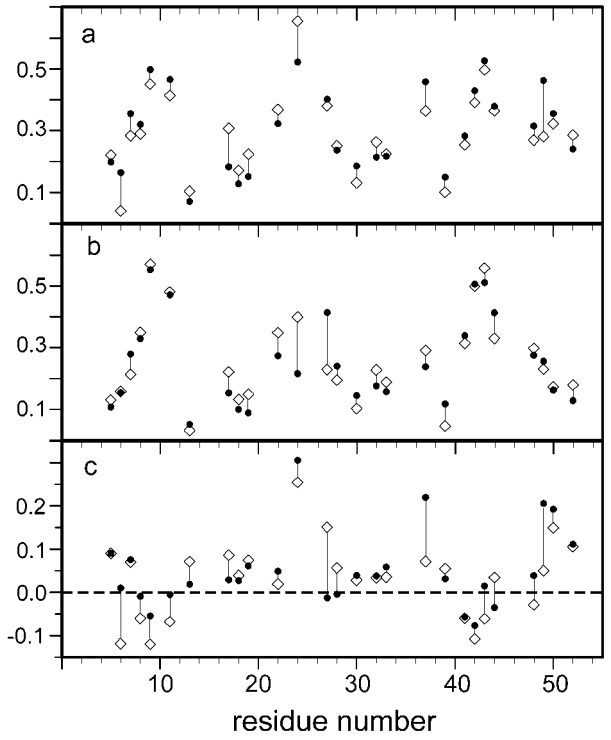

Figure 1. Normalized O_2- and 4-hydroxy-TEMPO-induced paramagnetic relaxation of static solvent inaccessible amide protons in *Cp* (●) and *Pf* (◇) rubredoxins. In panel a is shown the difference for both proteins in the amide ^1H spin−lattice relaxation rates under 10 atm O_2 vs 10 atm N_2, normalized to the corresponding difference in the solvent ^1H_2O spin−lattice relaxation rate. Data are shown only at sequence positions for which the amide proton is static solvent inaccessible in the X-ray structures of both proteins. Panel b presents the analogous data for the rubredoxin samples containing nominally 5 mM of the nitroxide 4-hydroxy-TEMPO. In panel c is plotted the difference between the data of panels a and b (i.e. O_2 − TEMPO).

compared to the corresponding resonance of *Cp* rubredoxin. The fifth elevated residue of *Cp* rubredoxin (Gly 45) has a proline in the homologous position of the *Pf* protein. The effects of these weak binding interactions appear to be highly localized as evidenced by the fact that in no case is the normalized TEMPO-induced relaxation rate of an amide sequentially adjacent to these binding sites more than 0.1 higher than its O_2-induced relaxation value.

As seen in Figure 1b, with the exception of residues 24 and 27 discussed below, the TEMPO-induced relaxation effects are very similar between the static inaccessible amides of the two proteins. More surprisingly, for the large majority of the residues, both proteins show a difference between the O_2-induced and TEMPO-induced relaxation of less than 0.1 (Figure 1c). Excluding the four residues for which this (O_2 − TEMPO) difference is greater than 0.15, the remaining 23 of 27 static inaccessible amides of *Cp* rubredoxin (●) have a mean O_2-induced relaxation value that is only 0.019 higher than the mean TEMPO-induced relaxation value. The analogous calculation for the hyperthermophile *Pf* rubredoxin (◇) yields a mean (O_2 − TEMPO) difference of 0.013 for 88% (24/27) of the buried amides.

The relaxation effects of O_2 and TEMPO are strikingly similar for the great majority of buried amide sites. The simplest plausible explanation of this behavior is a similar spatial distribution of the two different paramagnets with respect to the rubredoxin structures. However, for this model to be true, two distinct conditions must be met. Since the bulky TEMPO radical is not expected to permeate the protein structure to a significant degree, this model predicts that both paramagnets interact with the buried amides primarily by their distribution in the bulk aqueous phase surrounding the native protein structure. The second, more subtle, condition is that normalization against the paramagnetically induced solvent proton relaxation serves to divide out any explicit diffusional dependence of the O_2 or TEMPO effect. This condition is straightforwardly satisfied for O_2-induced relaxation. The O_2 electron spin−lattice relaxation time T_{1e} is ∼7.5 ps which is more rapid than either the rotational and translational correlation times in this system.[43] Therefore, the electron relaxation will dominate the effective correlation time for the electron−nuclear coupling, and the observed paramagnetic relaxation effect will be insensitive to the O_2 diffusion rate in the aqueous phase. Furthermore, the O_2 T_{1e} is nearly independent of solvent and local noncovalent environment,[43] so that oxygen molecules which penetrate the protein interior should induce relaxation with equivalent efficiency to those in the bulk phase. As a result, the paramagnetically induced relaxation effect of dioxygen can be directly estimated by summation over the time-averaged O_2 distribution weighted by the $1/r^6$ distance dependence of a simple dipole−dipole interaction.

It should be noted that deviations from single-exponential decay behavior might be anticipated if penetrating O_2 molecules contribute significantly to the ^1H amide relaxation, and if these internal oxygen molecules are not in rapid equilibrium with the aqueous phase. In marked contrast to the fluorescence quenching experiments mentioned previously, the intrinsic relaxation time of the monitored ^1H signals is ∼1 s. As noted in the Introduction, the residence times for internal bound water and various nonpolar small molecules are several orders of magnitude shorter. More specific to the present application, the rate of solvent (and hence presumably O_2) access to the various amide sites in these rubredoxin proteins is rapid on the time scale of a second.[22,33] Hence, the relaxation contribution arising from internalized oxygen molecules will be fully averaged into the observed single-exponential relaxation decay.

It is less obvious that normalization against the solvent ^1H T_1 value should remove the explicit diffusional dependence from the paramagnetically induced relaxation caused by the much more slowly relaxing TEMPO radical. In this case, translational and rotational diffusion dominates the modulation of the electron−nuclear dipolar interaction.[44] In particular, previous studies have indicated that the translational diffusion of water along the protein surface is ∼5 times slower than in bulk phase.[45,46] If this same positional−orientational dependent correlation time behavior were to hold for the nitroxide diffusion, it would be expected that, relative to the ^1H_2O relaxation effect, the nitroxide would relax the surface exposed amide protons much more efficiently than would O_2. However, in contrast to this prediction, the median values of the normalized O_2- and TEMPO-induced relaxation effects for the surface exposed amide protons of *Cp* rubredoxin are 0.52 and 0.53, respectively.

(43) Teng, C. L.; Hong, H.; Kiihne, S.; Bryant, R. G. *J. Magn. Reson.* **2001**, *148*, 31−34.
(44) Hornak, J. P.; Moscicki, J. K.; Schneider, D. J.; Freed, J. H. *J. Chem. Phys.* **1986**, *84*, 3387−3395.
(45) Polnaszek, C. F.; Byant, R. G. *J. Am. Chem. Soc.* **1982**, *106*, 428−429.
(46) Kimmich, R.; Klammler, F.; Skirda, V. D.; Serebrennikova, I. A.; Maklakov, A. I.; Fatkullin, N. *Appl. Magn. Reson.* **1993**, *4*, 425−440.

The corresponding values for the hyperthermophile rubredoxin are 0.46 and 0.52.

In fact, the dipolar relaxation contributions from the angular and radial components of translational diffusion tend to cancel against each other. This mutual cancellation effect is largely responsible for the comparative robustness of the "rigid dipole" model commonly used in the analysis of macromolecular ^{1}H−^{1}H NOE data.[47−51] In the case of rapid nonoverlapping uncorrelated isotropic diffusion, the relative effects of angular vs radial fluctuations in the dipolar interaction cancel identically, thus yielding a cross-relaxation rate which depends only on the $1/r^6$ separation of the mean ^{1}H positions.[48] Under more general conditions, the dipole−dipole interaction for spins undergoing translational diffusion scales approximately as the inverse cube of the assumed spherical excluded radius.[52] This $1/R^3$ dependence corresponds to a $1/r^6$-weighted summation over the exterior volume.

The evidence for minimal specific binding interactions and the apparent lack of a major contribution from radicals penetrating the protein interior suggests that the paramagnetically induced relaxation effects should be satisfactorily modeled by a uniform distribution of paramagnets in the bulk phase. To test this hypothesis, the coordinates of each rubredoxin structure were embedded in a cubic lattice with 0.1 Å grid spacing that extended at least 8.4 Å (6 times the 1.4 Å probe radius) beyond all protein atoms in each direction. All lattice points less than the van der Waals radius + 1.4 Å from any protein atom were excluded from the summations. The interior cavity positions present in *Cp* rubredoxin were excluded as well. For each amide hydrogen, the $1/r^6$ distance factor for each of the nonexcluded lattice points was calculated and summed. To generate the best experimental estimate of the relaxation effects of a nonpenetrating, uniformly distributed paramagnetic species, a combination of the O_2 and TEMPO data was used. If, in addition to the distribution on the protein exterior, O_2 penetration significantly contributes to the observed relaxation, the O_2-induced effect should be larger than the TEMPO-induced effect. On the other hand, any specific binding of TEMPO to the protein surface will tend to yield relaxation values that are higher than those of the generally more nonspecific O_2 interactions. To compensate for both of these effects, the minimum of the O_2- and TEMPO-induced relaxation values were used for comparison to the calculations based on the model of a nonpenetrating uniformly distributed paramagnet.

To compare the average $1/r^6$ calculation to the experimental data of Figure 1, a scale factor reflecting the grid spacing and probe radius (0.0300 for a 0.1 Å grid) was applied to both the *Cp* and *Pf* rubredoxin calculations so as to minimize the root mean square deviation to the paramagnetically induced relaxation data. In Figure 2a is shown the comparison between the minimum of the O_2- and TEMPO-induced relaxation values (●) and the average $1/r^6$ calculation (◇) for *Cp* rubredoxin. Three residues (9, 17, and 42) differ by more than 0.1. For the other 24 buried amides the root mean square deviation between the observed and predicted values is 0.050. Figure 2b presents the analogous data for *Pf* rubredoxin. The observed and predicted values for these same 24 buried amide positions have a root mean square deviation value of 0.039 in the hyperthermophile protein. Here as well, only the same three residues show differences exceeding 0.1, and the differences are of the same sign in both proteins for all three residues. For the case of residue 17, a structural interpretation of the differing O_2- and TEMPO-induced relaxation values is offered below.

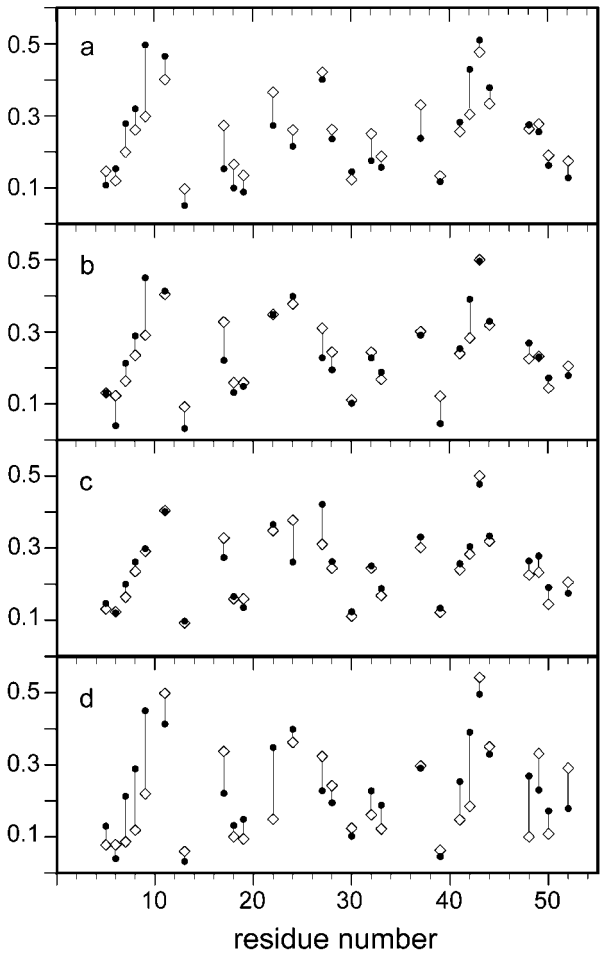

Figure 2. Prediction of the bulk-phase paramagnetically induced relaxation effects on the static solvent inaccessible amide protons in *Cp* and *Pf* rubredoxins. To model the time-averaged dipolar interaction between the amide protons and a paramagnet uniformly distributed in the bulk phase, a $1/r^6$-weighted sum (◇) was carried out over all lattice points (cubic with boundaries at least 8.4 Å beyond any protein atom) which are exterior to the protein van der Waals envelope + a 1.4 Å probe radius. Internal cavity space was also excluded. For each buried amide position, these calculations were compared to the minimum of the O_2- and TEMPO-induced relaxation effects for *Cp* (● in panel a) and *Pf* (● in panel b) rubredoxins. A comparison between the $1/r^6$-weighted lattice sums for *Cp* (●) and *Pf* (◇) rubredoxins is given in panel c. To illustrate the sensitivity of these calculations to the local steric asymmetry of the individual amide sites, in panel d is shown the *Pf* rubredoxin (●) relaxation data compared to the relaxation effect predicted for a closest approach model using the distance to the nearest exterior lattice point (◇).

The excellent agreement between the $1/r^6$-weighted lattice summation and the observed normalized relaxation rates demonstrates that a uniform distribution of exterior paramagnets adequately models the large majority of these data. Furthermore, as predicted for O_2 and proposed for the TEMPO radical, normalization against the paramagnetically induced ^1H_2O T_1

(47) Olejniczak, E. T.; Dobson, C. M.; Karplus, M.; Levy, R. M. *J. Am. Chem. Soc.* **1984**, *106*, 1923−1930.
(48) LeMaster, D. M.; Kay, L. E.; Brunger, A. T.; Prestegard, J. H. *FEBS Lett.* **1988**, *236*, 71−76.
(49) Bruschweiler, R.; Roux, B.; Blackledge, M.; Greisinger, C.; Karplus, M.; Ernst, R. R. *J. Am. Chem. Soc.* **1992**, *114*, 2289−2302.
(50) Post, C. B. *J. Mol. Biol.* **1992**, *224*, 1087−1101.
(51) LeMaster, D. M. *J. Biomol. NMR* **1997**, *9*, 79−93.
(52) Freed, J. *J. Chem. Phys.* **1978**, *68*, 4034−4037.

values appears to effectively remove any explicit dependence on diffusion rates.

For nearly all of the buried amides in both *Cp* and *Pf* rubredoxin, the experimental paramagnetically induced relaxation values reflect the protein tertiary structure as represented by the $1/r^6$ lattice sum. Since this lattice sum is hardly a familiar monitor of detailed tertiary structure, it is useful to compare the $1/r^6$ lattice sums to a more conventional measure of differences in tertiary structure. The root mean square deviation of a superposition between the C^α atoms of the *Cp* and *Pf* rubredoxins is 0.6 Å. As illustrated in Figure 2c, the root mean square deviation of the differences between the $1/r^6$ calculations for the corresponding amide hydrogens of these two rubredoxins is 0.039. When the two large deviations for residues 24 and 27 are removed, the root mean square deviation variation between the two $1/r^6$ calculations drops to 0.025. Thus, the root mean square deviation between the experimental paramagnetically induced relaxation values and the lattice calculations for a given rubredoxin (0.039 and 0.050 for *Pf* and *Cp* rubredoxins, respectively) is approximately twice that between the $1/r^6$ lattice calculations for these two homologous proteins.

Another measure of the tertiary structural detail provided by an experimentally estimated $1/r^6$-weighted averaging is indicated by comparison to the commonly used distance of closest approach model. The oversimplicity of the closest approach model can be illustrated by determining the radius of the sphere which bounds 50% of the $1/r^6$ exterior lattice sum for each amide hydrogen. For Glu 15 of *Pf* rubredoxin, which is buried 6.1 Å from the nearest exterior lattice point, the radius for the 50% sum is just under 11 Å. Two-thirds of the protein C^α atoms lie within this sphere, which bounds only half of the predicted exterior relaxation effect. For a more detailed comparison of the predictions based on a closest approach model, the distance to the nearest nonexcluded grid point was determined for each buried rubredoxin amide. The inverse cube of this closest distance for each amide was plotted against the experimental data, as illustrated in Figure 2d for *Pf* rubredoxin. The same 24 buried amide positions considered in the discussion of Figure 2a,b yield a root mean square deviation between the observed (●) and predicted (◇) paramagnetically induced relaxation values of 0.105. The markedly poorer fit to the experimental data given by the spherical model approximation clearly demonstrates the sensitivity of paramagnetically induced relaxation effect to the local nonspherical distribution of the excluded volume at different sites in the protein.

These lattice calculations provide a robust estimate of the observed paramagnetically induced relaxation effects which in turn appear to be highly sensitive to the details of the tertiary structure. Of particular note are residues 24 and 27 which yield markedly different $1/r^6$ lattice sum values for the *Cp* and *Pf* rubredoxins (Figure 2c), despite the fact that the difference in the distances between the homologous amide nitrogens in the superimposed structures is only 0.68 and 0.46 Å, respectively. As seen in the comparison between the TEMPO-induced relaxation of the two proteins (Figure 1b), the magnitude and relative sign of the differences at residues 24 and 27 correspond closely to that predicted from the $1/r^6$-weighted calculation. At least some of the other apparent outliers also appear to be understandable in terms of a model of uniformly distributed external paramagnets. Although a 1.4 Å probe appears to be a satisfactory representation for the nearly spherical O_2 molecule, it is clearly a less appropriate model for 4-hydroxy TEMPO. On one hand, quantum mechanical calculations indicate that much of the nitroxide radical spin density resides on the oxygen,[53] so that a 1.4 Å probe would appear appropriate for convex segments of the protein surface. However, when confronted with a concave protein surface, the two vicinal dimethyl groups may prevent the radical from reaching positions that are accessible to dioxygen. In these cases the normalized TEMPO-induced relaxation value would be systematically smaller than that of the O_2-induced relaxation. Amides fitting this pattern include residue 17 of both *Cp* (Figure 2a) and *Pf* (Figure 2b) as well as residue 37 of *Cp* rubredoxin (Figure 2a).

It must be emphasized that the residual discrepancies between the predicted and observed paramagnetically induced relaxation effects of Figure 2a,b represent a systematic underestimate of the accuracy for the model of relaxation via a uniformly distributed external paramagnet. As noted above, a 1.4 Å probe is an oversimplification for the steric constraints of the TEMPO radical. Lattice points near concave protein surfaces could be assessed using a more realistic molecular representation of the TEMPO molecule. Conversely, the fixed hard sphere X-ray coordinates assumed in these calculations surely exclude some lattice point contacts which are in fact readily accessible for the conformationally dynamic protein molecule. When warranted, molecular dynamics simulations could be used to assign more realistic weightings for the surface lattice points.

Discussion

Paramagnetically induced relaxation provides a particularly effective means of estimating the time-averaged O_2 spatial distribution. The rapid modulation of the electron−nuclear coupling by the short T_{1e} yields relaxation effects which are largely independent of differential diffusion rates. The insensitivity of the O_2 T_{1e} to the local noncovalent environment implies that the oxygen molecule will exert similar relaxation interactions whether it is buried in the protein interior or diffusing in the bulk solvent. The success of the $1/r^6$-weighted lattice summations in predicting the relaxation contribution arising from paramagnets in the bulk phase warrants a closer examination of the potential relaxation contributions arising from oxygen molecules which have penetrated the protein interior. When sites of specific binding interactions are accounted for as discussed above, the normalized TEMPO-induced relaxation effects represent a lower limit estimate of the relaxation contribution which arises from the component of the total oxygen that is uniformly distributed in the bulk phase. Hence, the experimentally observed difference between the O_2- and TEMPO-induced relaxation represents an upper limit estimate of the relaxation contribution arising from oxygen which has penetrated the protein interior.

Of the 27 buried amides monitored in this study, the differential normalized paramagnetically induced relaxation (O_2 − TEMPO) is greater than 0.1 for only three positions in *Pf* rubredoxin and four positions in the *Cp* protein (Figure 1c). For the other buried amides, the mean O_2-induced relaxation exceeds the mean TEMPO-induced value by 0.013 and 0.019 for *Pf* and *Cp* rubredoxins, respectively. These results indicate

(53) Eastman, M. P.; Kooser, R. G.; Das, M. R.; Freed, J. H. *J. Chem. Phys.* **1969**, *51*, 2690−2709.

that the upper limit of the effective partition coefficient of dioxygen between the interior of both *Pf* and *Cp* rubredoxins and the aqueous bulk phase is at most 0.1 and is potentially as low as 0.01−0.02. This is in marked contrast to the far higher solubility of O_2 in the nonpolar organic solvents that have been used for phase-transfer measurements as a means of estimating free energies of hydrophobic burial in protein interiors. Indeed, the partition coefficient of 2.9 for dioxygen into lipid bilayers[54] has been used to measure immersion depths of nitroxide-labeled membrane proteins[55] and fluorinated detergents[56] via paramagnetic-induced relaxation.

In the experiments monitoring the dynamic quenching of tryptophan fluorescence mentioned above, O_2 quenched a series of protein tryptophan side chains 20−50% as efficiently as for free tryptophan.[5] When interpreted as an internal diffusion process, this 2- to 5-fold attenuation arises from the product of both the local O_2 concentration and O_2 diffusion constant in the protein interior. The low O_2 partition coefficient observed for the interior of the rubredoxins implies a significant limitation of the efficiency of a dioxygen diffusion process through such a protein matrix relative to that in the bulk aqueous phase.

As predicted by scaled particle theory, the relative solubility of weakly interacting solutes is dominated by the free energy cost of forming an empty cavity in the solvent phase.[57,58] This can be estimated experimentally from the isothermal compressibility which has been most directly measured in proteins by X-ray diffraction as a function of pressure. Studies on hen egg white lysozyme[59] have demonstrated significant local variations in the responsiveness to pressure with an average compressibility of 4.7×10^{-3} kbar^{-1}, which is an order of magnitude below that of bulk water.

Since the upper limit of 0.1 for the partition coefficient of dioxygen in the interior of the rubredoxins correlates with the lower compressibility of the protein matrix, it is worth considering the potential generality of these observations. Compared to most proteins,[9] the rubredoxins have a low volume fraction of internal cavities that are large enough to accommodate an O_2 (or water) molecule. The presence of such large cavities potentially offers increased free volume suitable for mobile defects to facilitate O_2 diffusion.[20,21] The pressure dependence of NMR chemical shifts provides an independent experimental monitor of protein compressibility. The linear term in pressure dependence has been found to be rather uniform among a wide range of protein structures.[60] In particular, it was found to be independent of the volume fraction of cavities large enough to accommodate a water molecule. Hence, the regions containing water-sized cavities do not appear to exhibit substantially higher compressibility than the remainder of the protein interior. In contrast, the quadratic term of the pressure-dependent chemical shift changes was found to correlate with the volume fraction of water-sized cavities. This is consistent with a model in which water molecules are forced into preformed cavities for which the enthalpy of water binding is unfavorable at atmospheric pressure. Analogous behavior as a function of O_2 pressure is anticipated. The majority of the closely packed protein interior exhibits a fairly low partitioning of oxygen due to the energetic cost of cavity formation. Although largely absent in the rubredoxins, proteins exhibiting larger preformed cavities will presumably fill those cavities as a function of O_2 pressure depending on the details of the noncovalent interactions and the possible need to displace internalized water molecules.

An estimate of the differential O_2 solubility in the interiors of the mesophile vs hyperthermophile rubredoxins is provided by comparison of the mean (O_2 − TEMPO) paramagnetic-induced relaxation values for the two proteins. Averaging over all 27 buried amide positions, the mean (O_2 − TEMPO) difference is 0.01 larger for *Cp* rubredoxin vs that for *Pf* rubredoxin with half of this variation arising from the three residues exhibiting the largest differences. These data indicate that the difference in O_2 solubility for the interior of *Cp* vs *Pf* rubredoxin is approximately 1% of the solubility for O_2 in the bulk aqueous phase. The expected correlation with isothermal compressibility is consistent with a rather modest differential rigidity of the mesophile and hyperthermophile proteins, at least as reflected in the high-frequency, low-amplitude fluctuations sampled by compressibility and solubility measurements.

Having found that a $1/r^6$-weighted summation over a uniform external distribution provides a reasonably accurate prediction of the experimentally determined minimum (O_2, TEMPO)-induced relaxation effect, the practical utility of the reverse prediction was considered. With the recent dramatic expansion in the availability of genomic DNA sequence information, increased interest has arisen in the problem of recognizing structural homology between highly divergent protein sequences. This problem is of particular significance for protein pairs that are below the so-called "twilight zone" of sequence homology prediction.[61] Over 90% of protein sequences exhibiting more than 30% sequence identity are structurally homologous, while less than 10% of protein sequences exhibiting under 25% sequence identity are structural homologues.[62]

In practical terms, the desire to rapidly categorize proteins according to families of folding motifs places a premium on establishing this characterization as early as possible in the structural analysis process. As residual dipolar couplings provide orientational constraints with respect to the protein molecular frame, various NMR studies have examined their utility in early stages of solution structure analysis for folding motif recognition. The most straightforward approaches to utilizing residual dipolar couplings for folding motif recognition require the simultaneous determination of the bond vector orientations with respect to the molecular orientation tensor. Efforts have been made to carry out this orientational determination for a series of segments along the protein sequence[63,64] or for the entire set of protein residual dipolar couplings simultaneously.[65,66] Unfortunately, the structural segments which yield the most characteristic residual dipolar coupling patterns in the oriented structure (e.g. α-helices), are often the least robust in providing the initial orienta-

(54) Kimmich, R.; Peters, A. *Chem. Phys. Lipids* **1975**, *14*, 350−362.
(55) Altenbach, C.; Greenhalgh, D. A.; Khorana, H. G.; Hubbell, W. L. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1667−1671.
(56) Prosser, R. S.; Luchette, P. A.; Westermann, P. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9967−9971.
(57) Pierotti, R. A. *J. Phys. Chem.* **1963**, *67*, 1840−1845.
(58) Nandi, N.; Basumallick, I. N. *Z. Phys. Chem.* **1991**, *173*, 179−189.
(59) Kundrot, C. E.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 157−170.
(60) Akasaka, K.; Li, H. *Biochemistry* **2001**, *40*, 8665−8671.
(61) Doolittle, R. F. *Of URFs and ORFs: A Primer on How to Analyze Derived Protein Sequences;* University Science Books: Mill Valley, CA, 1986.
(62) Rost, B. *Protein Eng.* **1999**, *12*, 85−94.
(63) Delagio, F.; Kontaxis, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 2142−2143.
(64) Andrec, M.; Du, P.; Levy, R. M. *J. Am. Chem. Soc.* **2001**, *123*, 1222−1229.

tional constraints.[67] Recently, procedures have been developed for residual dipolar coupling-based structural refinement which utilize frame-independent orientational restraints.[68] However, to date these approaches have not yet been applied to the homologous protein structure recognition problem.

A necessary precursor to any 3D alignment of an unknown protein structure onto a known target structure is the initial (at least partial) alignment of their sequences. Central to the commonly used global (e.g., Needleman−Wunsch[69]) and local (e.g., BLAST[70]) sequence alignment algorithms is a scalar residue-by-residue comparison between a sample and target sequence. The percent of sequence identity is obtained using a unit matrix, while more generally substitution matrices (e.g., PAM[71] or BLOSUM[72]) contain the expected probabilities for substitution of one amino acid type for another based on mutational frequencies within evolutionary protein family trees. By construction, fold recognition experiments are directed toward proteins for which sequence alignment based on mutational frequencies is insufficient. In marked contrast to the residual dipolar coupling data, the paramagnetically induced relaxation measurements yield scalar data which can be compared on a residue-by-residue basis to estimated values derived from calculations on target protein structures.

The $1/r^6$-weighted lattice summations for each of the two rubredoxins predicted the experimental paramagnetically induced relaxation measurements to a precision approximately half of that between the $1/r^6$-weighted lattice summations for the two proteins. This prediction can be applied in reverse to generate a two-step algorithm for comparing an unknown protein structure to a set of known target structures. The experimental relaxation data are first used to predict the $1/r^6$-weighted lattice sums for the unknown protein structure which in turn are compared to the lattice summations obtained for the target protein structures.

The potential utility of this class of experiments in predicting structural homology can be illustrated by model calculations on several recently determined helix-turn-helix DNA-binding proteins. NMR analysis of the DNA-binding domain from the *E. coli* lexA protein[73] demonstrated the expected structural homology to various previously known helix-turn-helix proteins, although a one residue insertion at the junction between the helices yields an expanded turn region. In contrast, structural homology to the C-terminal domain of the human ssDNA-binding replication protein A (RPA) (2.4 Å main chain root mean square deviation vs lexA domain) was not anticipated on the basis of either sequence analysis (15% sequence identity when structurally aligned with the lexA domain) or the absence of any known DNA binding function.[74] Similarly, the Z-DNA-binding domain Zα from the dsRNA adenosine deaminase type I was found to possess a similar conformation (2.0 Å main chain root mean square deviation vs lexA domain) despite having a similarly weak 16% sequence identity with the lexA domain and a markedly different DNA conformational specificity.[75,76]

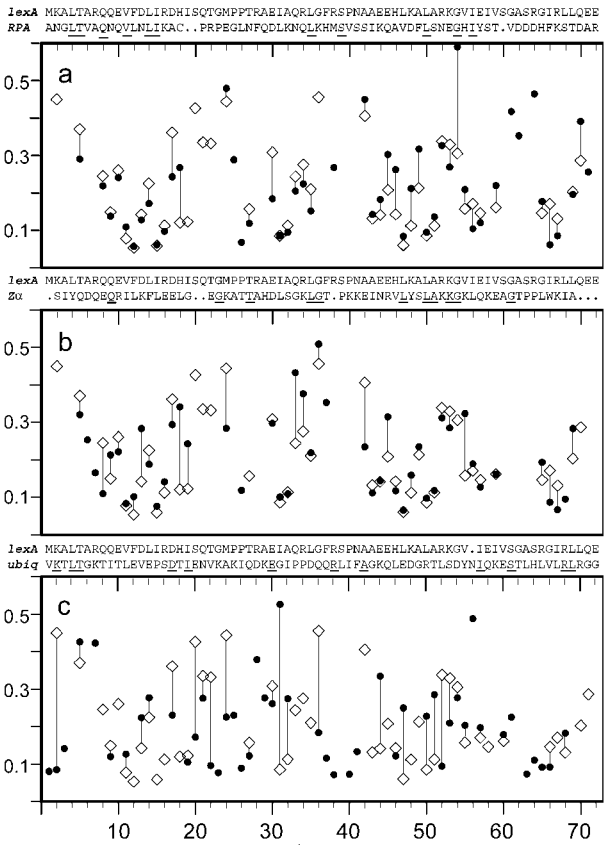

Figure 3. Comparison of the $1/r^6$-weighted lattice summations for the helix-turn-helix DNA binding domain of *E. coli* lexA (◇) to structurally homologous and nonhomologous proteins. In panel a is shown calculations from the lexA domain (1LEA)[73] compared to the C-terminal domain of the human ssDNA-binding replication protein A (1DPUa).[74] Analogous comparisons for the homologous Z-DNA-binding domain Zα from the dsRNA adenosine deaminase type I (1QBJa)[75,76] and the nonhomologous ubiquitin (1UBQ)[77] are given in panels b and c, respectively. Using structure-based sequence alignment[78,79] for RPA and Zα and direct sequence alignment[69] for ubiquitin, essentially equivalent sequence identities with lexA are observed (15% for RPA, 16% for Zα, and 17% for ubiquitin). Sequence identities are underlined, and in panel c the last residue of lexA is excluded from the plot to accommodate the one residue insertion at position 56.

In Figure 3 are illustrated comparisons of the $1/r^6$-weighted summations over the exterior volumes of the known structures of lexA domain (◇) vs the RPA domain (Figure 3a), the Zα domain (Figure 3b), and the nonhomologous ubiquitin (Figure 3c). For each protein, data from only the static solvent inaccessible amides are plotted. In comparing the calculations for the lexA domain vs the RPA domain in Figure 3a, the

(65) Annila, A.; Aitio, H.; Thulin, E.; Drakenberg, T. *J. Biomol. NMR* **1999**, *14*, 223−230.
(66) Meiler, J.; Peti, W.; Griesinger, C. *J. Biomol. NMR* **2000**, *17*, 283−294.
(67) Fushman, D.; Ghose, R.; Cowburn, D. *J. Am. Chem. Soc.* **2000**, *122*, 10640−10649.
(68) Meiler, J.; Blomberg, N.; Nilges, M.; Greisinger, C. *J. Biomol. NMR* **2000**, *16*, 245−252.
(69) Needleman, S. B.; Wunsch, C. D. *J. Mol. Biol.* **1970**, *48*, 443−453.
(70) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403−410.
(71) Dayhoff, M. O.; Schwartz, R. M.; Orcutt, B. C. *Atlas of Protein Sequence and Structure*; National Biomedical Research Foundation: Washington, D.C., 1978; Vol. 5.
(72) Henikoff, S.; Henikoff, J. G. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10915−10919.
(73) Fogh, R. H.; Ottleben, G.; Ruterjans, H.; Schnarr, M.; Boelens, R.; Kaptein, R. *EMBO J.* **1994**, *13*, 3936−3944.
(74) Mer, G.; Bochkarev, A.; Gupta, R.; Bochkareva, E.; Frappier, L.; Ingles, C. J.; Edwards, A. M.; Chazin, W. J. *Cell* **2000**, *103*, 449−456.
(75) Schade, M.; Turner, C. J.; Lowenhaupt, K.; Rich, A.; Herbert, A. *EMBO J.* **1999**, *18*, 470−479.
(76) Schwartz, T.; Rould, M. A.; Lowenhaupt, K.; Hebert, A.; Rich, A. *Science* **1999**, *284*, 1841−1845.
(77) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544.
(78) Holm, L.; Sander, C. *Nucleic Acids Res.* **1994**, *22*, 3600−3609.
(79) Holm, L.; Sander, C. *Science* **1996**, *273*, 595−602.
(80) Marck, C. *Nucleic Acids Res.* **1988**, *16*, 1829−1836.

predicted paramagnetic relaxation effects for the two proteins track each other rather closely in the sequence intervals for residues 5−16 and 27−35, as well as for residues 42−59 with the exception of Gly 54. It should be noted that Gly54 is solvent exposed in the Zα domain (Figure 3b). Furthermore, the comparison between the lexA and Zα domains in Figure 3b indicate correlated $1/r^6$-weighted summation values over the same three segment regions. Both sets of comparisons show larger discrepancies, including burial of amide protons in only one of the two protein structures, near the sites of insertion/deletion around residues 20, 38, and 60. Since all of the segments between the sites of insertion/deletion show significant correlation of the predicted paramagnetic relaxation, correct sequence alignment should be feasible with such data.

The structurally nonhomologous ubiquitin aligns to the lexA domain with a 17% identity using the Needleman−Wunsch algorithm[69] with a maximum of two gaps (Figure 3c). However, the correlation between the $1/r^6$-weighted summation values for these two proteins is clearly worse. In marked contrast to the RPA and Zα domain comparisons, for lexA vs ubiquitin no sequence segment shows more than two adjacent positions, yielding similar estimated paramagnetic relaxation effects (Figure 3c). Furthermore, only 30 residues in the lexA and ubiquitin sequence alignment are buried in both X-ray structures, consistent with a random distribution of the observed fraction of static solvent inaccessible amides.

As is readily apparent, the predicted differences in relaxation behavior between the pairs of proteins follow a highly non-Gaussian distribution. To make a more quantitative comparison between these calculations, robust statistical methods help reduce the skewing effect of small outlier populations.[42] For this type of application, the residues exhibiting highly similar predicted relaxation effects will predominantly drive the structural homology analysis. Using the buried amide positions of the lexA domain as the reference set, median absolute differences for the $1/r^6$-weighted sums of 0.045, 0.047, and 0.184 were observed for RPA, Zα, and ubiquitin, respectively. The larger difference seen for the nonhomologous ubiquitin reflects a greater spread in relaxation values at amide positions that are buried in both structures. Equally importantly, in the latter comparison there is a much larger fraction of buried lexA amides that are paired to solvent exposed positions in ubiquitin which exhibit systematically higher $1/r^6$-weighted summations.

Analogous $1/r^6$-weighted calculations were carried out for additional members of the rubredoxin, helix-turn-helix and ubiquitin families as well as for members of the thioredoxin family. As summarized in Table 1, intra- and interfamily comparisons were carried out to assess the degree of confidence with which assignment to the proper family can be made. The intrafamily comparisons indicate a clear correlation between median absolute differences for the $1/r^6$-weighted sums and the root mean square deviation between the C^α positions in the corresponding structures. In contrast, optimal sequence alignment of the nonhomologous proteins yields nearly constant values for the median absolute differences of the $1/r^6$-weighted sums, consistent with a lack of statistical correlation. The 3- to 4-fold difference between the $1/r^6$-weighted sums for the nonhomologous and the highly diverged homologous structures demonstrates a substantial dynamical range for this statistical measure of structural homology. Application of the paramagnetically induced relaxation measurements to the problem of structural homology detection implies the additional imprecision of using experimental relaxation data to estimate the corresponding (unknown) $1/r^6$-weighted summation. For the *Cp* and *Pf* rubredoxin data given above, the average of the median absolute differences between the paramagnetically induced relaxation data and the $1/r^6$-weighted sums is 0.035, equivalent to the quality of estimate derived from an homologous protein structure which differs by a C^α root mean square deviation of 1.5−2.0 Å. Since these errors can be expected to be largely uncorrelated with the structure-dependent variations in the $1/r^6$-weighted sums, a nearly 3-fold difference in the median absolute deviation is estimated for the structure prediction of the least favorable homologous comparison (1LEA-1QBJa) vs the most favorable nonhomologous calculation (5RXN-1THX).

Table 1. Median Absolute Deviation for $\Sigma 1/r^6$ within[a] and between[b] Protein Fold Families

PDB file	% identity	Med$\lvert\Delta\Sigma 1/r^6\rvert$	C^α rmsd
5RXN			
1BQ8	57	0.021	0.6
1RB9	70	0.022	0.6
1LEA			
1DPUa	15	0.045	2.4
1QBJa	16	0.047	2.0
1UBQ			
1NNDa	57	0.021	0.8
1H8Ca	17	0.036	2.1
1THX			
2TRXa	42	0.021	1.2
1ERU	23	0.029	1.3
1LEA-1UBQ	17	0.184[c]	
1LEA-5RXN	13	0.179	
1LEA-1THX	16	0.175	
5RXN-1UBQ	20	0.183	
5RXN-1THX	7	0.167	
1UBQ-1THX	14	0.171	

[a] Structural alignments and C^α root mean square deviation values from the FSSP database.[78,79] [b] Sequence alignment by Needleman−Wunsch algorithm[69] with two gaps allowed (DNA Strider[80]). [c] Mean of calculations using each protein as reference.

Conclusions

In the absence of specific binding to a preformed internal cavity, the effective solubility of O_2 in the rubredoxin interior is at least 1 order of magnitude less than that of bulk water, consistent with published isothermal compressibility results from other proteins. The quite small difference observed for the penetration of O_2 into the interiors of the mesophile vs hyperthermophile rubredoxins suggests a modest difference in conformational flexibility as sensed by the high-frequency fluctuations which dominate solubility effects. As paramagnetically induced relaxation offers a robust means of measuring relative O_2 concentration, its potential for characterizing the binding of oxygen to internal protein cavities should be noted.

The ability to accurately predict the detailed paramagnetically induced relaxation effects for buried protein protons indicates considerable potential for facile recognition of structural motifs. As only $^1H^N$ and ^{15}N resonances are utilized in this approach, HN(CA)N and HN(COCA)N type experiments[81,82] can prove adequate for providing the necessary sequence specific assign-

(81) Lohr, F.; Pfeiffer, S.; Lin, Y. J.; Hartlieb, J.; Klimmek, O.; Ruterjans, H. *J. Biomol. NMR* **2000**, *18*, 337−346.
(82) Panchal, S. C.; Bhavesh, N. S.; Hosur, R. V. *J. Biomol. NMR* **2001**, *20*, 135−147

ments without any ^{13}C assignment required. Correlations of $1/r^6$-weighted lattice calculations and paramagnetically induced relaxation effects for burial depths up to at least 6 Å indicate applicability to appreciably larger proteins. Lattice calculations on RNase A and phage T4 lysozyme, proteins which are approximately 2- and 3-fold larger than the rubredoxins, indicate that the fraction of the static solvent inaccessible amides which are buried more deeply than 6 Å is only 6 and 3%, respectively.

The potential applicability of paramagnetic relaxation for determining the effective protein burial depth is not limited to its use in rapid determination of folding family members. Solution structural analysis with a limited set of conformational constraints, as is typically the case for larger proteins, is generally confronted with the problem of narrowing down a broad distribution of initial starting structures for the subsequent refinement process. It is anticipated that the paramagnetic relaxation data may prove highly useful in this regard. It should be noted that, in contrast to most other forms of NMR structural constraints, the paramagnetic relaxation data provide spatial information about protein atoms which do not give rise to observable NMR signals. Although indirect in nature, this information on the local packing environment should prove valuable in orienting otherwise poorly constrained protein groups. The paramagnetic relaxation analysis described herein is not limited to amide resonances. Appropriate combinations of selective ^{13}C and ^{2}H labeling will allow for analogous measurements at side chain positions which may prove of increasing value as larger molecular weight systems are considered.

Acknowledgment. We thank Marly Eidsness for providing the expression plasmids used in this work. We acknowledge the use of the Wadsworth Center NMR Facility and Protein Mass Spectrometry Facility.

JA017340K