J|A|C|S
ARTICLES
Published on Web 04/17/2004

Ensemble Approach for NMR Structure Refinement against ^{1}H Paramagnetic Relaxation Enhancement Data Arising from a Flexible Paramagnetic Group Attached to a Macromolecule

Junji Iwahara,[†] Charles D. Schwieters,[‡] and G. Marius Clore*,[†]

Contribution from the Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health, Bethesda, Maryland 20892-0520, and Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892-5624

Received December 8, 2003; E-mail: mariusc@intra.niddk.nih.gov

Abstract: Paramagnetic relaxation enhancement (PRE) measurements on ^{1}H nuclei have the potential to play an important role in NMR structure determination of macromolecules by providing unique long-range (10−35 Å) distance information. Recent methodological advances for covalently attaching paramagnetic groups at specific sites on both proteins and nucleic acids have permitted the application of the PRE to various biological macromolecules. However, because artificially introduced paramagnetic groups are exposed to solvent and linked to the macromolecule by several freely rotatable bonds, they are intrinsically flexible. This renders conventional back-calculation of the ^{1}H-PRE using a single-point representation inaccurate, thereby severely limiting the utility of the ^{1}H-PRE as a tool for structure refinement. To circumvent these limitations, we have developed a theoretical framework and computational strategy with which to accurately back-calculate ^{1}H-PREs arising from flexible paramagnetic groups attached to macromolecules. In this scheme, the ^{1}H-PRE is calculated using a modified Solomon−Bloembergen equation incorporating a "model-free" formalism, based on a multiple-structure representation of the paramagnetic group in simulated annealing calculations. The ensemble approach for ^{1}H-PRE back-calculation was examined using several SRY/DNA complexes incorporating dT-EDTA−Mn^{2+} at three distinct sites in the DNA, permitting a large data set comprising 435 experimental backbone and side-chain ^{1}H-PREs to be obtained in a straightforward manner from 2D through-bond correlation experiments. Calculations employing complete cross-validation demonstrate that the ensemble representation provides a means to accurately utilize backbone and side-chain ^{1}H-PRE data arising from a flexible paramagnetic group in structure refinement. The results of ^{1}H-PRE based refinement, in conjunction with previously obtained NMR restraints, indicate that significant gains in accuracy can be readily obtained. This is particularly significant in the case of macromolecular complexes where intermolecular translational restraints derived from nuclear Overhauser enhancement data may be limited.

Introduction

Paramagnetic relaxation enhancement (PRE) provides information on distances between a paramagnetic center and nuclei. For an electron−nucleus distance r, the magnitude of the PRE is proportional to r^{-6}, which is similar to the relationship between the nuclear Overhauser effect (NOE) and interproton distance. Unlike the NOE which is limited to interproton distances less than ∼6 Å, the PRE can provide significantly longer distance information, up to ∼35 Å for Mn^{2+}−^{1}H interactions, due to the strong interaction between a paramagnetic electron and a nucleus arising from the large magnetic moment of the unpaired electron. Such long-range distance information derived from PRE measurements can potentially be extremely valuable in NMR structure determination because all other NMR observables that provide translational information are dependent on relatively short-range interactions between proximal nuclei. In addition, ^{1}H-PRE data can be readily acquired using through-bond correlation experiments in which cross-peaks can be easily assigned to unique interactions. This is in marked contrast to NOE experiments where unambiguous cross-peak assignment to unique through-space interactions may be severely hindered by chemical shift degeneracy. Despite its potential, the application of the PRE has not gained widespread general use in the NMR structure determination of biological macromolecules because the vast majority do not possess intrinsic paramagnetic centers.

The exceptions are metalloproteins with rigid paramagnetic centers. The application of the ^{1}H-PRE for structural analyses of these systems has a long history, and distances between the metal ion and protons have been successfully used in structure determination of metalloproteins and their bound ligands.[1]

In principle, the same kind of long-range distance restraints can be obtained for any macromolecule by introducing extrinsic

[†] National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.
[‡] Center for Information Technology, National Institutes of Health.

paramagnetic centers through appropriate chemical modification. This was first demonstrated about 20 years ago on spin-labeled lysozyme and bovine pancreatic trypsin inhibitor.[2a,b] Examples of extrinsic paramagnetic centers include covalent attachment of nitroxide labels[2] and metal chelators.[3,4] EDTA-derivatized deoxythymidine (dT-EDTA), whose phosphoreamidite derivative is commercially available, can be easily incorporated into DNA using conventional solid-phase synthesis, thereby enabling PRE measurements on any protein−DNA complex.[4] These technical advances have expanded the potential applicability of the PRE for obtaining long-range structural information in biological macromolecules.

An important issue, however, still remains to be resolved which, to date, has precluded the widespread use of PRE measurements in macromolecular structure determination: how to treat ^{1}H-PRE data arising from a flexible paramagnetic group. Because artificially introduced paramagnetic groups are invariably attached to the macromolecule of interest through linkers that have several rotational degrees of freedom, the paramagnetic centers are intrinsically flexible and, hence, the ensemble space occupied by the populated conformers must be relatively large. In such cases, it will be inappropriate to apply the Solomon−Bloembergen equation[5,6] (referred to as the SB equation) in which the paramagnetic center is represented as a single point in the structure calculations, because some nuclei will be strongly affected by a particular region of the conformational space occupied by the paramagnetic center and other nuclei by another region of the space. In the context of direct structure refinement against ^{1}H-PRE data, this is absolutely critical because the single-point representation of the paramagnetic center is likely to reduce the coordinate accuracy of the resulting structures because of distortions introduced through inappropriate treatment of the ^{1}H-PRE data.

In this paper, we present a theoretical and computational approach to quantitatively make use of ^{1}H-PRE data arising from a flexible paramagnetic group as a source of restraints for NMR structure calculations. First, we introduce a model-free formalism into the Solomon−Bloembergen theory and show how internal motions affect the longitudinal and transverse ^{1}H-PREs (Γ_1 and Γ_2, respectively). We then set out the computational strategies employed to deal with the intrinsic flexibility of an extrinsic paramagnetic group attached to a macromolecule. Finally, we show that, using this approach, it is possible to take full advantage of PRE data on both backbone and side-chain ^{1}H nuclei in structure calculations on macromolecules with artificially introduced paramagnetic groups. We demonstrate the applicability of this method to SRY/DNA complexes in which EDTA−Mn^{2+} has been attached to three distinct positions on the DNA.

Theory

PRE by a Flexible Paramagnetic Group Attached on a Macromolecule. We consider macromolecules with chemically stable paramagnetic groups that are either covalently attached or tightly bound. In addition, we focus on paramagnetic systems in which the electronic g-tensor is isotropic (i.e., the g-factor is close to 2.0) and Curie spin relaxation mechanism is negligible. The unpaired electrons of Cu^{2+}, Mn^{2+}, and nitroxides satisfy these conditions.

The PRE rate Γ is measured as a difference in relaxation rates for the paramagnetic and diamagnetic states of the macromolecule:

$$\Gamma = R_{\text{paramagnetic}} - R_{\text{diamagnetic}} \tag{1}$$

This subtraction cancels out all relaxation mechanisms common to both states, including exchange contributions to the transverse relaxation rate R_2, such that the only remaining relaxation mechanism arises from electron−nucleus interactions.

There are two reasons that ^{1}H nuclei are ideally suited for PRE analysis to derive distance information. First, the PRE on ^{1}H nuclei is significantly larger than that on ^{13}C and ^{15}N nuclei, because the PRE is proportional to the square of the nuclear gyromagnetic ratio. Second, the point−point approximation used in the equations describing the PRE is not valid for ^{13}C and ^{15}N nuclei due to delocalization of an unpaired spin to the $2p_z$ orbital.[7−9]

In the case where Fermi contact interactions are negligible (i.e., the paramagnetic center−^{1}H distance is long enough), the PREs on longitudinal and transverse relaxation rates are conventionally described by the Solomon−Bloembergen (SB) equations:[5]

$$\Gamma_1 = \frac{2}{5}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2 g^2 \mu_B^2 s(s+1) J_{SB}(\omega_I) \tag{2}$$

$$\Gamma_2 = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2 g^2 \mu_B^2 s(s+1)\{4J_{SB}(0) + 3J_{SB}(\omega_I)\} \tag{3}$$

where s is the electron spin quantum number, g is the electron g-factor, γ_I is the proton gyromagnetic ratio, μ_0 is the permeability of a vacuum, μ_B is the magnetic moment of the free electron, and $\omega_I/2\pi$ is the Larmor frequency of the proton. $J_{SB}(\omega)$ is the generalized spectral density function for the reduced correlation function:

$$J_{SB}(\omega) = r^{-6}\frac{\tau_c}{1 + (\omega\tau_c)^2} \tag{4}$$

The correlation time τ_c is defined as $(\tau_r^{-1} + \tau_s^{-1})^{-1}$, where τ_r is the rotational correlation time of the macromolecule, and τ_s

(1) (a) Cheng, H.; Markley, J. L. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 209−237. (b) Bertini, I.; Lichinat, C.; Piccioli, M. *Methods Enzymol.* **2001**, *339*, 314−340. (c) Ubbink, M.; Worrall, J. A. R.; Canters, G. W.; Groenen, E. J. J.; Huber, M. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 393−422.

(2) (a) Schmidt, P. G.; Kuntz, I. D. *Biochemistry* **1984**, *23*, 4261−4266. (b) Kosen, P. A.; Scheek, R. M.; Naderi, H.; Basus, V. J.; Manogaran, S.; Schmidt, P. G.; Oppenheimer, N. J.; Kuntz, I. D. *Biochemistry* **1986**, *25*, 2356−2364. (c) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 158−169. (d) Ramos, A.; Varani, G. *J. Am. Chem. Soc.* **1998**, *120*, 10992−10993. (e) Battiste, J. L.; Wagner, G. *Biochemistry* **2000**, *39*, 5355−5365. (f) Gaponenko, V.; Howarth, J. W.; Columbus, L.; Gasmi-Seabrook, G.; Yuan, J.; Hubbel, W. L.; Rosevear, P. R. *Protein Sci.* **2000**, *9*, 302−309. (g) Jain, N. U.; Venot, A.; Umeto, K.; Leffler, H.; Prestegard, J. H. *Protein Sci.* **2001**, *10*, 2393−2400.

(3) (a) Gaponenko, V.; Dvoretsky, A.; Walsby, C.; Hoffman, B. M.; Rosevear, P. R. *Biochemistry* **2000**, *39*, 15217−15224. (b) Donaldson, L. W.; Skrynnikov, N. R.; Choy, W. Y.; Muhandiram, D. R.; Sarkar, B.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 9843−9847. (c) Dvoretsky, A.; Gaponenko, V.; Rosevear, P. R. *FEBS Lett.* **2002**, *528*, 189−192. (d) Gaponenko, V.; Sarma, S. P.; Altieri, A. S.; Horita, D. A.; Li, J.; Byrd, R. A. *J. Biomol. NMR* **2004**, *28*, 205−212.

(4) Iwahara, J.; Anderson, D. E.; Murphy, E. C.; Clore, G. M. *J. Am. Chem. Soc.* **2003**, *125*, 6634−6635.

(5) (a) Solomon, I. *Phys. Rev.* **1955**, *99*, 559−565. (b) Solomon, I.; Bloembergen, N. *J. Chem. Phys.* **1956**, *25*, 261−266. (c) Bloembergen, N. *J. Chem. Phys.* **1957**, *27*, 572−573. (d) Bloembergen, N.; Morgan, L. O. *J. Chem. Phys.* **1961**, *34*, 842−850.

(6) Westlund, P.-O. *Mol. Phys.* **1995**, *85*, 1165−1178.

(7) Gottlieb, H. P. W.; Barfield, M.; Doddrell, D. M. *J. Chem. Phys.* **1977**, *67*, 3785−3794.

is the effective electron relaxation time. The Solomon−Bloembergen theory makes the approximation that electron relaxation does not couple with molecular tumbling. This is reasonable because the electron relaxation lifetime is comparable to or shorter than the rotational correlation time of a macromolecule. In the case of Mn^{2+} ($s = {}^5/_2$), electron relaxation is multiexponential.[10] Theoretical considerations, however, have shown that the SB equation employing a single effective electron relaxation rate τ_s^{-1} does not introduce any significant errors at high magnetic field strength (>10 T).[6]

The Solomon−Bloembergen theory assumes that dipole−dipole interaction vectors are rigid in the molecular frame. Because PRE interaction vectors may be quite long (up to 35 Å), they are less susceptible to small fluctuations in atomic positions, which makes this assumption reasonable for many cases. However, this approximation will break down if the ensemble space sampled by the paramagnetic group is quite large.

Here, we incorporate a "model-free" formalism[11] into the Solomon−Bloembergen theory to evaluate the influence of internal motions on the PRE. We assume that internal motions are not coupled with the overall tumbling of the molecule so that the correlation function, $C_I(t)$, for the internal motion of the interaction vector can be approximated by:

$$C_I(t) = S^2 + (1 - S^2)\exp(-t/\tau_i) \tag{5}$$

where S^2 is the generalized order parameter, and τ_i is the correlation time for internal motion. This formalism does not require any physical model of the internal motions (hence the term "model-free"), but has proved to be extremely useful in the interpretation of heteronuclear relaxation data.[11] Although the "model-free" formalism is commonly used for fixed-length interaction vectors such as $^{15}N-^1H$ or $^{13}C-^1H$ bond vectors, it can also be applied to vectors of variable lengths such as those for homonuclear $^1H-^1H$ dipolar interactions.[12,13] In that case, the order parameter is defined as:[12]

$$S^2 = \frac{4\pi}{5}\langle r^{-6}\rangle^{-1}\sum_{m=-2}^{2}\left|\left\langle\frac{Y_2^m(\Omega^{mol})}{r^3}\right\rangle\right|^2 \tag{6}$$

where $Y_2^m(\Omega)$ are second-order spherical harmonics, and Ω^{mol} are Euler angles in the molecular frame. As Brüshweiler et al. demonstrated for $^1H-^1H$ dipolar interactions,[13] it is useful to approximate the order parameter by decomposition into its radial and angular components:

$$S^2_{PRE} \approx S^2_{PRE,angular}S^2_{PRE,radial} \tag{7}$$

in which the angular ($S^2_{PRE,angular}$) and radial ($S^2_{PRE,radial}$) order parameters are defined as:

$$S^2_{PRE,angular} = \frac{4\pi}{5}\sum_{m=-2}^{2}|\langle Y_2^m(\Omega^{mol})\rangle|^2 \tag{8}$$

$$S^2_{PRE,radial} = \langle r^{-6}\rangle^{-1}\langle r^{-3}\rangle^2 \tag{9}$$

We assume that the effective electron relaxation rate τ_s^{-1} is not influenced by either overall or internal motion correlation times of the interaction vector. For a transition metal ion, electron relaxation primarily arises from modulation of the zero-field splitting tensor as a consequence of collisions with solvent molecules and is therefore governed by a very short lifetime between collisions τ_v (~5 ps for water at physiological temperature).[10] Because internal motion faster than τ_v would affect the effective electron relaxation rate τ_s^{-1}, we focus on cases where the correlation time τ_i for internal motion is significantly longer than τ_v. Under these conditions, incorporation of the correlation function $C_I(t)$ for internal motions into the correlation function for the PRE transforms $J_{SB}(\omega)$ in the SB equation into:

$$J_{SBMF}(\omega) = \langle r^{-6}\rangle\left\{\frac{S^2\tau_c}{1+\omega^2\tau_c^2} + \frac{(1-S^2)\tau_t}{1+\omega^2\tau_t^2}\right\} \tag{10}$$

where τ_t is the total correlation time defined as $(\tau_r^{-1} + \tau_s^{-1} + \tau_i^{-1})^{-1}$. When the original SB equations (eqs 2−4) are expressed in the form:

$$\Gamma_m = r^{-6}f_{SB,m}(\tau_c), \quad m = 1 \text{ or } 2 \tag{11}$$

incorporation of the model-free formalism transforms eq 11 into:

$$\Gamma_m = S^2\langle r^{-6}\rangle f_{SB,m}(\tau_c) + (1-S^2)\langle r^{-6}\rangle f_{SB,m}(\tau_t) \tag{12}$$

For simplicity, we refer to the SB equations incorporating the model-free formalism as the SBMF equations. The SBMF equations reduce to the SB equations under conditions where the internal motion is either very slow (and eventually, $\tau_t \approx \tau_c$) or highly restricted in space ($S^2 \approx 1$).

Comparing the SB (eq 11) and SBMF (eq 12) equations, one can ascertain that two additional effects need to be taken into account for the PRE arising from a flexible paramagnetic group attached to a macromolecule: the ensemble effect of $\langle r^{-6}\rangle$ and the motional effect related to S^2 and τ_i. If the ensemble space of the paramagnetic center is relatively large and it is hard to find a unique effective position for all PRE interaction vectors, over- or underestimation of the effective distance ($\langle r^{-6}\rangle^{-1/6}$) will lead to significant errors in the calculated values of Γ. For example, if an effective distance with a true value of 20 Å is overestimated by only 2 Å, the error in Γ will be as large as 44%. This error increases as the distance decreases. As shown below, however, the ensemble effect of $\langle r^{-6}\rangle$ can be readily treated by ensemble averaging using a multiple-structure representation for the paramagnetic group.

The impacts of motional effects on $^1H-\Gamma_1$ and $^1H-\Gamma_2$ are very different. Because of the large distances involved, the variation in the values of the order parameters for PRE interaction vectors is rather small. However, the internal motion correlation time τ_i is likely to have a wide range of values depending on the internal dynamics of individual ^1H nuclei in the macromolecule

(8) Mispelter, J.; Momenteau, M.; Lhoste, J.-M. In *Biological Magnetic Resonance: NMR of Paramagnetic Molecules*; Berliner, L. J., Reuben, J., Eds.; Plenum Press: New York, 1993; Vol. 12, pp 299−355.
(9) Ma, L.; Jorgensen, A.-M. M.; Sorensen, G. O.; Ulstrup, J.; Led, J. J. *J. Am. Chem. Soc.* **2000**, *122*, 9473−9485.
(10) Rubinstein, M.; Baram, A.; Luz, Z. *Mol. Phys.* **1971**, *20*, 67−80.
(11) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(12) Olejniczak, E. T.; Dobson, C. M.; Karplus, M.; Levy, R. M. *J. Am. Chem. Soc.* **1984**, *106*, 1923−1930.
(13) Brüshweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. *J. Am. Chem. Soc.* **1992**, *114*, 2289−2302.

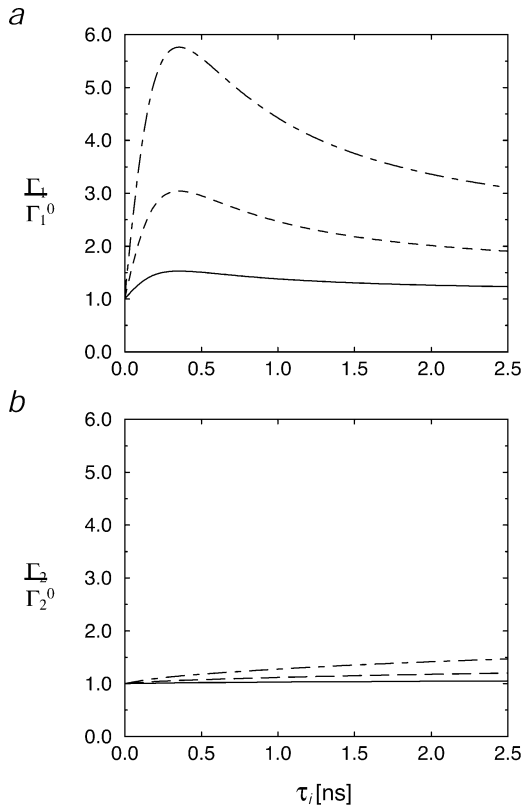

Figure 1. Influence of the correlation time τ_i for internal motion on (a) ^{1}H-Γ_1 and (b) ^{1}H-Γ_2 at 500 MHz. The vertical axis displays the Γ/Γ^0 ratio, where Γ is calculated with the full SBMF equation (eq 12) and Γ^0 is calculated with only the first term of the SBMF equation. Γ^0 corresponds to Γ when $\tau_i \rightarrow 0$. Curves, calculated with the correlation time τ_c [$=(\tau_r^{-1} + \tau_s^{-1})^{-1}$] set to 3 ns, are shown with three different values of the order parameter S^2 for PRE interaction vectors: $S_{Mn-H}^2 = 0.9$ (solid line), 0.7 (dotted line), and 0.5 (dashed line). The dependencies of ^{1}H-Γ_1 and ^{1}H-Γ_2 on τ_i are field-dependent, and the maximum of the ^{1}H-Γ_1/^{1}H-Γ_1^0 curve is located around $\tau_i \approx 1/\omega_I$.

(e.g., side chain versus backbone). The contribution of τ_i can be evaluated using the second term of the SBMF equation (eq 12). The dependencies of ^{1}H-Γ_1 and ^{1}H-Γ_2 on τ_i are shown in Figure 1. In the case where the order parameters S^2 for the PRE interaction vectors are as small as 0.5, ^{1}H-Γ_1 exhibits a strong dependence on τ_i (in particular for $\tau_i < 1/\omega_H$); indeed, the value of Γ_1 can be up to 5 times larger than that of Γ_1 in the absence of internal motion, depending on the value of τ_i (Figure 1a). The ^{1}H-Γ_2, on the other hand, is significantly less sensitive to the internal correlation time τ_i (Figure 1b). Thus, for structure analysis using ^{1}H-PRE data arising from flexible paramagnetic groups, the analysis of ^{1}H-Γ_1 requires accurate information on the values of τ_i for the individual PRE interaction vectors, whereas a simple estimation of τ_i is sufficient for analysis of ^{1}H-Γ_2 data. Consequently, ^{1}H-Γ_2 data provide far more useful structural restraints because they are not affected by the motional effects within the macromolecule, particularly those involving side chains. In addition, proton T_1 relaxation is not a single-exponential process because of cross-relaxation (and water-exchange in the case of amide protons), and hence the observed ^{1}H-Γ_1 values will also be perturbed by such effects. For these reasons, we restrict ourselves to the use of ^{1}H-Γ_2 data for structure calculation and only make use of ^{1}H-Γ_1 data to estimate the apparent values of the correlation times.

The SB (eq 11) and SBMF (eq 12) equations are valid for systems with isotropic tumbling. For nonglobular systems exhibiting significant diffusion anisotropy, the PRE also depends on the angles between the principal axes of the diffusion tensor and the electron−nucleus interaction vector. The effect of anisotropic tumbling can be readily incorporated according to previous literature.[14] In this Article, however, we have not pursued this effect because the calculated PRE will be minimally perturbed under conditions where the electron relaxation time dominates the overall correlation time (i.e., the electron relaxation time is significantly shorter than the apparent rotational correlation time, which is generally the case). Thus, for the purposes of this paper, we restrict ourselves to the use of eqs 11 and 12.

Computational Strategy. Our goal is to appropriately use PRE data arising from flexible paramagnetic groups as structural restraints in simulated annealing calculations. In back-calculating other NMR parameters (e.g., NOEs and dipolar couplings), two averaging techniques have been proposed to deal with flexibility: time-averaging during the course of restrained molecular dynamics trajectories[15] and ensemble-averaging.[16] Here, we adopt the latter approach. Back-calculation of ^{1}H-PRE based on the theory above can be achieved using a multiple-conformer representation for the paramagnetic group. The target function for the PRE restraints, E_{PRE}, is defined as:

$$E_{PRE} = k_{PRE} \sum_i w_i \{\Gamma_2^{obs}(i) - \Gamma_2^{calc}(i)\}^2 \qquad (13)$$

where k_{PRE} is a force constant, w_i is a weighting factor that is defined for each restraint, $\Gamma_2^{obs}(i)$ and $\Gamma_2^{calc}(i)$ are the observed and calculated values, respectively, and the summation is over individual ^{1}H-Γ_2 PRE data. The weighting factor should be based on the individual experimental errors. We choose weighting factors w_i to be:

$$w_i = \frac{1}{\delta_i^2} \frac{\Gamma_2^{obs}(i)}{\Gamma_2^{obs,max}} \qquad (14)$$

where δ_i is the experimental error associated with the measurement of $\Gamma_2^{obs}(i)$, and $\Gamma_2^{obs,max}$ is the maximum observed value of Γ_2 in the same data set. For data points with large values of Γ_2^{obs}, which have the most impact in structure refinement, the experimental errors are significantly larger because of lower sensitivity due to faster relaxation. Consequently, noninformative data points with small values of Γ_2^{obs} would be overly weighted if the weighting factors were simply set to $1/\delta_i^2$. The second scaling by $\Gamma_2^{obs}(i)/\Gamma_2^{obs,max}$ in eq 14 compensates for this. For CH_2 or NH_2 protons that exhibit two resolved peaks but are not stereospecifically assigned, the PRE target function uses the summation and difference of the two PRE values, employing

(14) (a) Woessner, D. E. *J. Chem. Phys.* **1962**, *37*, 647−654. (b) Benetis, N.; Kowalewski, J. *J. Magn. Reson.* **1985**, *65*, 13−33.
(15) (a) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289−294. (b) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223−235. (c) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991**, *220*, 457−479. (d) Torda, A. E.; Brunne, R. M.; Huber, T.; Kessler, H.; van Gunsteren, W. F. *J. Biomol. NMR* **1993**, *3*, 55−66.
(16) (a) Brüschweiler, R.; Blackledge, M.; Ernst, R. R. *J. Biomol. NMR* **1991**, *1*, 3−11. (b) Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. *J. Biomol. NMR* **1994**, *4*, 143−149. (c) Bonvin, A. M. J. J.; Brünger, A. T. *J. Mol. Biol.* **1995**, *250*, 80−93. (d) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923−2938.

the same technique that was previously used for nonstereospecifically assigned proton chemical shift restraints.[17]

To account for the finite conformational space occupied by a flexible paramagnetic group, N different conformers are used to represent the paramagnetic group in the calculations, while the remainder of the molecule is represented by a single structure. Individual conformers are treated independently under the restraints of the ^{1}H-PRE target function and can overlap with each other because interactions between the conformers are excluded from the nonbonded interaction energy term. The N conformers represent a discrete jump model where fractions of individual sites and the transition probability over a long period of time are uniform and equal to $1/N$. Because members of the ensemble can overlap structurally, this model is also valid for a case with fewer sites and uneven distributions, providing N is large enough. To obtain the ensemble average of Γ_2, the ensemble-averaged quantity $\langle r^{-6}\rangle$ for the electron−proton distance r must be calculated as follows:

$$\langle r^{-6}\rangle = \frac{1}{Nn_p}\sum_{h}^{N}\sum_{s}^{n_p} r_{hs}^{-6} \tag{15}$$

where n_p is the number of equivalent protons (i.e., 1 for NH or CH; 2 for equivalent aromatic protons that are degenerate because of fast ring flipping; 3 for methyl protons).

In simulated annealing calculation, we back-calculate Γ_2 with $\langle r^{-6}\rangle$ using two different approaches. The first is an approximation using the SB equation (eq 11) and the apparent correlation time τ_c^{app} given by:[4]

$$\tau_c^{app} = \frac{1}{\omega_H}\sqrt{\frac{3}{2}\frac{\Gamma_1}{\Gamma_2} - \frac{7}{4}} \tag{16}$$

For flexible paramagnetic groups, τ_c^{app} is influenced by the order parameter S^2 and the internal motion correlation time τ_i and is always smaller than the true value of τ_c in the SBMF equation (eq 9). Because, as shown later, S^2 for the PRE interaction vector tends to be relatively large due to the long distances involved, we examine the case of S^2 values between 0.5 and 0.9. When the correlation time for internal motions of the PRE interaction vector is larger than 50 ps (which is probably the case for most paramagnetic groups judging from their size), this approximation results in errors in ^{1}H-Γ_2 up to 11%, 21%, and 26% for S^2 values of 0.9, 0.7, and 0.5, respectively. However, the maximum corresponding error in the effective distances (proportional to $\Gamma_2^{-1/6}$) is only 4%. Thus, this approximation is reasonable and will not result in any large geometrical distortions in a structure calculation. We refer to this approach as the SB mode.

The second approach for back-calculating Γ_2 involves the direct application of the SBMF equation (which we refer to as the SBMF mode). This requires the explicit use of order parameters S^2 for PRE interaction vectors as defined by eq 6. These order parameters are calculated directly from the N-site discrete jump model of the paramagnetic group used in the simulated annealing calculations, assuming that positional fluctuations of the analyzed ^{1}H nucleus are small as compared to the paramagnetic center−^{1}H distance and hence do not contribute to S^2 for the ^{1}H-PRE interaction vector. For the N-site jump model, the order parameters are calculated as follows:

$$S^2_{PRE,angular} = \frac{1}{N^2 n_p^2}\sum_{h,k}^{N}\sum_{s,t}^{n_p}\left\{\frac{3}{2}\left(\frac{\vec{r}_{hs}\cdot\vec{r}_{kt}}{r_{hs}r_{kt}}\right)^2 - \frac{1}{2}\right\} \tag{17}$$

$$S^2_{PRE,radial} = \frac{\left(\sum_{h}^{N}\sum_{s}^{n_p} r_{hs}^{-3}\right)^2}{Nn_p\sum_{h}^{N}\sum_{s}^{n_p} r_{hs}^{-6}} \tag{18}$$

It should be noted that in this case S^2 is also a function of the coordinates of the ^{1}H nuclei and individual paramagnetic centers. Because of the increased complexity, the SBMF mode of calculation is computationally more expensive than the SB mode.

The correlation time τ_c can be readily optimized within a predefined range (between τ_c^{min} and τ_c^{max}) during simulated annealing using three pseudo-atoms, TO, TA, and TB:

$$\tau_c = \frac{1}{2}\{\tau_c^{min} + \tau_c^{max} + (\tau_c^{min} - \tau_c^{max})\cos\theta\} \tag{19}$$

where θ is the angle between vectors TO−TA and TO−TB. The coordinates of the pseudo-atoms are optimized along with those of all other atoms during the course of simulated annealing and minimization.

Experimental Section

PRE Measurements on SRY/DNA−EDTA−Mn^{2+} Complexes. Three DNA fragments with dT-EDTA at different sites were used for the SRY/DNA−EDTA−Mn^{2+} complexes (Figure 2). NMR samples of these SRY/DNA−EDTA complexes chelating either Ca^{2+} or Mn^{2+} were prepared as described previously.[4,18] The complexes at low concentrations were extensively washed with a high ionic strength buffer of 20 mM Tris·HCl (pH 6.8) and 500 mM NaCl to completely remove excess divalent ions at sites other than dT-EDTA. The solution conditions for NMR measurements comprised 0.3 mM SRY/DNA complex, 20 mM Tris·HCl (pH 6.8), 20 mM NaCl, and 7% D_2O. All buffers were treated with Chelex-20 (Sigma) prior to use.

^{1}H-Γ_2 and ^{1}H-Γ_1 PREs on ^{15}N-attached protons were measured as described previously[4] at 500 MHz on Bruker DMX-500 spectrometers equipped with cryogenic probes. ^{1}H-Γ_2 PREs on ^{13}C-attached protons were measured at 750 or 800 MHz on Bruker DMX-750 or DRX-800 spectrometers using the experiments described in the Supporting Information. ^{1}H-Γ_2 values were determined from two time points (T_a and T_b) for transverse relaxation as follows:

$$\Gamma_2 = \frac{1}{T_b - T_a}\ln\frac{I_{Ca}(T_b)I_{Mn}(T_a)}{I_{Ca}(T_a)I_{Mn}(T_b)} \tag{20}$$

where I_{Ca} and I_{Mn} are the peak intensities for the Ca^{2+}- and Mn^{2+}-chelated states, respectively. It should be noted that effects of homonuclear J_{HH}-modulation on peak intensities are canceled out by using the identical times and taking ratios for the two states. The errors in Γ_2 were estimated using the equation:

$$\delta\Gamma_2 = \frac{1}{T_b - T_a}\sqrt{\left\{\frac{\sigma_{Ca}}{I_{Ca}(T_a)}\right\}^2 + \left\{\frac{\sigma_{Ca}}{I_{Ca}(T_b)}\right\}^2 + \left\{\frac{\sigma_{Mn}}{I_{Mn}(T_a)}\right\}^2 + \left\{\frac{\sigma_{Mn}}{I_{Mn}(T_b)}\right\}^2} \tag{21}$$

(17) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson., Ser. B* **1996**, *112*, 79−81.

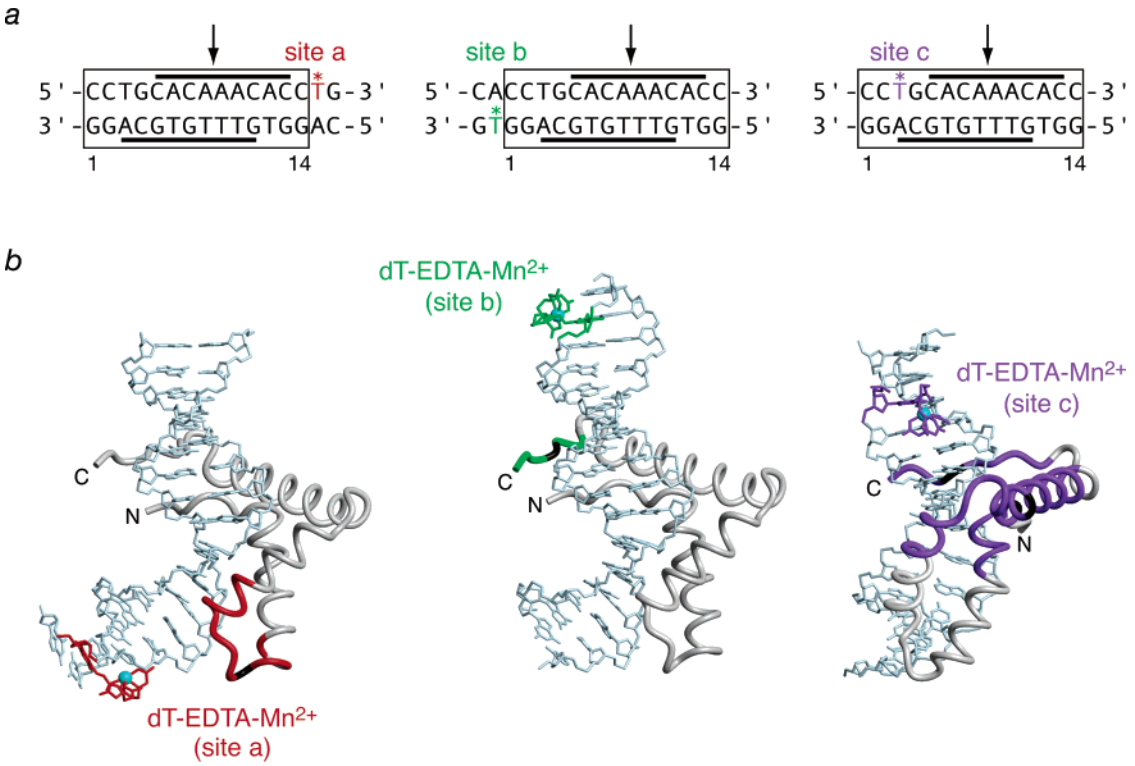

Figure 2. DNA fragments incorporating the dT-EDTA residue used in this study and overall summary of ^{1}H-PRE data recorded on the SRY/DNA complexes. (a) Sequences of the three oligonucleotide duplexes with the location of dT-EDTA indicated by an asterisk and color coded (red, site a; green, site b; purple, site c). The location of the SRY binding site on the top and bottom strands is indicated by the solid bars, and the site of partial intercalation of Ile13 is shown by an arrow.[18] The 14 base pairs common to all three oligonucleotides and used in the original structure determination of the SRY/DNA complex[18] are delineated by a rectangular box. (b) Residues that exhibit ^1H_N-Γ_2 values greater than 15 s^{-1} for dT-EDTA−Mn^{2+} at sites a, b, and c are shown in red, green, and purple, respectively, on a ribbon diagram representation of the SRY/DNA complex. (Residues depicted in black within these regions correspond to prolines). The Mn^{2+} ions are represented by cyan spheres. The coordinates of the SRY/DNA complex are taken from ref 18, and the additional base pairs at the ends of the original 14 base pair sequence were added as described previously.[4]

where σ_{Ca} and σ_{Mn} are the standard deviations of the noise in the spectra recorded with Ca^{2+} and Mn^{2+}, respectively.

To estimate the spin density of unpaired electrons delocalized onto the $2p_z$ orbital of nitrogen atoms, ^{15}N-Γ_1 was determined at 500 MHz on the complex with dT-EDTA−Mn^{2+} located at site c (Figure 2). The ^{15}N-T_1 relaxation rate measurements[19] were performed using the Ca^{2+}- and Mn^{2+}-chelated states of this complex.

Simulated Annealing with the PRE Target Function. The PRE target function was incorporated into the molecular structure determination package Xplor-NIH (which can be downloaded from http://nmr.cit.nih.gov/xplor-nih).[20] The module for the PRE target function was coded in C++, and the protocols for the structure calculations were written using the Python interface of Xplor-NIH. All simulated annealing calculations were carried out in torsion angle space using the IVM module[21] of Xplor-NIH. For simulated annealing calculations, a hybrid 18-bp DNA containing all three dT-EDTA−Mn^{2+} sites was used, and back-calculation of the three PRE data sets recorded on the three SRY/DNA complexes was performed simultaneously. Multiple structures for an EDTA−Mn^{2+} group and its linker region were used, whereas the rest of the molecule was represented by a single structure. The portion of the EDTA−Mn^{2+} group comprising the six coordination bonds to the metal ion was treated as a rigid body with coordinates taken from the EDTA−Mn^{2+} crystal structure,[22] while the linker region was given full torsional degrees of freedom. The van der Waals energies between members of the ensemble for the EDTA−Mn^{2+} moiety and its linker were excluded from the energy calculation during simulated annealing and minimization, thereby permitting atomic overlap of the ensemble members. The force-field parameters for the EDTA−Mn^{2+} group are provided in the Supporting Information.

The electron g-factor for dT-EDTA−Mn^{2+} was set to 2.0. To avoid over-weighting beyond the accuracy of the measurements, δ_i (eq 14) used in the calculation of E_{PRE} (eq 13) was set to a minimum value of 1.0 s^{-1} in those cases where the error in the measured ^{1}H-Γ_2 value was estimated to be smaller.

During the high-temperature (3000 K) stage of the simulated annealing calculations, PRE back-calculation was performed in the SB mode using the apparent correlation time τ_c^{app} calculated using eq 16. The force constant k_{PRE} in eq 13 was initially set to 0.05 kcal mol^{-1}. During the cooling phase (3000 → 25 K), the PRE back-calculation was switched to the SBMF mode and the value of k_{PRE} was gradually increased from 0.05 to 1.0 kcal mol^{-1} s^2 for the backbone and 0.025 to 0.5 kcal mol^{-1} s^2 for the side chains. The force constant for the side chains was set lower than that for the backbone because back-calculation of the PRE is expected to be less accurate for side chains due to larger effects of internal motions. At the beginning of the cooling phase, the range of τ_c for the SBMF equation was set to lie between $\tau_c^{app,min}$ and $\tau_c^{app,max}$ divided by the average value of the PRE order parameters S^2 calculated at the end of the high-temperature stage. Because of its low contribution to ^{1}H-Γ_2 (Figure 1), the second term of the SBMF equation for ^{1}H-Γ_2 was neglected under the assumption that $\tau_i \ll \tau_c$. This assumption was experimentally confirmed by fitting τ_i against ^{1}H-Γ_1 and Γ_2 data at multiple magnetic fields. The other conditions for the simulated annealing are the same as those in the previous publica-

(18) Murphy, E. C.; Zhurkin, V. B.; Louis, J. M.; Cornilescu, G.; Clore, G. M. *J. Mol. Biol.* **2001**, *312*, 481−499.
(19) Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E. *J. Biomol. NMR* **1994**, *4*, 727−734.
(20) Schwieters, C. D.; Kuszewski, J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 66−74.
(21) Schwieters, C. D.; Clore, G. M. *J. Magn. Reson.* **2001**, *152*, 288−302.
(22) Richards, S.; Pedersen, B.; Selverton, J. V.; Hoard, J. L. *Inorg. Chem.* **1964**, *3*, 27−33.

tions.[4,18] The agreement between calculated and observed values of ^{1}H-Γ_2 was evaluated using a PRE Q-factor[4] defined as:

$$Q = \sqrt{\frac{\sum_i \{\Gamma_2^{obs}(i) - \Gamma_2^{calc}(i)\}^2}{\sum_i \Gamma_2^{obs}(i)^2}} \quad (22)$$

Results

^{1}H-PRE for Three SRY/DNA Complexes with dT-EDTA−Mn^{2+} at Three Distinct Sites. We analyzed the three SRY/DNA complexes that share the same 14-bp DNA sequence and possess a dT-EDTA at three distinct sites a, b, and c as shown in Figure 2. The EDTA groups are covalently attached to the C5 position of a thymine base through a linker with five rotatable bonds (see Supporting Information for the chemical structure). Judging from the three-dimensional structure of the SRY/DNA complex (PDB accession code 1J46),[18] conjugation of EDTA at these sites should have no impact on complex formation. This was confirmed experimentally by the observation that the ^{1}H, ^{13}C, ^{15}N chemical shifts of the bound SRY protein are essentially identical to those for the original complex.[18] PREs on backbone and side-chain ^{1}H nuclei were measured using 2D HSQC-type experiments on the Mn^{2+}- and Ca^{2+}-chelated states. In the case of the complex with dT-EDTA−Mn^{2+} located at site c, ^1H_N-Γ_2 data were collected on two independently prepared samples and the observed ^{1}H-Γ_2 values agreed with each other within 2 times the estimated error δ in the measurement. The distributions of the observed ^{1}H-Γ_2 values arising from the dT-EDTA−Mn^{2+} at locations a, b, and c are quite different. Of the three sites, site c is situated closest to the bound SRY protein and yields the largest number of backbone amide protons with ^1H_N-$\Gamma_2 > 15$ s^{-1}; the most distantly locate site, site b, yields the smallest number of backbone amide protons with ^1H_N-$\Gamma_2 > 15$ s^{-1} (Figure 2). For the structure calculations, we used ^{1}H-PRE data that satisfied the following criteria: ^{1}H-$\Gamma_2 \geq 2\delta$ for PREs with ^{1}H-$\Gamma_2 \geq 6$ s^{-1}; $\delta < 3.0$ s^{-1} for PREs with ^{1}H-$\Gamma_2 < 6$ s^{-1}. The resulting ^{1}H-PRE data set comprised 435 restraints: 255 for backbone protons and 180 for side-chain protons.

Influence of Number of Conformations on Back-Calculation of the PRE. The ensemble approach described in the Theory section was examined by a series of simulated annealing calculations using ^{1}H-PRE restraints, changing the number of conformers employed to represent the EDTA−Mn^{2+} groups and their linkers. In this evaluation, we kept the coordinates of the original NMR structure of the SRY/DNA complex[18] (which includes the 14 bp delineated by the boxes in Figure 2) fixed to maintain the contribution of coordinate errors in all the calculations constant. The remaining portions of the DNA in the complex were given full torsional degrees of freedoms and optimized under the influence of the ^{1}H-PRE restraints, as well as nucleic acid torsion angle and DNA base−base positional database potentials of mean force.[23,24]

First, we carried out a set of calculations using only the PRE ^{1}H-Γ_2 data for the backbone amide protons to ensure that the analysis was not affected by any effects arising from large-scale internal motions. The assessment of the side-chain ^{1}H-Γ_2 data will be described below. The dependence of the calculated Q-factor on the number of conformers used to represent the covalently linked EDTA groups is shown in Figure 3a (closed circles). The overall Q-factors calculated using the multiple-conformer representations are clearly smaller than that obtained with single-conformer representation. When the ^{1}H-PRE data are further analyzed on the basis of the location of dT-EDTA−Mn^{2+} in the oligonucleotide sequence, two trends are apparent (Figure 3b, closed circles).

In the two cases where dT-EDTA−Mn^{2+} is located only one base pair in from the ends of the DNA (sites a and b), the multiple-conformer representation does not afford any improvement in the Q-factor (Figure 3b, red closed circles). Moreover, when the number of conformers is increased beyond four, the overall Q-factor increases slightly, presumably due to less effective minimization of the PRE target function as a consequence of the increased complexity of the energy hypersurface. Thus, the simplest explanation for these results is that the EDTA−Mn^{2+} groups at sites a and b are locked in a single conformation. Alternatively, as a consequence of a particular spatial relationship between the paramagnetic group and the protons of the protein, only a single narrow region of the ensemble space occupied by the paramagnetic group is responsible for the PRE observed on the majority of protons, and hence the position of the paramagnetic group can be described by a single effective position (see Discussion along with Figure 12).

When dT-EDTA−Mn^{2+}, on the other hand, is positioned as little as two base pairs in from the ends of the DNA (site c), the Q-factor is significantly reduced (by ca. 33% from 0.32 to 0.21) upon the introduction of a multiple-conformer representation (Figure 3b, blue closed circles). Interestingly, although the Q-factor for the ^1H_N-Γ_2 data obtained with dT-EDTA−Mn^{2+} at site c is reduced dramatically by increasing the number of conformers from one to three, further increases in conformer number yield no further improvements in Q-factor. In addition, the correlation between calculated and observed values of ^1H_N-Γ_2 for site c is nonlinear in the case of the single-conformer representation (Figure 4a), but very close to linear for the multiple-conformer representations with the correlations for 3- and 10-conformer representations being almost identical (Figure 4b,c). These results imply that the three-conformer model provides a realistic representation for EDTA−Mn^{2+} at site c located at the edge of the SRY binding site. The positions of the Mn^{2+} ion at site c in the calculated structures of the complex for the 1-, 3-, and 10-conformer models are shown in Figure 4d−f. When a three-conformer model for EDTA−Mn^{2+} at site c is employed, three distinct locations for the three Mn^{2+} ions are observed: two are close to the major groove surface and separated by ~11 Å; the third is far from the surface of the major groove (Figure 4e). In the 10-conformer model, the positions of Mn^{2+} can be grouped into three clusters that correspond to the three sites occupied in the three-conformer model (two on the major groove surface; one far from the surface) (Figure 4f).

Because inclusion of large numbers of EDTA−Mn^{2+} conformers in the back-calculation of the ^{1}H-PRE could result in over-fitting, we also performed a series of calculations with complete cross-validation,[24−26] randomly partitioning the ^1H_N-Γ_2 data into 10 pairs of working (70%) and test (30%) data

(23) Kuszewski, J.; Schwieters, C. D.; Clore, G. M. *J. Am. Chem. Soc.* **2002**, *123*, 3903−3918.
(24) Clore, G. M.; Kuszewski, J. *J. Am. Chem. Soc.* **2003**, *125*, 1518−1525.

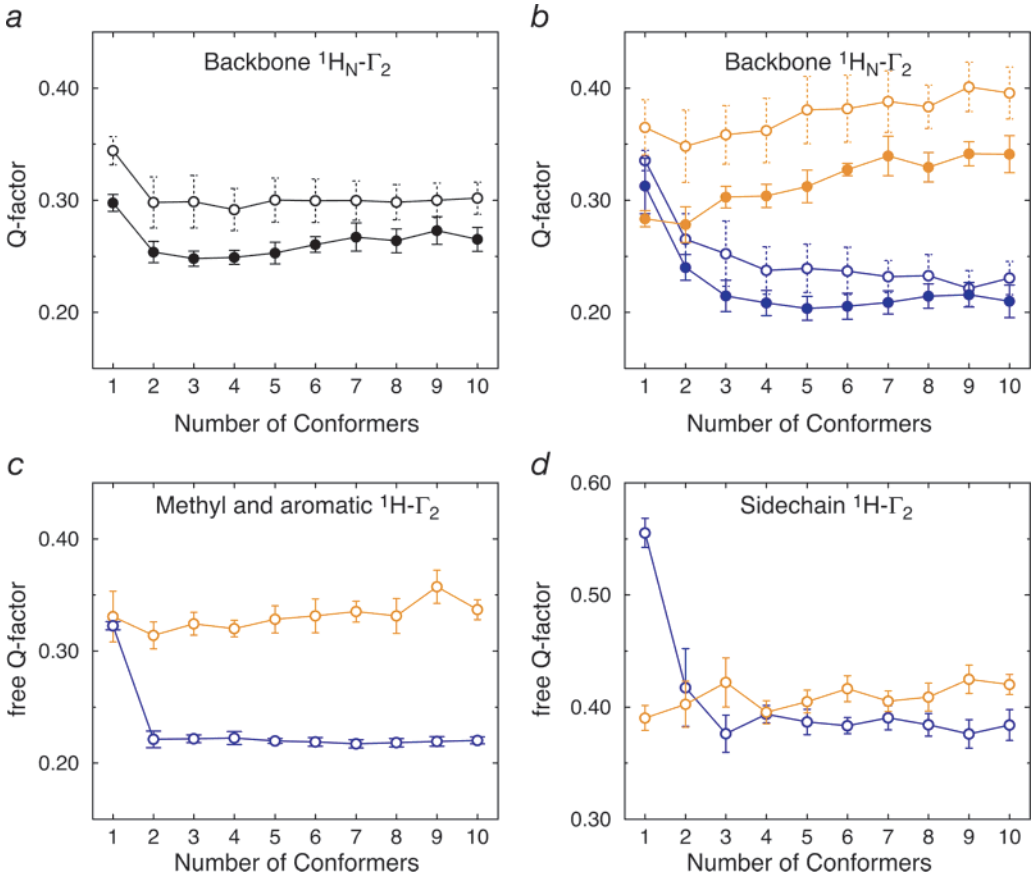

Figure 3. Influence of the number of EDTA−Mn^{2+} conformers on the PRE Q-factors. (a and b) Overall Q-factors (filled circles) and cross-validated free Q-factors (open circles) for backbone amide ^{1}H-Γ_2. (c) Free Q-factors for buried methyl and aromatic ^{1}H-Γ_2. (Note the different scale for the Q-factor axis in panel d). (d) Free Q-factors for overall side-chain ^{1}H-Γ_2. In panel a, the average Q-factors for all dT-EDTA−Mn^{2+} groups combined are shown, whereas in panels b, c, and d the Q-factors for dT-EDTA−Mn^{2+} groups are subdivided into two groups: those related to sites a and b, located at the edges of the DNA (cf. Figure 2), are shown in orange; those related to site c, located more internally in the DNA sequence, are displayed in blue. Ten structures were calculated using only backbone ^1H_N-Γ_2 data; SRY and the central 14 base pairs of DNA are held rigid, while the terminal two base pairs at either end together with the paramagnetic groups are given full torsional degrees of freedom; the values shown are the averages of the 10 calculations. Free Q-factors for ^1H_N-Γ_2 were calculated using complete cross-validation in which the ^1H_N-Γ_2 data set was randomly partitioned into 10 pairs of 70% working and 30% test groups; free Q-factors for side-chain ^{1}H-Γ_2 were calculated from structures determined using only ^1H_N-Γ_2 data. Error bars indicate standard deviations for the individual points. Note that the Q-factors for site c are dramatically reduced upon the introduction of a multiple-conformer representation for the paramagnetic group.

sets. Free Q-factors for the test sets (omitted from the refinement) were calculated for the structures refined against the corresponding working sets, and the average values of the free Q-factors for the 10 different test sets, plotted as a function of conformer number, are shown in Figure 3a and b (open circles). The trends observed for the free Q-factors are similar to those for the overall Q-factors. For site c, the free Q-factor is significantly reduced upon increasing the number of the conformers from one to two and from two to three. No change in the free Q-factor, however, is apparent for further increases in conformer number. We therefore conclude that the improved agreement between observed and calculated values of ^1H_N-Γ_2 as the number of conformers is increased to three is due to improved accuracy rather than to over-fitting.

Assessment of Back-Calculation of Side-Chain ^{1}H-Γ_2. On the basis of the ensembles of the paramagnetic groups derived from refinement against the backbone amide ^1H_N-Γ_2 data, free Q-factors for the side-chain ^{1}H-Γ_2 data were calculated and the agreement with the experimental data was evaluated (Figure 3c and d). This provides a good measure of accuracy of ^{1}H-Γ_2 back-calculations for side-chain protons. As seen in Figure 3d, the free Q-factors for all side chains are reasonably low, indicating that the accuracy of the back-calculation is good. Indeed, when only buried methyl and aromatic groups are considered, the free Q-factors are as low as those for the backbone amide protons (cf., compare Figure 3c to 3b). In the case of the ^{1}H-Γ_2 data derived from site c, the decrease in the free Q-factors for the side chains upon increasing the number of conformers follows that observed for the backbone amide protons.

The poorer agreement with the experimental ^{1}H-PRE data for exposed side-chain protons is due to two factors. First, only the paramagnetic groups were allowed to take on multiple conformations. Although motional effects on ^{1}H-Γ_2 related to S^2 and τ_i are not significant for side chains as shown in Figure 1b, the calculation of $\langle r^{-6} \rangle$ for a side-chain proton could be inaccurate if the finite conformational space sampled by that proton is not taken into account. Second, the accuracy of the exposed surface side-chain coordinates themselves is intrinsically lower, and cross-validation was performed without altering

(25) Brünger, A. T.; Clore, G. M.; Gronenborn, A. M.; Saffrich, R.; Nilges, M. *Science* **1993**, *261*, 328−331
(26) Clore, G. M.; Garrett, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 9008−9012.

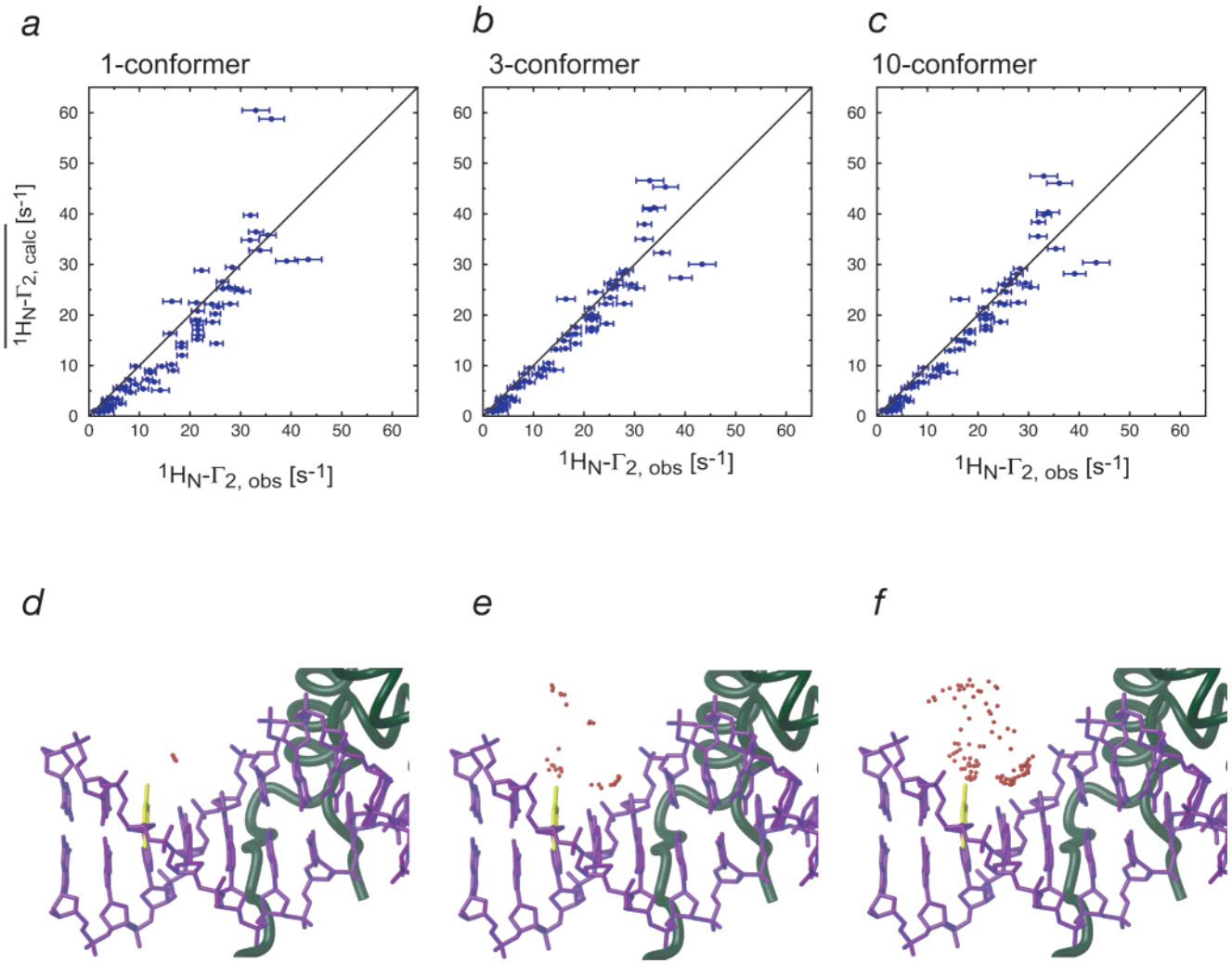

Figure 4. Correlation between observed and calculated values of backbone ^1H_N-Γ_2 arising from the dT-EDTA−Mn^{2+} at site c obtained with (a) single, (b) 3-, and (c) 10-conformer models of the EDTA−Mn^{2+} group. The calculated values represent the average for 10 calculated ensembles of simulated annealing structures. The experimental errors in the observed values are shown as horizontal bars. The correlation coefficients between observed and calculated values for (a), (b), and (c) are 0.89, 0.94, and 0.94, respectively. The location of Mn^{2+} (represented as red spheres) at site c in the structures of the SRY/DNA complex calculated with the (d) 1-, (e) 3-, and (f) 10-conformer representations of the flexible paramagnetic group. The DNA (purple) and protein backbone (green tube) are shown for a single structure, while the Mn^{2+} ion (red spheres) is shown as a superposition from 10 ensembles of simulated annealing structures (i.e., there are 10, 30, and 100 Mn^{2+} ions superimposed in panels d, e, and f, respectively).

the original structure of the SRY/DNA complex. This is investigated below by direct refinement against all ^1H-Γ_2 data.

Order Parameters and Effective Distances. The order parameters and effective distances of the ^1H-PRE interaction vectors are not independent of each other. Rather, there is a strong correlation between them as shown in Figure 5a. As the distance increases so the angles between the interaction vectors of individual states are reduced, the values $\langle r^{-3}\rangle^2$ and $\langle r^{-6}\rangle$ become closer, and hence $S^2_{\mathrm{PRE,radial}} \rightarrow 1$ (eq 18). Similarly, $S^2_{\mathrm{PRE,angular}} \rightarrow 1$ for large distances (cf. eq 17). Consequently, the longer the effective distance, the larger the PRE order parameter will be. Because the PRE interaction vectors observed in the SRY/DNA complex are long, the corresponding order parameters are relatively large with small variations ($0.68 \leq S^2_{\mathrm{PRE}} \leq 0.96$) despite the large spatial distributions of the paramagnetic centers.

As shown above, back-calculation of the ^1H-Γ_2 data for site c with 3- and 10-conformer models results in almost the same values of working and free Q-factors. The question then arises as to whether the corresponding ensembles are equivalent in terms of order parameters and effective distances. Figure 5b compares the order parameters for the ensembles with the 3- and 10-conformer representations for EDTA−Mn^{2+} at site c. Despite the different numbers of conformers, the values of the order parameters for these two ensembles are actually close to each other (Figure 5b), although those for the 10-conformer model are systematically smaller than those for the 3-conformer model, probably due to increased conformational freedom (see Figure 4). In addition, the effective $\langle r^{-6}\rangle^{-1/6}$ distances are essentially identical for the 3- and 10-conformer models (Figure 6a). This is in clear contrast to the comparison between the three-conformer and single-conformer representations (Figure 6b). Considering the similarity of both the order parameters and the effective distances, we conclude that ensembles with 3- and 10-conformer representations for the PRE data set obtained for site c are in essence completely equivalent.

Structure Refinement with ^1H-PRE Restraints Based on the Ensemble Approach. The arguments thus far are based on the coordinates of the SRY/DNA complex that had been determined with extensive use of other NMR-derived data

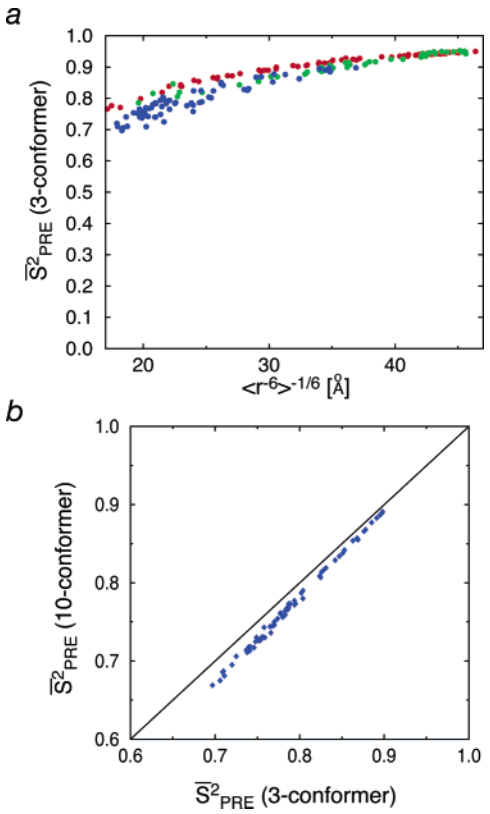

Figure 5. Order parameters for Mn−^1H_N vectors (S_{Mn-H}^2) for the SRY/DNA complexes with dT-EDTA−Mn^{2+} located at sites a (red), b (green), and c (purple). (a) Dependence of S_{Mn-H}^2 on the effective distance $\langle r_{Mn-H}^{-6}\rangle^{-1/6}$ calculated using eqs 7, 15, 17, and 18 for a 10-state discrete jump model. Note that S_{Mn-H}^2 decreases as the effective distance is reduced. (b) Correlation between S_{Mn-H}^2 calculated for 3- and 10-conformer discrete jump models.

including 375 residual dipolar coupling (RDC) restraints for both protein and DNA.[18] The low values of the PRE Q-factors for the backbone ^1H_N protons and buried methyl and aromatic protons for the structure (Figure 3a−c) are indicative of good coordinate accuracy for these regions. The overall free Q-factor, however, for the side-chain protons is quite high (Figure 3d), which suggests that further improvements in coordinate accuracy are possible. We therefore performed further structure refinement using the complete set of backbone and side-chain ^{1}H-PRE restraints based on the ensemble approach with a three-conformer representation, allowing full torsional degrees of freedom for the entire complex. In this calculation, all other NMR-derived restraints that had been used in the original structure determination[18] were also included. A stereoview showing a best-fit superposition of 40 simulated annealing structures refined with ^{1}H-PRE restraints on the restrained regularized mean coordinates derived from the simulated annealing structures calculated without ^{1}H-PRE restraints is shown in Figure 7. We also carried out a set of simulated annealing calculations without RDC restraints. Structural statistics and coordinate rms differences are provided in Tables 1 and 2, respectively. As seen in Table 2, the structures calculated with and without ^{1}H-Γ_2 PRE data are almost identical in the presence of the RDC restraints: the atomic rms difference between the two sets of mean coordinates is ~0.6 Å for the protein backbone (residues 4−79 of SRY) and DNA (central 12 bp) heavy atoms. Refinement against the ^{1}H-PRE data does not affect the agreement with the other experimental NMR restraints, all of which are satisfied within their experimental errors (Table 1). There is a significant improvement in the agreement between observed and calculated values of ^{1}H-Γ_2: the overall PRE Q-factors for backbone and side-chain protons after full refinement were 0.20 ± 0.01 and 0.26 ± 0.01, respectively, as compared to values of 0.24 ± 0.01 and 0.39 ± 0.02, respectively, when only the coordinates of the paramagnetic group are optimized. Thus, upon full refinement, the agreement between calculated and observed ^{1}H-Γ_2 for the side chains is significantly improved and is almost comparable to that for the backbone, as can be seen from the correlation plots displayed in Figure 8. In addition, the presence or absence of RDC restraints does not impact the agreement with the PRE data (Table 1).

Although the impact on the mean coordinate positions as a consequence of including PRE restraints is comparable to that of including the RDC restraints, inclusion of PRE restraints does not result in an increase in coordinate precision (Table 2). This is due to the fact that the PRE restraints are based on long (>17 Å) distances that can easily allow for subtle changes in the coordinates. The PRE data, however, can result in significant improvements in coordinate accuracy as shown in a later section. In contrast, inclusion of RDC restraints, which permit the orientations of short internuclear vectors relative to an external

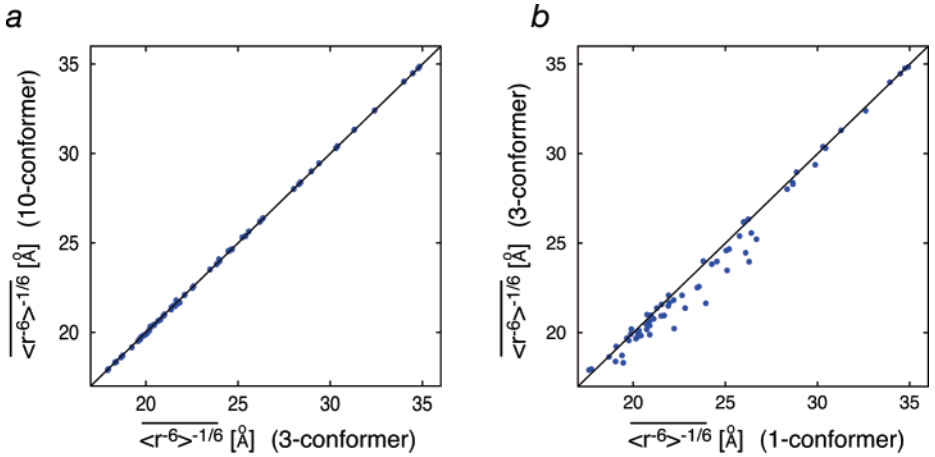

Figure 6. Comparison of Mn^{2+}−^1H_N effective distances $\langle r^{-6}\rangle^{-1/6}$ for 1-, 3-, and 10-conformer representations of the flexible paramagnetic group. The values plotted represent the averages of 10 independent calculations on the SRY/DNA complex with dT-EDTA−Mn^{2+} located at site c.

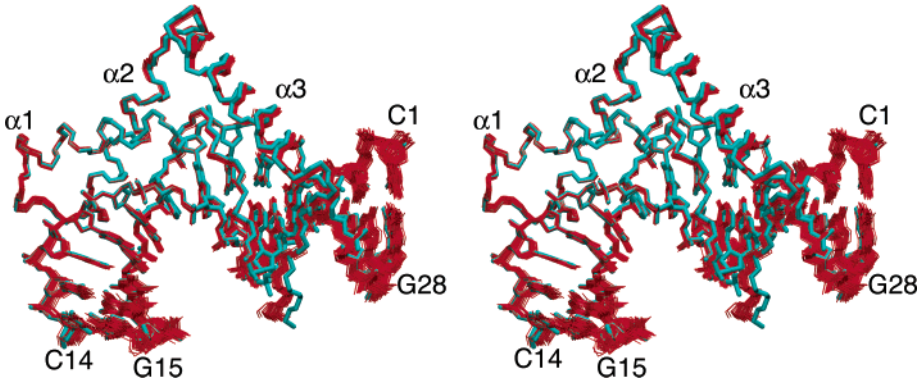

Figure 7. Stereoview of a best-fit superimposition of 40 simulated annealing structures refined against the complete set of 435 ^{1}H-PRE restraints (red) on the restrained regularized mean coordinates generated from structures calculated without ^{1}H-PRE restraints (cyan). The protein backbone of residues 4−79 of SRY and the DNA heavy atoms (central 14 base pairs) are displayed.

Table 1. Structural Statistics for the 40 Simulated Annealing Structures Calculated with and without ^{1}H-PRE Restraints[a]

	⟨SA$_{PRE,RDC}$⟩	⟨SA$_{RDC}$⟩	⟨SA$_{PRE}$⟩	⟨SA⟩
		^{1}H-Γ_2 Q-factors		
overall (438)	0.23 ± 0.01		0.22 ± 0.01	
backbone (258)	0.20 ± 0.00		0.20 ± 0.01	
side chain (180)	0.26 ± 0.01		0.25 ± 0.01	
		Root Mean Square Deviation from Other Experimental Restraints		
distance restraints (Å) (1755)	0.04 ± 0.00	0.04 ± 0.00	0.04 ± 0.00	0.03 ± 0.00
torsion angle restraints (deg) (433)	0.58 ± 0.06	0.36 ± 0.01	0.51 ± 0.05	0.29 ± 0.03
$^3J_{HNH\alpha}$ coupling restraints (Hz) (70)	0.86 ± 0.03	0.84 ± 0.02	0.67 ± 0.02	0.64 ± 0.02
^{13}Cα/^{13}Cβ shift restraints (ppm) (165)	1.02 ± 0.01	1.02 ± 0.01	1.01 ± 0.01	1.01 ± 0.01
DNA D_{HH} dipolar couplings (Hz) (55)	0.52 ± 0.01	0.57 ± 0.01	1.25 ± 0.12[c]	1.54 ± 0.11[c]
		Fixed Distance Dipolar Coupling R-factors (%)[b]		
protein $^1D_{NH}$ (71)	5.4 ± 0.3	4.4 ± 0.3	26.5 ± 1.8[c]	24.2 ± 2.3[c]
protein $^1D_{CH}$ (67)	5.0 ± 0.5	5.2 ± 0.5	22.4 ± 1.8[c]	22.4 ± 2.1[c]
protein $^1D_{NC'}$ (68)	22.5 ± 1.4	20.5 ± 2.1	28.5 ± 2.2[c]	25.9 ± 2.2[c]
protein $^2D_{HNC'}$ (68)	20.1 ± 0.3	19.8 ± 0.5	29.0 ± 1.4[c]	27.7 ± 1.5[c]
DNA $^1D_{NH}$ (9)	10.8 ± 0.2	10.0 ± 0.5	28.5 ± 7.4[c]	28.4 ± 6.7[c]
DNA $^1D_{CH}$ (37)	13.0 ± 0.2	12.3 ± 0.2	42.2 ± 7.0[c]	35.1 ± 3.4[c]
		Root Mean Square Deviation from Idealized Covalent Geometry		
bonds (Å)	0.006 ± 0.000	0.004 ± 0.000	0.006 ± 0.000	0.003 ± 0.000
angles (deg)	0.90 ± 0.00	0.74 ± 0.00	0.82 ± 0.01	0.65 ± 0.01
improper torsions (deg)	0.60 ± 0.02	0.60 ± 0.01	0.44 ± 0.02	0.39 ± 0.03

[a] Four sets of 40 simulated annealing structures were calculated using the full set of NOE, torsion angle, ^{13}C chemical shift, and ^{3}J coupling restraints,[18] and are represented with the following notation: ⟨SA$_{PRE,RDC}$⟩, structures calculated with PRE and RDC restraints; ⟨SA$_{RDC}$⟩, with RDC but without PRE restraints; ⟨SA$_{PRE}$⟩, with PRE but without RDC restraints; ⟨SA⟩, without PRE and RDC restraints. The structures of the SRY/DNA complex reported previously[18] correspond to ⟨SA$_{RDC}$⟩. The PRE restraints were applied using three-conformer representation for individual dT-EDTA−Mn^{2+} groups. The numbers in parentheses indicate the number of restraints in each category. The simulated annealing protocol employed is the same as that used in ref 18. [b] The dipolar coupling R-factor, R_{dip}, is given by the ratio of the rms difference between observed and calculated dipolar couplings and the expected value of the rms difference if the vectors are randomly oriented. The latter is given by $[2\lambda D_a^{NH}(4 + 3\eta^2)/5]^{1/2}$, where D_a^{NH} is the magnitude of the axial component of the alignment tensor for the N−H dipolar couplings, η is the rhombicity, and λ is a normalization factor for the other dipolar couplings given by $\gamma_m\gamma_n r_{NH}^3/\gamma_N\gamma_H r_{mn}^3$, where γ are the gyromagnetic ratios for the different nuclei (proton, nitrogen, or carbon), and the bond distances r_{mn} are taken from ref 34. The values of D_a^{NH} and η are taken from ref 18. [c] The dipolar coupling R-factors and the rms for the D_{HH} dipolar couplings for the two sets of structure calculated without dipolar couplings, ⟨SA$_{PRE}$⟩ and ⟨SA⟩, are determined after the structure calculations by optimizing the orientation of the alignment tensor.

alignment tensor to be defined, results in large increases in precision (Table 2).

Comparison of Single- and Three-Conformer Representations for dT-EDTA−Mn^{2+} Groups in Simulated Annealing. In the previous section on structure refinement, we employed a three-conformer representation for all dT-EDTA−Mn^{2+} groups, because cross validation with the coordinates of the complex held fixed indicated that three conformers were sufficient to describe the site c data and did not compromise agreement with the sites a and b data, even though the latter could be accounted for by a single conformer (Figure 4). What are the structural consequences if a single-conformer representation is used for all three sites instead? This is critical because it is self-evident that the purpose of adding PRE restraints into NMR structure refinement is to increase rather than decrease the coordinate accuracy. The results of the complete cross-validation calculations shown in Figure 4 in which the original RDC-based NMR coordinates of the complex (SRY plus central 14 DNA base pairs) were held rigid clearly indicate that the PRE Q-factor is significantly lower for the three-conformer representation than the single-conformer representation of the EDTA−Mn^{2+} groups as a consequence of the EDTA−Mn^{2+} group at site c occupying multiple positions. Thus, if the SRY/DNA complex is given full torsional degrees of freedom, one would predict that refinement using a single-conformer representation would only result in a reduction in the PRE Q-factor at the expense of coordinate accuracy for the SRY/DNA complex; that is to say the inappropriate treatment of the PRE data originating from site c would be predicted to distort the structure of the complex.

Table 2. Comparison of the NMR Structures[a] of the SRY/DNA Complex Refined with and without PRE Restraints

	atomic rms difference (Å)		
	protein[b] backbone + DNA[c] heavy atoms	protein[b] heavy atoms	protein[b] heavy atoms + DNA[c] heavy atoms
	Coordinate Precision[d]		
$\langle SA_{PRE,RDC} \rangle$	0.21 ± 0.06	0.66 ± 0.07	0.53 ± 0.06
$\langle SA_{RDC} \rangle$	0.16 ± 0.04	0.72 ± 0.05	0.57 ± 0.04
$\langle SA_{PRE} \rangle$	0.46 ± 0.16	0.79 ± 0.08	0.70 ± 0.11
$\langle SA \rangle$	0.40 ± 0.14	0.81 ± 0.07	0.69 ± 0.08
	Impact of PRE Restraints on Mean Coordinates[e]		
$\overline{SA}_{PRE,RDC}$ vs $\overline{SA}_{RDC}$	0.61	0.66	0.72
$\overline{SA}_{PRE}$ vs $\overline{SA}$	0.90	0.84	0.95
	Impact of RDC Restraints on Mean Coordinates[e]		
$\overline{SA}_{PRE,RDC}$ vs $\overline{SA}_{PRE}$	1.02	1.04	1.07
$\overline{SA}_{RDC}$ vs $\overline{SA}$	0.88	0.93	0.96

[a] The notation for the simulated annealing structures is the same as that in Table 1. The bar indicates the mean coordinates. [b] Residues 4−79 of SRY. [c] Central 12-bp of the DNA. [d] Average atomic rms difference between the individual 40 simulated annealing structures and the corresponding mean coordinates. [e] Root mean square difference between mean coordinates.

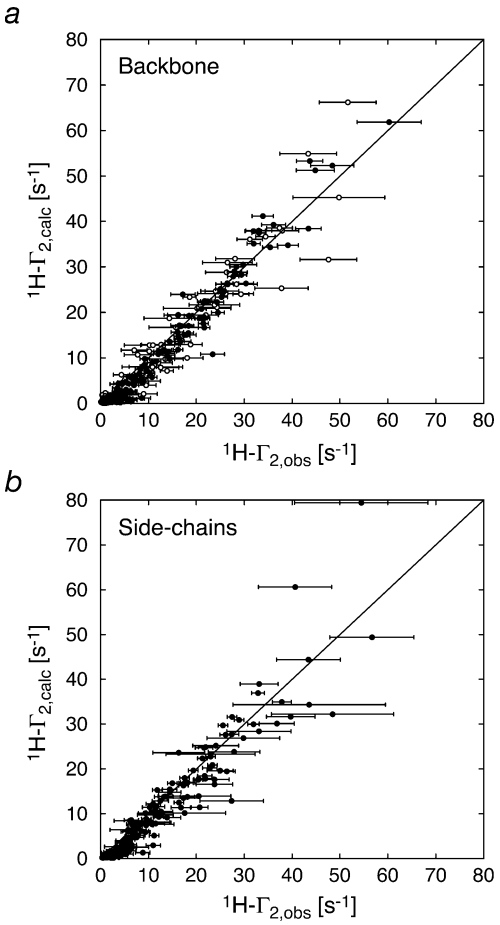

Figure 8. Correlations between the observed and calculated values of ^{1}H-Γ_2 for backbone (a) and side-chain (b) ^{1}H-PREs after refinement. A three-conformer representation for the paramagnetic groups was employed for the ^{1}H-PRE back-calculation. In panel a, data for ^1H_N and $^1H\alpha$ protons are represented by filled-in and open circles, respectively.

Using a single-conformer representation for all EDTA−Mn^{2+} groups results in significant deviations in the SRY coordinates from the original RDC-based structure.[18] In the plot shown in Figure 9a (top panel), this is manifested when the backbone rms difference between the two mean coordinates (with and without PRE restraints) is substantially larger than the sum of the precision of the two sets of coordinates (as measured by the average rms difference between the members of each ensemble of simulated annealing structures and their respective mean coordinate positions).[26] In contrast, no substantial differences in backbone coordinates, with the possible exception of the C-terminal three residues, are observed between the structures refined with and without the PRE data when the three-conformer representation of the EDTA−Mn^{2+} groups is employed (Figure 9a, bottom panel).

Table 3. Free Residual Dipolar Coupling R-factors for Residues 9−14, 37−46, and 66−74 of SRY for Structures Calculated with PRE Data Using the Single- or Three-Conformer Representations for EDTA−Mn^{2+} at Site C, and with No PRE Data[a]

	free dipolar coupling R-factor (%)		
dipolar coupling	one-conformer	three-conformer	without PRE
$^1D_{CH}$	34.1 ± 2.9	21.8 ± 3.4	20.4 ± 4.5
$^1D_{NH}$	25.1 ± 1.8	20.3 ± 1.5	21.3 ± 3.3
$^1D_{NC'}$	33.3 ± 1.4	27.5 ± 1.9	24.1 ± 2.5
$^2D_{HNC'}$	28.5 ± 1.0	25.9 ± 1.3	27.5 ± 1.3

[a] For each set of calculations, 32 simulated annealing structures were calculated excluding the residual dipolar couplings for the three regions of SRY, residues 9−14, 37−46, and 66−74, which exhibit significant structural differences between the one- and three-conformer representations (see Figure 9e). The free R-factors (for the dipolar couplings left out of the calculations) were calculated using the principal axes of the alignment tensor determined in the simulated annealing calculations. The magnitude of the alignment tensor and the rhombicity were taken from ref 18. The values listed represent the average and standard deviation for the 32 simulated annealing structures.

If the calculations are repeated using only the PRE data from sites a and b, no significant deviations from the original RDC-based SRY coordinates are observed for either the single- or three-conformer representations of the EDTA−Mn^{2+} groups (Figure 9b). This is an important result because it demonstrates that even if the EDTA−Mn^{2+} occupies a single position, as in the case of sites a and b, use of a multiple-structure representation for the paramagnetic group is not detrimental.

When the calculations are carried out using only the PRE data from site c, however, the pattern of deviations from the original RDC-based SRY coordinates (Figure 9c) is almost identical to that observed when all PRE data are employed (Figure 9a). Moreover, the same pattern of structural deviations is preserved when the RDC data are omitted from the calculations and the coordinate precision is lower (Figure 9d). Thus, the structural deviations arise from the single-conformer representation of the EDTA−Mn^{2+} group at site c.

There are three regions of SRY where there are substantial structural differences in the backbone coordinates calculated with the single- and three-conformer representations: residues 9−14, 37−46, and 66−74 (Figure 9e). To assess whether these deviations arise from structural distortions and decreased coordinate accuracy as a consequence of employing a single-conformer representation, we carried out cross-validation calculations against the RDC data using the single- and three-conformer representations to refine against the PRE data for site c, omitting the RDC data for the three regions specified above (Table 3). The free dipolar coupling R-factors for the NH and CαH RDCs (i.e., those for the three regions not included in the calculations) were substantially lower for the structures calculated with the three-conformer representation than for those calculated with the single-conformer representation: 21.8 ± 3.4% versus 34.1 ± 2.9% for the CαH RDCs, and 20.3 ± 1.5%

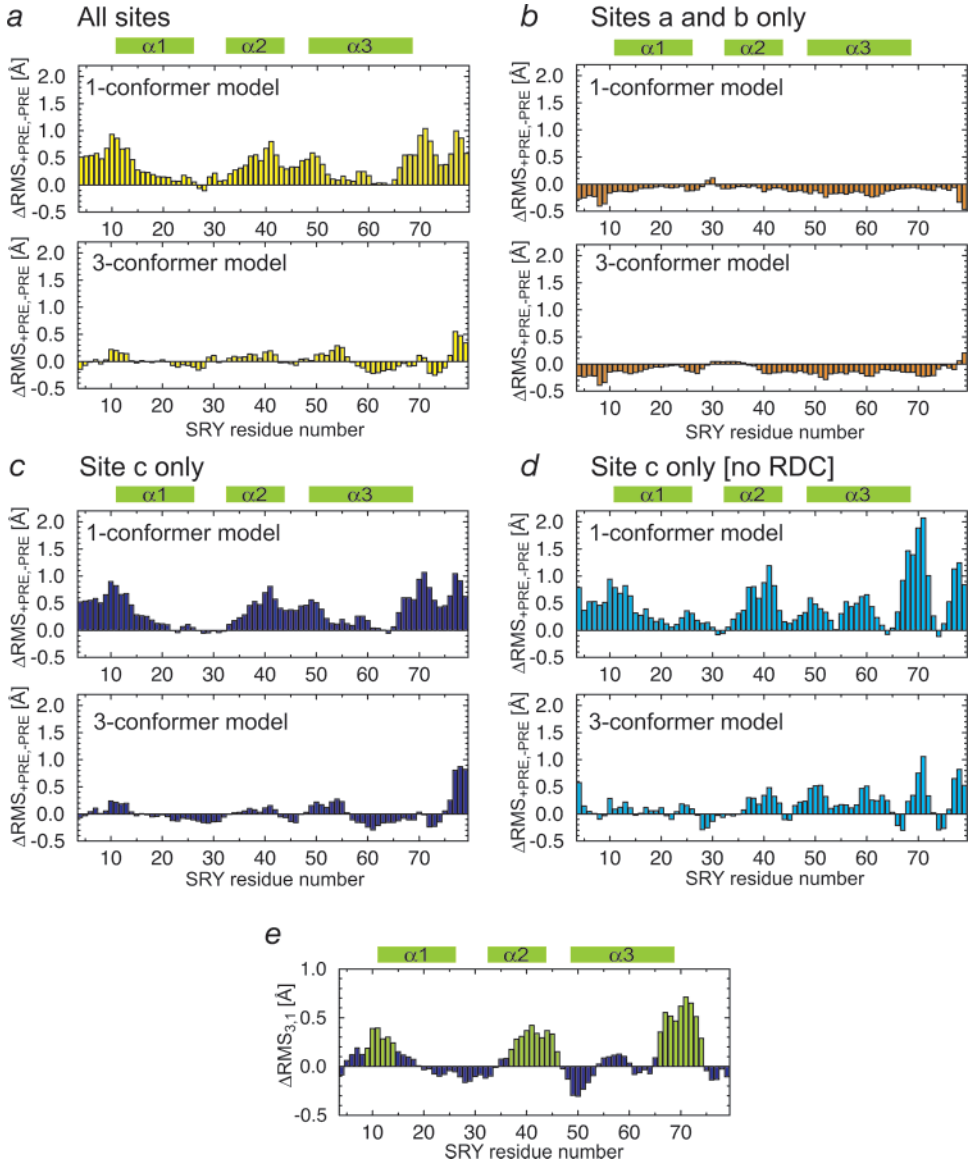

Figure 9. Structural consequences of PRE refinement with single- or three-conformer representations for the dT-EDTA−Mn^{2+} groups. Panels a−d provide a quantitative description of the differences between the simulated annealing structures calculated with and without PRE data in the form of plots of $\Delta RMS_{+PRE,-PRE}$ versus SRY residue number, where $\Delta RMS_{+PRE,-PRE}$ is defined as the rms difference between the mean backbone coordinates minus the sum of the precision of the coordinates for the two ensembles.[26] The comparison set of coordinates constitutes those calculated with the complete set of original experimental NMR restraints (i.e., without PRE data).[18] Positive values of $\Delta RMS_{+PRE,-PRE}$ indicate regions of real difference between the two sets of coordinates; negative values indicate the absence of any differences.[26] The results for the single- and three-conformer models for dT-EDTA−Mn^{2+} are displayed separately, at the top and bottom, respectively, of panels a−d. (a) $\Delta RMS_{+PRE,-PRE}$ plot for structures calculated using the PRE data from all three sites; (b) $\Delta RMS_{+PRE,-PRE}$ plot for structures calculated using only the PRE data from sites a and b. (c) $\Delta RMS_{+PRE,-PRE}$ plot for the structures calculated using only PRE data from site c. (d) $\Delta RMS_{+PRE,-PRE}$ plot for the structures calculated using only PRE data from site c but with the residual dipolar coupling (RDC) restraints omitted. The PRE refined structures in (a), (b), and (c) were calculated with all of the original NMR experimental restraints, including the RDCs;[18] the PRE-refined structures in (d) were calculated omitting the RDC restraints. (e) Regions of significant structural difference between the one- and three-conformer representation ensembles for structures refined with PRE data from site c, expressed as $\Delta RMS_{1,3}$. The backbone coordinates of three regions of the protein differ significantly between the two sets of structures (defined as $\Delta RMS_{1,3} > 0.2$ Å for a contiguous stretch of residues): residues 9−14, 37−46, and 66−74 (indicated by the green bars). Each ensemble consists of 30 simulated annealing structures.

versus 25.1 ± 1.8% for the NH RDCs. Further, the free dipolar coupling R-factors for the three-conformer representation are comparable to those (same set of RDCs) for the structures calculated without PRE data. One can therefore conclude that the use of a single-conformer representation for the PRE data at site c results in a significant decrease in protein backbone coordinate accuracy. As predicted at the beginning of this section, full refinement against the PRE data from site c using the single- and three-conformer representation yields similar PRE Q-factors of 0.19 and 0.17, respectively. Thus, although a single-conformer representation can result in good agreement between observed and calculated PRE data, this is achieved at the expense of coordinate accuracy.

Impact of ^{1}H-PRE Restraints on Coordinate Accuracy of the SRY/DNA Complex under Conditions where Only a Single Intermolecular NOE Restraint Is Employed. The original set of NMR restraints includes a large number of intermolecular NOEs and residual dipolar couplings which define the structure of the complex with high precision,[18] so that the full impact of the addition of ^{1}H-PRE restraints is not immediately apparent. We therefore carried out another set of calculations with and without ^{1}H-PRE restraints in which the

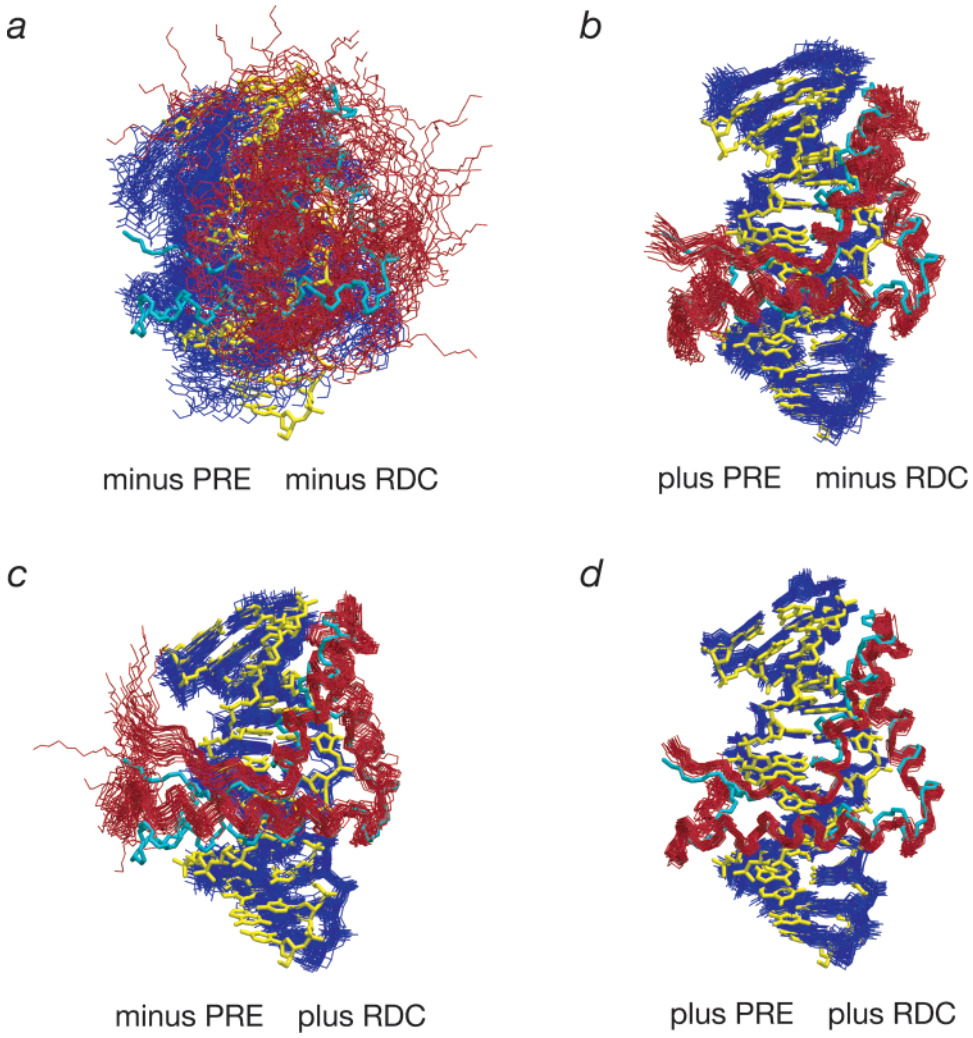

Figure 10. Impact of ^{1}H-PRE restraints on coordinate accuracy of the SRY/DNA complex under conditions where only a single intermolecular NOE restraint located at the center of the protein−DNA interface is employed. Best-fit superpositions of 30 simulated annealing structures calculated (a) without ^{1}H-PRE and RDC restraints, (b) with ^{1}H-PRE and without RDC restraints, (c) without ^{1}H-PRE and with RDC restraints, and (d) with ^{1}H-PRE and RDC restraints. PRE restraints were applied with the three-conformer representation for the dT-EDTA−Mn^{2+} groups. The protein backbone (residue 4−79) and DNA (central 14 base pairs) of the simulated annealing structures are drawn in red and blue, respectively. The original restrained regularized mean structure of the SRY/DNA complex[18] determined using the complete set of 168 intermolecular NOE restraints is shown in cyan (protein) and yellow (DNA). The single intermolecular NOE restraint used in the calculations is between the CδH$_3$ methyl group of Ile13 and the H2 proton of A9 with a distance range of 1.8−3.2 Å. Apart from the 167 omitted intermolecular NOE restraints, all other NMR-derived restraints from the original structure determination of the SRY/DNA complex[18] were employed. Coordinate accuracy as assessed by the average of the rms difference between the simulated annealing structures and the mean coordinates derived from calculations employing the complete set of experimental NMR restraints is as follows: (a) 14.6 ± 0.9 Å, (b) 2.9 ± 0.3 Å, (c) 3.8 ± 0.6 Å, and (d) 1.9 ± 0.2 Å for the complete protein backbone (residues 4−79) and DNA heavy atoms (central 12 base pairs); (a) 11.7 ± 1.0 Å, (b) 2.6 ± 0.3 Å, (c) 2.7 ± 0.4 Å, and (d) 1.7 ± 0.1 Å for the backbone of the SRY core (residues 11−68) and DNA heavy atoms (central 12 base pairs).

number of intermolecular NOE restraints was reduced from 168 to only a single NOE. For this purpose, we chose the intermolecular NOE between the CδH$_3$ methyl group of Ile13 and the H2 proton of A9, which is located approximately at the center of the binding site and can be unambiguously assigned in the 3D ^{13}C-separated/^{12}C-filtered NOE spectrum. Using this single intermolecular NOE restraint, structures of the SRY/DNA complex were calculated (a) without both PRE and RDC restraints, (b) without PRE but with RDC restraints, (c) with PRE but without RDC restraints, and finally (d) with both PRE and RDC restraints. In all calculations, a three-conformer representation was used for all three EDTA−Mn^{2+} groups. The resulting structures are shown in Figure 10.

In the absence of both PRE and RDC restraints, the relative orientation between the protein and DNA cannot be determined (Figure 10a). In the presence of PRE restraints, the relative orientation is determined at a reasonable level of accuracy: the rms difference from the original NMR coordinates (calculated with all NMR restraints other than the PRE data) is 2.6 ± 0.3 Å for the protein backbone core (resides 11−68) plus the heavy atoms of the central 12 bp of DNA (Figure 10b). Application of RDC restraints also permits the orientation to be determined at a similar level of accuracy (2.7 ± 0.4 Å for the protein backbone core and central 12 bp of DNA; Figure 10c). Although the RDC data, measured in a single alignment medium and characterized by an asymmetric alignment tensor, are consistent with four possible protein−DNA orientations, the relative orientation in this instance is uniquely determined because the other three would exhibit intermolecular atomic overlap and are therefore excluded by the repulsive van der Waals term. A

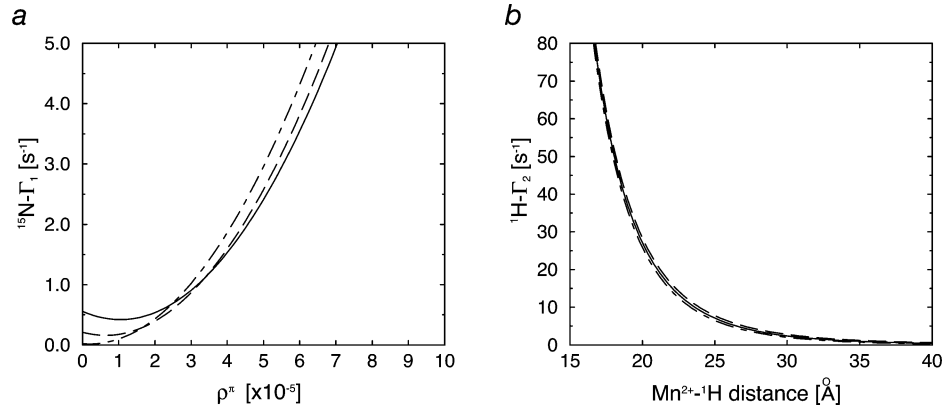

Figure 11. Effect of electron spin density delocalized to the $2p_z$ orbital. (a) Relationship between ^{15}N-Γ_1 induced by Mn^{2+} and the delocalization density ρ^π. The overall correlation time τ_c and Larmor frequency for ^{15}N were set to 2.9 ns and 50.6 MHz, respectively. Curves for three different $Mn^{2+}-^{15}N$ distances are shown: 17 Å, solid line; 25 Å, dashed line; and 30 Å, dotted line. Note that the distance-independent term of eq 23 (second term of Δ^2) is dominant in situations where ρ^π is large. (b) Dependence of ^1H_N-Γ_2 on electron$-^1H$ distance in the absence (solid line) and presence of the delocalization effect ($\rho^\pi = 1 \times 10^{-5}$) with the maximum and minimum limits depicted by dashed and dotted lines, respectively.

comparison of Figure 10b and c reveals some interesting differences as a consequence of the different information content afforded by the distance-based PRE data and the orientation-based RDC restraints. Thus, both the N (residues 4−10)- and the C (residues 69−79)-terminal tails of SRY that makes extensive contacts with the DNA are well defined by the PRE restraints but not by the RDC restraints. This is due to the fact that translational displacements are restricted by the PRE restraints but do not impact the RDC restraints, providing the orientations of the corresponding bond vectors can still be satisfied, which is readily the case for extended segments of structure. The DNA, however, is determined with higher accuracy in the structures calculated with RDC restraints, which is readily rationalized because the current PRE data are restricted to ^1H nuclei located in the protein portion of the complex. Thus, the translational PRE and orientational RDC restraints are complementary, and the combined use of both PRE and RDC restraints increases the accuracy still further (1.7 ± 0.1 Å for the protein core plus central 12 bp of DNA; Figure 10d). Similar improvements in accuracy were also observed when a single NOE was chosen from either edge of the protein−DNA interface.

An additional set of calculations, with both PRE and RDC restraints, was also carried out using a single-conformer representation for the EDTA−Mn^{2+} groups. Although the accuracy with which the relative orientation of the protein and DNA is determined is only slightly worse than that obtained using the three-conformer representation (2.1 ± 0.1 Å versus 1.7 ± 0.1 Å for the protein core plus central 12 bp of DNA), the coordinate accuracy of the individual protein and DNA components is significantly reduced. Thus, the distortions in the protein structure remain localized to the same three regions of the proteins discussed in the preceding section (residues 9−14, 37−46, and 66−74), and the backbone accuracy for these three regions is reduced from 0.4 to 0.9 Å. In addition, the accuracy of the DNA coordinates (central 12 bp) is also reduced from 1.0 to 1.3 Å. These structural distortions necessarily impact the overall accuracy of the complex because they involve regions of the protein that contact the DNA.

Effect of Unpaired Electron Spin Density Delocalized to the $2p_z$ Orbital. In this section, we consider potential errors resulting from unpaired electron spin density delocalized to the $2p_z$ orbital of carbon or nitrogen atoms, which could result in a breakdown of the point−dipole approximation of the SB equation. Recently, this effect has been shown to be significant in the case of the small copper protein plastocyanin that functions as an electron carrier in the electron-transfer process in photosynthesis.[9] However, as discussed in ref 9, the large effect may be related to the electron-transfer processes and may not be significant for redox-inactive systems. We have therefore estimated the magnitude of the effect for EDTA−Mn^{2+} systems.

If unpaired electrons are delocalized onto the $2p_z$ orbital of carbon or nitrogen atoms, the SB equation becomes:[7,8]

$$\Gamma_m = \Delta^2 f_{SB,m}(\tau_c) = \{r^{-6} + a(\rho^\pi)^2 r_L^{-6} + b\rho^\pi r_L^{-3} r^{-3}\} f_{SB,m}(\tau_c) \quad (23)$$

The original SB equation (eq 11) corresponds to the case where the second and third terms of Δ^2 are negligible. Parameters for the additional terms are as follows: ρ^π is the unpaired π-spin density located in the $2p_z$ atomic orbital; $a = 0.16$ and $b = -0.4$ for ^{13}C and ^{15}N nuclei; $a = 0.116$ and $b = 0.59 \cos 2\phi$ for protons, where ϕ is the angle between electron$-^1H$ and ^1H-C (or N) vectors; $r_L = 0.420$ Å for carbon and 0.345 Å for nitrogen; r_L is the ^1H-C (or N) distance.[8,9] The second and third terms of Δ^2 could become dominant for distances larger than 10 Å, if ρ^π is greater than 10^{-4}. The effect is especially strong for ^{13}C and ^{15}N, and, therefore, ρ^π can be estimated from PRE data measured for these nuclei.

To this end, we measured a ^{15}N-Γ_1 data set at 500 MHz on the SRY/DNA complex with dT-EDTA−Mn^{2+} located at site c. ^{15}N-Γ_1 rather than ^{15}N-Γ_2 was chosen, because, according to eqs 2 and 3, the range of ^{15}N-Γ_1 is expected to be about 10% that of ^1H-Γ_1 for the same electron−nuclear distance whereas the range of ^{15}N-Γ_2 is only 1% that of ^1H-Γ_2. (Note that the $J(\omega_I)$ values for ^{15}N and ^1H are very different for the correlation time of our system.) The range of observed ^{15}N-Γ_1 values was 0.0−0.4 s^{-1}. The $Mn^{2+}-^{15}N$ distances in the structure ensemble of the complex range from 17 to 38 Å. Figure 11a shows the theoretical relationships between ρ^π and ^{15}N-Γ_1 for a τ_c value of 3 ns. The observed range of ^{15}N-Γ_1 is accounted for by a value of $\rho^\pi \leq 1 \times 10^{-5}$. The theoretical dependence of ^1H-Γ_2 on the $Mn^{2+}-^1H$ distance with and without the second and third terms of eq 23 is plotted in Figure 11b for $\rho^\pi = 1 \times 10^{-5}$.

Clearly, as far as ^{1}H-Γ_2 is concerned, the effect of taking ρ^τ into account is negligible, particularly when considering the experimental errors in the measurement of ^{1}H-Γ_2. Thus, we conclude that the point−dipole approximation using $\Delta^2 = r^{-6}$ for back-calculation of ^{1}H-PREs is entirely valid for the dT-EDTA−Mn^{2+} system.

Discussion

Importance of Taking Paramagnetic Group Flexibility into Account in the Analysis of PRE Data. With the coordinates of the original RDC-based NMR structure of the SRY/DNA complex held fixed, we have demonstrated that a marked improvement in the agreement between observed and calculated ^{1}H-PREs arising from EDTA−Mn^{2+} at site c is obtained using a multiple-conformer representation of the paramagnetic group. It seems highly likely that this observation reflects an actual multiplicity of Mn^{2+} positions. Indeed, Dreyer and Dervan proposed a two-conformer model for dT-EDTA−Fe^{2+} located in the middle of a stretch of DNA on the basis of the observed DNA cleavage patterns generated by a hydroxyl radical produced by Fe^{2+} in the presence of dithiothreitol.[27] Such conformational multiplicity is probably related to the number of anchoring positions for the seventh coordination ligand to the metal ion. In the crystal structure of EDTA−Mn^{2+}, six atoms of EDTA are involved in coordination to Mn^{2+} and a water molecule was identified as the seventh coordination ligand.[22] A DNA-bound water molecule could readily serve as the seventh coordination ligand because numerous bound water molecules are present in the major groove of the DNA where EDTA−Mn^{2+} is located.[28] A dT-EDTA−Mn^{2+} located in the middle of a DNA oligonucleotide could readily find two anchoring positions because of the pseudosymmetric nature of the major groove, whereas dT-EDTA−Mn^{2+} located at the outer edges of a DNA oligonucleotide may find only one major anchoring position because of asymmetry. Because the spatial distribution of the paramagnetic group of dT-EDTA−Mn^{2+} at site c is as wide as ∼11 Å, consideration of the ensemble average is necessary for accurate back-calculation of ^{1}H-PREs arising from dT-EDTA−Mn^{2+} located in the interior of a DNA sequence.

It is probably appropriate to use the ensemble approach even in cases where the linker connecting the paramagnetic group is relatively short in length. Hubbell and co-workers recently determined high-resolution crystal structures of spin-labeled T4 lysozyme mutants in which a paramagnetic nitroxide ring was attached to the Sγ atom of a surface cysteine residue via a three-bond linker; the electron density for the nitroxide rings was weak, indicative of high mobility of the paramagnetic group; moreover, two distinct conformers with the nitroxide oxygen atoms placed 8.3 Å apart were observed for the 119R1 mutant.[29] These observations also support the necessity of the ensemble approach in the quantitative analysis of ^{1}H-PRE data on a macromolecule with an extrinsic, covalently attached, paramagnetic group.

Bearing this in mind, it is our view that a multiple-conformer representation of the paramagnetic group represents the safest and most appropriate approach for the refinement of PRE data arising from any potentially flexible paramagnetic group attached to a macromolecule. As was seen in the case of the paramagnetic group attached to site c of the SRY/DNA complex, structural distortion and a reduction of coordinate accuracy are clearly apparent when the paramagnetic group is inappropriately represented by a single-conformer model and the conventional SB equation is applied (Figure 9c,d). Conversely, in cases where the paramagnetic center occupies a unique position (cf. sites a and b in the present case), a multiple-conformer representation does not adversely impact the back-calculation of the ^{1}H-PRE or coordinate accuracy (Figure 9b). Hence, operationally, the ensemble approach should always be used as the "default" in structure refinement against PRE data arising from a potentially flexible paramagnetic group. It is worth noting that the large increase in accuracy of the PRE back-calculation afforded by the ensemble approach requires only a small amount of additional computational time. In the examples presented here, the total calculation time was only increased by ∼15% for the three-conformer representation relative to the single-conformer one. This is because the ensemble average for the ^{1}H-PRE is calculated with the SBMF equation (eq 12) using a multiple-structure representation restricted to the paramagnetic center while employing only a single structure for the rest of the macromolecule. This is perfectly reasonable for relatively rigid portions of the macromolecule such as the backbone, the hydrophobic core, and macromolecular interfaces, because any positional fluctuations of ^{1}H nuclei in such regions will not contribute significantly to ensemble effects of $\langle r^{-6}\rangle$ and S^2 for the ^{1}H-PRE. For highly mobile, exposed surface side chains, however, it may be necessary to use an ensemble representation for the side chains as well.

Number of Conformers Required for Accurate ^{1}H-PRE Back-Calculations. The effective distance $\langle r^{-6}\rangle^{-1/6}$ for a PRE interaction is highly biased toward the minimum distance, which restricts the effective positions of the paramagnetic center to regions of space closest to the observed protons. Consequently, the optimal number of conformers used to represent a paramagnetic group in ^{1}H-PRE back-calculations will depend on both the size and the shape of the conformational space sampled by the paramagnetic center and the spatial distribution of the observed protons. For example, even for a relatively large spatial ensemble of the paramagnetic center, a single-structure representation of the paramagnetic group may still yield reasonable accuracy, providing, as illustrated in Figure 12a, only a particular region of the ensemble space is effective in terms of the PRE observed on most protons. This situation is typified, for example, by back-calculation of the ^{1}H-PRE arising from the dT-EDTA−Mn^{2+} groups located at the edges of a DNA oligonucleotide (i.e., sites a and b). If, on the other hand, a wide portion of the ensemble space of the paramagnetic center is within a similar distance from the distribution of observed protons, as is the case for site c and illustrated schematically in Figure 12b, a multiple-structure representation for the paramagnetic group is required for accurate back-calculation of the ^{1}H-PRE. The total number of conformers required to represent the paramagnetic group may vary from case to case, but the present calculations suggest that in the case of intermolecular ^{1}H-PREs it is likely that a three-conformer model will be generally sufficient because the intermolecular PRE interaction vectors are biased to one particular direction.

(27) Dreyer, G. B.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 968−972.
(28) Schneider, B.; Cohen, D.; Berman, H. M. *Biopolymer* **1992**, *32*, 725−750.
(29) Langen, R.; Oh, K. J.; Cascio, D.; Hubbell, W. L. *Biochemistry* **2000**, *39*, 8396−8405.

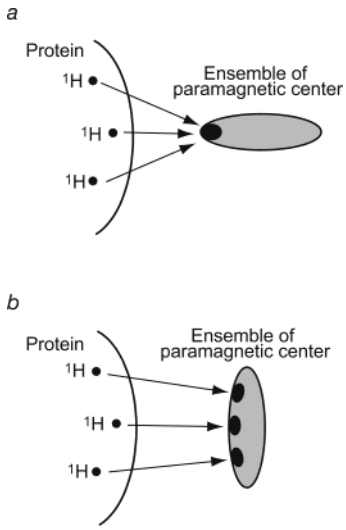

Figure 12. The number of conformers required to represent the flexible paramagnetic center depends on the shape and relative orientation of the ensemble space occupied by the paramagnetic center. Because effective $\langle r^{-6}\rangle^{-1/6}$ distances are close to the minimum electron−proton distance, a single-point representation of the paramagnetic center will be adequate for geometry a but not geometry b. The gray region represents the conformational space in which the paramagnetic center can be located within the stereochemical restrictions imposed by the covalent linker.

Although one might have anticipated that over-fitting would present a problem with the multiple-structure representation, in practice we find the approach insensitive to this issue. In the case of NOE analysis using a multiple-structure representation of the protein,[16b] over-fitting can readily occur by adjusting one conformation to exclusively fit several bad NOE restraints. As a consequence, this phenomenon effectively restricts the number of conformers that can be employed for refinement against NOE-derived interproton distance data to two or three at the most. In ^{1}H-PRE back-calculations presented here, on the other hand, only the paramagnetic groups are represented by multiple structures. It is therefore virtually impossible to generate a single conformer that just fits a small number of bad data points, because the position of the paramagnetic center affects back-calculation of every ^{1}H-PRE data point. Indeed, the results of the complete cross-validation calculations (Figure 3) indicate that even a 10-conformer representation for the paramagnetic group can be used without over-fitting.

Potential Role of ^{1}H-PRE in NMR Structure Determination of Macromolecular Complexes. The ensemble representation for extrinsic paramagnetic groups presented in this paper permits accurate quantitative use of ^{1}H-PRE data in macromolecular structure calculations, even if the artificially attached paramagnetic group is flexible. Probably the most useful application of the ^{1}H-PRE will involve macromolecular complexes, both protein−nucleic acid and protein−protein. Currently, most NMR structure determinations of macromolecular complexes rely on intermolecular NOEs,[30] supplemented in some cases by orientational information provided by residual dipolar coupling data.[18,31] Thus, the only experimental distance information between the two macromolecules is provided by NOE-derived interproton distance restraints. In many instances, however, intermolecular NOE cross-peaks may be broadened due to global or local exchange processes at the molecular interface.[32] Moreover, the assignment of NOE cross-peaks to unique interactions is nontrivial in the presence of extensive chemical shift overlap. In such instances, the ^{1}H-PRE can provide highly informative, long-range intermolecular distance information to protons distant from the actual interface, thereby considerably enhancing the accuracy of the structure determination, as was demonstrated above (cf. Figure 10). Our results on the SRY/DNA complex suggest that the combination of ^{1}H-PRE and residual dipolar coupling restraints significantly reduce reliance on intermolecular NOE data such that reasonable accuracy can be attained even if only a few intermolecular NOE restraints are available and the DNA is highly distorted. Moreover, analysis of intermolecular ^{1}H-PRE data is straightforward. In our analysis of the three SRY/DNA complexes with EDTA conjugated to three different sites on the DNA, no significant changes in ^{1}H-, ^{15}N-, or ^{13}C-chemical shifts for the bound protein were observed relative to the original SRY/DNA complex without the dT-EDTA modification. Consequently, reassignment of NMR signals was not required for the analysis of the intermolecular ^{1}H-PRE data. Thus, a large number of intermolecular translational restraints can be obtained with minimal effort by analyzing ^{1}H-PRE data using sensitive 2D HSQC-type experiments on a complex in which an extrinsic paramagnetic center is attached to one component and the other component is uniformly labeled with ^{15}N and/or ^{13}C. In addition, it should be readily feasible to measure ^{1}H-Γ_2 data on backbone amide protons with good precision on relatively large macromolecular complexes using 3D TROSY-HNCO-type experiments.[33] We therefore expect that ^{1}H-PRE by artificially attached paramagnetic groups will play a significant role in structure determination of macromolecular complexes by NMR, providing highly valuable long-range intermolecular restraints.

Concluding Remarks

In this paper, we have provided the theoretical framework and computational strategy required to accurately back-calculate ^{1}H-PRE data arising from a flexible paramagnetic group attached to a macromolecule. We show that it is essential to take into account the conformational space sampled by the flexible paramagnetic group, and to this end we make use of a multiple-structure representation of the paramagnetic group. In combination with recently developed chemical modification techniques, quantitative use of ^{1}H-PRE data can provide unique and highly valuable long-range distance information for NMR structure determination of macromolecular systems and is likely to be particularly valuable when dealing with macromolecular complexes.

Acknowledgment. We thank Attila Szabo for useful discussions, and Wayne Hubbell for providing the coordinates of the crystal structures of spin-labeled T4 lysozyme mutants. This

(30) Clore, G. M.; Gronenborn, A. M. *Trends Biotechnol.* **1998**, *16*, 22−34.
(31) (a) Garrett, D. S.; Seok, Y.-J.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1999**, *6*, 166−173. (b) Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9021−9025. (c) Wang, G.; Louis, J. M.; Sondej, M.; Seok, Y.-J.; Peterkofsky, A.; Clore, G. M. *EMBO J.* **2000**, *19*, 5635−5649. (d) Cornilescu, G.; Lee, B. R.; Cornilescu, C.; Wang, G.; Peterkosky, A.; Clore, G. M. *J. Biol. Chem.* **2002**, *277*, 42298−42298. (e) Williams, D. C.; Cai, M.; Clore, G. M. *J. Biol. Chem.* **2004**, *279*, 1449−1457.
(32) (a) Strzelecka, T. E.; Clore, G. M.; Gronenborn, A. M. *Structure* **1995**, *3*, 1087−95. (b) Foster, M. P.; Wuttke, D. S.; Radhakrishnan, I.; Case, D. A.; Gottesfeld, J. M.; Wright, P. E. *Nat. Struct. Biol.* **1997**, *4*, 605−608. (c) Iwahara, J.; Wojciak, J. M.; Clubb, R. T. *J. Biomol. NMR* **2001**, *19*, 231−241.
(33) Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13585−13590.
(34) Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 12334−12341.

work was supported in part by the AIDS Targeted Antiviral program of the Office of the Director of the NIH (to G.M.C.).

Supporting Information Available: Pulse sequences for measurement of ^{1}H-Γ_2 on ^{13}C-attached protons, and force-field parameters for the dT-EDTA−Mn^{2+} groups used in the simulated annealing calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

JA031580D