Probability Assessment of Conformational Ensembles: Sugar Repuckering in a DNA Duplex in Solution

Nikolai B. Ulyanov, Uli Schmitz, Anil Kumar, and Thomas L. James
Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446 USA

ABSTRACT Conformational flexibility of molecules in solution implies that different conformers contribute to the NMR signal. This may lead to internal inconsistencies in the 2D NOE-derived interproton distance restraints and to conflict with scalar coupling-based torsion angle restraints. Such inconsistencies have been revealed and analyzed for the DNA octamer GTATAATG·CATATTAC, containing the Pribnow box consensus sequence. A number of subsets of distance restraints were constructed and used in the restrained Monte Carlo refinement of different double-helical conformers. The probabilities of conformers were then calculated by a quadratic programming algorithm, minimizing a relaxation rate-based residual index. The calculated distribution of conformers agrees with the experimental NOE data as an ensemble better than any single structure. A comparison with the results of this procedure, which we term PARSE (Probability Assessment via Relaxation rates of a Structural Ensemble), to an alternative method to generate solution ensembles showed, however, that the detailed multiconformational description of solution DNA structure remains ambiguous at this stage. Nevertheless, some ensemble properties can be deduced with confidence, the most prominent being a distribution of sugar puckers with minor populations in the N-region and major populations in the S-region. Importantly, such a distribution is in accord with the analysis of independent experimental data—deoxyribose proton-proton scalar coupling constants.

INTRODUCTION

Because of its continuous improvement during the last decade, high-resolution nuclear magnetic resonance (NMR) methodologies have become a traditional tool for structural studies on biomolecules in solution. The introduction of complete relaxation matrix methods has increased the accuracy of interproton distances estimated from nuclear Overhauser effect (NOE) data (Keepers and James, 1984; Boelens et al., 1988; Borgias and James, 1989, 1990; Post et al., 1990; Madrid et al., 1991). The accuracy of distance restraints is of special importance for the high resolution structure determination of relatively short (1–1.5 helical turns) DNA oligonucleotides; in this case, the phenomenon under study is a fairly subtle sequence-conformation relationship in the DNA double helix, rather than overall folding geometry, which is the prime target of investigations of longer biopolymers. A persistent interest in this subtle relationship is explained by a general consensus that the dependence on nucleotide sequence of DNA conformation is involved in many biological processes, such as packaging of DNA in chromatin, transcription, replication, recombination, etc. (for reviews, see Trifonov, 1985; Hagerman, 1990; Steitz, 1990).

Several structures of short DNA duplexes have been recently solved to high resolution in this laboratory (Schmitz et al., 1991, 1992b; Stolarski et al., 1992; Mujeeb et al., 1993; Weisz et al., 1994). These structures, although belonging to the family of B-forms, display a sequence-dependent conformational microheterogeneity in solution. Some of the conformational features, revealed in these structures, reproduce independent data for DNA: both sequence-specific, e.g., a well known tendency of TG·CA dinucleotides to be bent into the major groove (Mujeeb et al., 1993; Weisz et al., 1994), as well as average characteristics for pseudo-random nucleotide sequences, such as an average helical repeat of 10.4 base pairs per turn of double helix (Ulyanov and James, 1994). Significantly, the latter is a fairly subtle structural feature of B-DNA in solution, which is generally not reproduced in high-resolution crystal structures of B-DNA (Baikalov et al., 1993). In addition, the structure of one of the duplexes, that of a DNA octamer d(GTATAATG)·d(CATTATAC), was determined using two different methods of refinement, namely, restrained molecular dynamics (rMD) (Schmitz et al., 1992b) and restrained Monte Carlo (rMC) (Ulyanov et al., 1993); both methods yielded essentially the same structures. This indicates that the resulting conformation is largely determined by experimental structural restraints rather than by a particular method of refinement employed.

All of the points mentioned above yield significant confidence in the structures of short DNA duplexes in solution, determined using the methodologies employed in this laboratory (James, 1991; James et al., 1991). However, it is well known that the DNA double helix is not rigid in *solution* (for a review, see Hagerman, 1988). Therefore, experimental NMR structural parameters (NOE intensities and scalar coupling constants) represent appropriate averages according to the population-weighted ensemble of conformers. Complicating the situation, averaging of NOE intensities and scalar coupling constants has a different nature; the derived distance and torsion angle restraints are not simple population-weighted, arithmetic averages. It has been understood for a

Received for publication 2 June 1994 and in final form 18 August 1994.
Address reprint requests to Dr. Thomas L. James, Department of Pharmaceutical Chemistry, University of California, 926 Medical Science, San Francisco, CA 94143-0446. Tel.: 415-476-1569; Fax: 415-476-0688; E-mail: james@picasso.ucsf.edu.
© 1995 by the Biophysical Society
0006-3495/95/01/13/12 $2.00

long time that a static "average" structure derived from the observed spectroscopic data may even be devoid of physical meaning (see, e.g., Jardetsky, 1980) if several distinct conformers contributed to the NMR signal. The presence of such conformational equilibria may lead to internal inconsistencies in NMR data. However, if the equilibria are interpreted in terms of a single conformation, such inconsistencies may be useful, because they can serve as a basis for assessment of the conformational ensemble. Such an approach has been previously applied to studies of conformational equilibria in peptides and small proteins (Fesik et al., 1986; Kessler et al., 1988; Kim and Prestegard, 1989; Blackledge et al., 1993). In the case of nucleic acids, it has been shown, initially for mononucleotides (Davies and Danyluk, 1975) and then for oligonucleotide duplexes (Rinkel et al., 1987; Celda et al., 1989; Schmitz et al., 1990, 1992b; Stolarski et al., 1992; Mujeeb et al., 1992; Weisz et al., 1992; Macaya et al., 1992; Bishop et al., 1994), that proton-proton scalar coupling constants are not consistent with any single conformation of sugar rings. However, they can be explained by interconversions between the northern- (N) and southern (S)-type sugar puckers.

This creates a methodological contradiction: scalar couplings indicate the presence of multiple deoxyribose conformers for short DNA duplexes in solution, whereas interproton distances extracted from the NOE data for the same molecules lead to single static structures.

This contradiction has been resolved in our recent study of a DNA octamer d(GTATAATG)·d(CATTATAC) utilizing MD simulations with exponentially weighted time-averaged distance restraints (MD-tar) (Schmitz et al., 1993). In contrast to standard rMD calculations, MD-tar does not require *simultaneous* satisfaction of all restraints, but it generates an ensemble of conformations which satisfies restraints on a *time-average* basis (Torda et al., 1990; Pearlman and Kollman, 1991). We have demonstrated in the work cited above that MD-tar generated a structural ensemble that agreed with experimental NOE data better than any single structure, giving lower NOE *R*-factors. Furthermore, despite the fact that torsion angle restraints were not used in simulations, this ensemble had a smaller root-mean-square (RMS) deviation of average calculated coupling constants from their experimental values, because the generated distribution of sugar puckers contained both N- and S-type conformers. Importantly, the most populated conformers (and the average structural parameters as well) were close enough to the static structure determined by the rMD refinement, justifying why this single conformation is a good approximation of the "solution ensemble" for the DNA octamer (Schmitz et al., 1993).

However, an important question arises regarding the dynamic description of DNA oligonucleotides, namely, is the generated structural ensemble unique? To answer this question, it is necessary to have independent methods producing such ensembles. This is analogous to the situation with the refinement of the average solution structure, where we needed to use two different methods—rMD and rMC—to demonstrate that the final structure is nearly independent of the method used. Therefore, in the present work, we developed an alternative approach for generation and assessment of conformational equilibria, which we call PARSE (Probability Assessment via Relaxation rates of a Structural Ensemble), for oligonucleotide duplexes in solution, and applied it to a DNA octamer which has been extensively studied in this laboratory (Schmitz et al., 1992a, b, 1993; Ulyanov et al., 1993):

```
5′ - G1  T2  A3  T4  A5  A6  T7  G8
     C16 A15 T14 A13 T12 T11 A10 C9 - 5′
```

At first, we analyzed satisfaction, or rather, violation of individual distance restraints in the single best conformation of the octamer, which was determined previously. Then, based on this analysis and on model calculations, we subdivided the restraint set into several overlapping subsets and repeated the refinement against each individual subset. The refined conformations, together with conformations obtained by energy minimization calculations, constituted a pool of several hundred conformers. Next, using a "quadratic programming algorithm" adapted by us for this special purpose, we found probabilities of conformers that minimize a relaxation rate-based figure of merit. In the last step, using a semi-manual search, we modified the distribution of probabilities to improve the NOE-based figures of merit. Similar to our MD-tar calculations (Schmitz et al., 1993), proton-proton scalar coupling constants were not included in the process of finding the optimal distribution, but they were used as an independent check of the agreement with the experimental data.

THEORY AND METHODS

Monte Carlo simulation and energy minimization calculations

All structural calculations for the DNA octamer, both Metropolis Monte Carlo and energy minimization, were performed with the DNAminiCarlo program (Ulyanov et al., 1989; Zhurkin et al., 1991), which uses internal coordinates—generalized helical parameters—as independent variables defining the double helical conformation. Conformational energy is calculated using empirical additive atom-atom potential functions described elsewhere (Zhurkin et al., 1981; Poltev and Shulyupina, 1986). All calculations were carried out on the Cray C90 at the Pittsburgh Supercomputing Center and on SUN SPARC workstations in our laboratory.

Interproton distance restraints

The full set of distance restraints was generated previously (Schmitz et al., 1993) via the MARDIGRAS program (Borgias and James, 1989, 1990) using 500 MHz proton-proton 2D NOE datasets acquired at mixing times of 100 and 150 ms (Schmitz et al., 1992b). The MARDIGRAS program computes lower and upper bounds rather than a single value for each restraint; these bounds are then used to construct a "flat well potential" as described previously (James, 1991; James et al., 1991). Restraints involving methyl groups refer to the geometric centers of three methyl protons; they were determined with an accounting for the fast rotation of methyl groups (Liu et al., 1992). In total, the restraint set consisted of 184 interproton distance restraints for 16 residues, which included 100 intra-residue, 78

sequential, and 6 inter-strand restraints with the average flat well width of 0.31 Å. In addition to experimentally determined distance restraints, we used pseudo-restraints with a flat well of 1.8–2.0 Å for all protons involved in Watson-Crick hydrogen bonds.

In the original structure determination, distances involving H3′ protons were excluded from the list of restraints, because of reduced accuracy due to spectral overlap and proximity of the residual HDO signal, but all measured NOE intensities, including those involving H3′ protons, were used for the calculation of 2D NOE-based figures of merit (Schmitz et al., 1992b). (The refined structure is available from the Brookhaven Protein Data Bank under accession number 1D70.) Later, relevant 2D NOE cross peaks were re-integrated with improved deconvolution software, and some additional estimates were made assuming a structural symmetry for the TATA·TATA stretch of the octamer (Schmitz et al., 1993).

Several *overlapping subsets* of restraints were constructed by excluding certain classes of distances from the full set. Namely, inter- and/or intra-residue distances H6/H8-H1′, H6/H8-H2′, H6/H8-H2″, H6/H8-H3′ and some others were excluded from the set, separately or in various combinations. Altogether, approximately 60 subsets were generated; some relevant subsets are described in more detail in the section of Results.

Generation of the pool of conformers

The DNA octamer was independently refined against each subset of interproton distance restraints with the Monte Carlo modules of the DNAmini-Carlo program; each refinement produced a single conformer. Protocols for the rMC refinement, including simulated annealing procedures, have been described previously (Ulyanov et al., 1993). In addition to rMC refinements, a number of restrained and unrestrained energy minimization calculations were carried out to produce a pool of structures broadly covering the conformational space. This was achieved by applying a "scanning procedure" described in detail elsewhere (Zhurkin et al., 1982; Ulyanov et al., 1992). In total, several hundred conformers were generated.

CORMA and MARDIGRAS calculations

Following Schmitz et al. (1992b, 1993), all CORMA (Borgias et al., 1993a) and MARDIGRAS (Borgias et al., 1993b) calculations were performed assuming a single rotational correlation time of 2.8 ns. The 18 site rapid jump model (Liu et al., 1992) was used for methyl protons. NOE intensities were normalized to the sum of all NOE intensities; no noise was added. The MARDIGRAS program was used to compute "experimental" relaxation rates on the basis of experimental 2D NOE datasets. The CORMA program was used (i) to calculate theoretical relaxation rates for each potential conformer, and (ii) to simulate NOE intensities for the ensembles of conformers (Schmitz et al., 1992a), assuming that the conformational equilibrium is fast on the NMR time-scale but that it is slower than the overall correlation time; simulated intensities were compared with the experimental data via various figures of merit, most frequently via crystallographic-type R-factor and sixth-root weighted R^x-factor (James, 1991):

$$R = \frac{\sum |I_o(i) - I_c(i)|}{\sum I_o(i)}, \qquad R^x = \frac{\sum |I_o(i)^{1/6} - I_c(i)^{1/6}|}{\sum I_o(i)^{1/6}}$$

where I_o and I_c refer to the observed and calculated NOE magnitudes, respectively. Experimental and theoretical relaxation rates, computed by the MARDIGRAS and CORMA programs, respectively, were used for the determination of optimal probabilities of conformers (*vide infra*).

Calculation of probabilities of conformers

Indices R and R^x are hardly applicable to the *global* search for optimal probabilities in the multiconformational equilibrium because of their complicated functional dependence on the probabilities of conformers. Therefore, we introduce here a relaxation rate-based index, a quadratic objective function Q^r:

$$Q^r(\mathbf{p}) = \sum_{k=1}^{n} w_k (T_k - E_k)^2 \tag{1}$$

where E_k stands for experimentally observed NMR parameters, which can be averaged in a linear fashion over an ensemble of conformers; T_k is a theoretical value of the kth parameter averaged over the ensemble:

$$T_k = \sum_{i=1}^{m} p_i t_{ki} \tag{2}$$

$\mathbf{p} = \{p_i\} \in \mathbf{R}^m$ is a column vector of probabilities; t_{ki} is a calculated value of the kth parameter for the ith conformer, $i = 1, \cdots, m$, m being the number of conformers. Parameters that can be used for the calculation of the objective function Q^r are dipolar relaxation rates and scalar coupling constants; weights w_k are introduced to regulate the relative contribution of parameters of a different kind as well as parameters measured with a different accuracy. However, in the present study, scalar couplings were not included in the Q^r-calculations (*vide infra*), and all weights were assumed to be equal. This objective function is analogous to the r^{-6}-based "mismatch parameter" used by Ernst and co-workers (Brüschweiler et al., 1991; Blackledge et al., 1993). The objective function Eqs. 1 and 2 can be re-written as

$$Q^r(\mathbf{p}) = \tfrac{1}{2}\mathbf{p}^T \cdot \mathbf{G} \cdot \mathbf{p} + \mathbf{g}^T \cdot \mathbf{p} + g_0 \tag{3}$$

where $\mathbf{G} = \{g_{ij}\} \in \mathbf{R}^m \times \mathbf{R}^m$ is a symmetric positively defined matrix and $\mathbf{g} = \{g_i\} \in \mathbf{R}^m$; the sign T stands for the transpose operation:

$$g_{ij} = g_{ji} = 2 \sum_{k=1}^{n} w_k t_{ki} t_{kj} \qquad i, j = 1, \cdots, m$$

$$g_i = -2 \sum_{k=1}^{n} w_k E_k t_{ki} \qquad i = 1, \cdots, m$$

$$g_0 = \sum_{k=1}^{n} w_k E_k^2$$

Now, the problem of the determination of optimal probabilities p can be formulated as

$$\text{Minimize} \quad Q^r(\mathbf{p}), \qquad \mathbf{p} \in \mathbf{R}^m \tag{4}$$

Subject to constraints

$$\sum p_i = 1 \tag{5}$$

$$p_i \geq 0, \qquad i = 1, \cdots, m \tag{6}$$

One apparent approach to solution of the problem Eqs. 4–6 would be to zero the derivatives of Q^r with respect to the probabilities p_i (Brüschweiler et al., 1991), which leads to a system of m linear equations with m variables. However, the solution of such equations generally violates the constraints Eqs. 5 and 6, which limits the applicability of this approach to the consideration of various combinations of a small number of conformers (Blackledge et al., 1993). On the other hand, the problem Eqs. 4–6 can be solved rigorously by the *quadratic programming algorithm* with the use of the *active set method* (Fletcher, 1981). We have written a FORTRAN program PDQPRO (Probability Determination by Quadratic PROgramming) implementing this algorithm for our special case. Below we outline this method as applied to the problem Eqs. 4–6; the notations of Fletcher (1981) are kept whenever possible.

First, we eliminate variable p_m using the condition Eq. 5; the new objective function $\tilde{Q}^r$ has the following form:

$$\tilde{Q}^r(\tilde{\mathbf{p}}) = \tfrac{1}{2}\tilde{\mathbf{p}}^T \cdot \tilde{\mathbf{G}} \cdot \tilde{\mathbf{p}} + \tilde{\mathbf{g}}^T \cdot \tilde{\mathbf{p}} + \tilde{g}_0, \tag{7}$$

where $\tilde{\mathbf{p}} \in \mathbf{R}^{m-1}$; matrix $\tilde{\mathbf{G}} = \{\tilde{g}_{ij}\} \in \mathbf{R}^{m-1} \times \mathbf{R}^{m-1}$, vector $\tilde{\mathbf{g}} = \{\tilde{g}_i\} \in \mathbf{R}^{m-1}$

and $\tilde{g}_0$ are related to $\mathbf{G}$, $\mathbf{g}$, and g_0 by

$$\begin{cases} \tilde{g}_{ij} = g_{ij} - g_{im} - g_{mj} + g_{mm}; & i,j = 1,\cdots,m-1 \\ \tilde{g}_i = g_i + g_{im} - g_{mm} - g_m; & i = 1,\cdots,m-1 \\ \tilde{g}_0 = g_0 + g_{mm}/2 + g_m \end{cases} \quad (8)$$

The problem Eqs. 4–6 is equivalent to the following:

$$\text{Minimize} \quad \tilde{Q}^r(\tilde{\mathbf{p}}), \qquad \tilde{\mathbf{p}} \in \mathbf{R}^{m-1} \quad (9)$$

Subject to constraints

$$c_i(\tilde{\mathbf{p}}) = p_i \geq 0, \qquad i = 1,\cdots,m-1 \quad (10)$$

$$c_m(\tilde{\mathbf{p}}) = 1 - \sum_{k=1}^{m-1} p_k \geq 0. \quad (11)$$

Vector $\tilde{\mathbf{p}} \in \mathbf{R}^{m-1}$ is called *feasible* if it satisfies the constraints Eqs. 10 and 11. The set of indices, A ($\tilde{\mathbf{p}}$), is called an *active set* if $c_i(\tilde{\mathbf{p}}) = 0\ \forall i \in$ A ($\tilde{\mathbf{p}}$). The index i, constraint c_i and variable p_i are called *active* if $i \in$ A ($\tilde{\mathbf{p}}$), and they are called *inactive* otherwise. We start from any feasible vector $\tilde{\mathbf{p}}^{(0)}$ and construct a series of feasible approximations $\tilde{\mathbf{p}}^{(k)}$ until the Kuhn-Tucker conditions are satisfied, which are sufficient conditions for the global minimizer in our case, because $\tilde{\mathbf{Q}}^r$ is a quadratic positively defined function and the feasible set defined by constraints Eqs. 10 and 11 is convex (Fletcher, 1981). In the following scheme of the algorithm, index k at $\tilde{\mathbf{p}}^{(k)}$ refers to the iteration number:

(i) Choose the initial feasible point $\tilde{\mathbf{p}}^{(0)}$ and determine the corresponding active set A. It can be done, e.g., by setting all $p_i = 1/m$; in this case the active set is empty.

(ii) Calculate the global minimizer $\mathbf{p}^*$ of the problem

$$\text{Minimize} \quad \tilde{Q}^r(\tilde{\mathbf{p}}), \qquad \tilde{\mathbf{p}} \in \mathbf{R}^{m-1}$$

$$\text{Subject to} \quad c_i(\tilde{\mathbf{p}}) = 0, \qquad i \in A$$

For that purpose, we reduce the dimensionality of the problem by taking into account all active (equality) constraints; let us denote the resulting quadratic function by $\hat{Q}$. When $m \notin$ A, this is done by simply eliminating all active variables: $p_i = 0$ for $i \in$ A. In the case of $m \in$ A, we have to take care of one additional constraint:

$$\sum_{j=1}^{m-1} p_j = 1, \quad (12)$$

which leads to elimination of one additional (inactive) variable, e.g., the last variable p_r in the inactive set ($r \notin$ A):

$$p_r = 1 - \sum_{j \notin A,\, j \neq r} p_j$$

The condition Eq. 12 is equivalent to $p_m = 0$ in the original problem Eq. 4; therefore, as can be easily shown, the resulting function $\hat{Q}$ can be obtained in this case from Eq. 3 by a formal substitution of index m for r in Eq. 8 followed by elimination of all active (zero) variables. After that, the problem is reduced to the unconstrained global minimization of a positively defined quadratic function $\hat{Q}$, which is done by zeroing partial derivatives $\partial\hat{Q}/\partial p_i$, for $i \notin$ A in the case of $m \notin$ A, and for $i \notin$ A, $i \neq r$ in the case of $m \in$ A. The resultant system of linear equations is solved by Gaussian elimination. The minimizer $\mathbf{p}^*$ may or may not be feasible. If it is feasible, it gives the next feasible approximation $\tilde{\mathbf{p}}^{(k+1)} = \mathbf{p}^*$, and we proceed to the next step. If $\mathbf{p}^*$ is not feasible, go to (iv).

(iii) Calculate Lagrange multipliers λ_i for all active constraints $i \in$ A. Here again, we have to consider separately the cases of the active and inactive mth constraint. If $m \notin$ A, obviously, $\lambda_i = \partial\tilde{Q}^r/\partial p_i$ for $i \in$ A. When $m \in$ A, it is easy to show that

$$\lambda_m = \frac{-1}{N_{in}} \sum_{j \notin A} \frac{\partial \tilde{Q}^r}{\partial p_j}, \qquad \lambda_i = \frac{\partial \tilde{Q}^r}{\partial p_i} + \lambda_m \quad \text{for} \quad i \in A, \quad i \neq m,$$

where N_{in} is the number of inactive constraints.

If all $\lambda_i \geq 0$, then $\tilde{\mathbf{p}}^{(k+1)}$ solves the problem Eqs. 9–11. If there are negative Lagrange multipliers, the constraint with minimum λ_i is removed from the active set A, and we return to step (ii).

(iv) Find the intersection between the line segment [$\tilde{\mathbf{p}}^{(k)}$, $\mathbf{p}^*$] and the boundary of the feasible set defined by the constraints of Eqs. 10 and 11. The intersection point gives the next feasible approximation $\tilde{\mathbf{p}}^{(k+1)}$, and one of the inactive constraints becomes active and is added to the active set A. Go to step (ii).

The outlined above algorithm was used to calculate the optimal probabilities $\mathbf{p}$ minimizing the relaxation rate-based index Q^r. After that stage, just a few conformers with nonzero probabilities were left, compared with several hundred conformers in the total pool. However, because of the exponential relationship between relaxation rates and NOE intensities, the optimal distribution of conformers with respect to the Q^r index is not necessarily optimal with respect to the NOE-based indices R and R^x. Therefore, the probabilities of conformers were additionally modified by a semi-manual search in order to improve NOE-based R- and R^x-factors.

Calculation of proton-proton scalar coupling constants

As we have mentioned before, it is possible to include scalar coupling constants in the Q^r-index. Experimental proton-proton coupling constants for the DNA octamer have been extracted from homonuclear 2QF-COSY spectra via simulation of the cross peaks (Schmitz et al., 1992b). However, in the present work, we chose to leave the coupling constants as an independent check of the agreement between the calculated conformational ensemble and the experimental data. Proton-proton coupling constants J(H1′,H2′), J(H1′,H2″), J(H3′,H2′), and J(H3′,H2″) were calculated for individual conformers using the modified Karplus equations (Haasnoot et al., 1980; de Leeuw et al., 1983), averaged over the ensemble, and compared with the experimental data.

RESULTS

Analysis of the distance restraint violations in the best single conformation

A number of very close B-DNA conformations for the DNA octamer have been previously calculated by restrained MD and MC protocols, with pair-wise atomic RMS deviations below 0.5 Å (Schmitz et al., 1992b; Ulyanov et al., 1993). For the following analysis, we will use the $A2_{fin}$ conformation determined by the rMC refinement (Ulyanov et al., 1993). As explained in Theory and Methods, this structure had been refined against a set of restraints not involving H3′ protons. The mean deviation between the actual interproton distances in the $A2_{fin}$ conformation and the closest of upper or lower bounds is 0.13 Å for this set of restraints. However, this index increased to 0.33 Å when calculated for the full set of restraints including fifty H6/H8/CH_3-H3′ intra- and inter-residue distances. The deviations are especially large for sixteen *intra-residue* H6/H8-H3′ restraints (1.05 Å on average); model distances are systematically overestimated compared with experimental restraints. These deviations are much too large and cannot be rationalized by decreased accuracy in the measurement of NOE cross peak intensities involving H3′ protons (*vide supra*). A similar situation, when intra-residue distances H6/H8-H3′ were overestimated while other intra-residue distances H6/H8-H1′/H2′/H2″ were satisfied well, has been previously reported for a DNA hexamer in solution; it has been hypothesized that this might

be indicative of a conformational equilibrium involving S-type (typical of B-DNA) and N-type (typical of A-DNA) sugar puckers (Ulyanov et al., 1992). It is well known (see, e.g., Wüthrich, 1986) that intra-residue H6/H8-H3′ distances are long in B-DNA and short in A-DNA, whereas distances H6/H8-H2′/H2″ are short in B-DNA and long in A-DNA. This statement is also valid for a local S-to-N repuckering of a single deoxyribose within the framework of B-DNA: the H8-H3′ distance decreases with decrease in sugar pseudorotation phase angle P, and the H8-H2′/H2″ distances simultaneously increase (Ulyanov et al., 1992). In the case of conformational equilibrium, averaged NOE intensities are strongly biased toward conformers with shorter distance; hence, the structure refined against H6/H8-H2′/H2″ restraints is bound to have overestimated intra-residue H6/H8-H3′ distances.

Further analysis showed that *inter-residue* experimental restraints H1′/H2′-H6/H8/CH_3 and H2″-CH_3 are generally consistent with B-type DNA conformations, whereas inter-residue H3′-H6/H8/CH_3 restraints are suggestive of A-DNA. Also, some other inter-residue restraints were found to be poorly satisfied by the refined structure of the DNA octamer; they were mainly localized around the A5 residue. Most notably, H2′T4-H8A5 and H2″T4-H8A5 NOE cross peaks had unusually low intensities which were not matched by the model structure. In addition, the distance restraints H8A5-H8A6 (2.69–2.99 Å) and H2A5-H2A6 (3.16–3.39 Å) are each relatively short; this pair of distances can hardly be matched simultaneously in any single conformation, which is suggestive of some kind of internal motion involving A5 and possibly A6.

Calculation of the "global" N-conformer

To test the hypothesis about N-type sugar puckers contributing to observed NOE intensities, we constructed a subset of restraints which included all H6/H8/CH_3-H3′ interproton distances, but excluded all H6/H8/CH_3-H2′/H2″ and inter-residue H6/H8/CH_3-H1′ distances. In a way, this set of restraints is complementary to the set used in the original refinement, although a number of restraints were present in both sets (e.g., inter-residue aromatic-to-aromatic distances, and intra-sugar distances). This new restraint set was used to repeat the refinement by the rMC simulated annealing protocol similar to that employed for the calculation of the $A2_{fin}$ conformation (Ulyanov et al., 1993). The superposition of the resultant structure, N1 conformation, with $A2_{fin}$ is shown in Fig. 1. The N1 conformation belongs to the A-family of forms; it is underwound with helical twist angle ranging from 24° to 30° for different dinucleotide steps, and it has a positive roll angle of 11° on average; glycosidic angles are also typical for A-DNA. At the same time, this is not a standard A form: deoxyriboses have mostly C4′-*exo* rather than C3′-*endo* pucker, with pseudorotation angle varying from 21° to 62°.

To assess the possibility of the N1 conformation contributing to the observed NOE signals, we carried out simulations of interproton NOE intensities for different equilibria

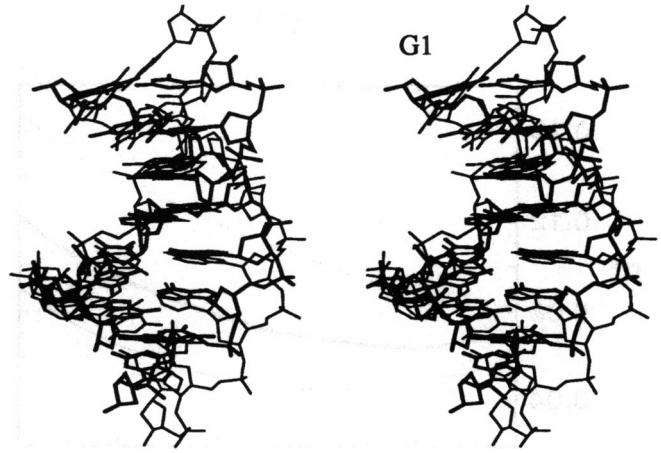

FIGURE 1 A superposition of conformations $A2_{fin}$ (*thick line*) and N1 (*thin line*). All protons are omitted for clarity.

of the two conformers, $A2_{fin}$ and N1. The intra-residue part of the sixth-root weighted NOE index R^x at 100 ms decreased for the 40–60% mixture of the N1 and $A2_{fin}$ conformations (Fig. 2 *a*). However, because of the inter-residue component, the total R^x value did not indicate any significant improvement over the pure $A2_{fin}$ conformation for any ratio of the N1 conformation, p(N1) (Fig. 2 *a*). Other residual indices, R^x and crystallographic-type index R calculated for the data sets at 100 and 150 ms displayed a similar behavior (not shown). Scalar coupling constants, which depend on sugar pucker only, agreed best with the experimental data at a ratio p(N1) = 20–30% (Fig. 2 *b*). Intra-residue base-to-sugar interproton distances depend on sugar pucker and glycosidic conformation; the corresponding distance restraints were also satisfied better for the $A2_{fin}$-N1 equilibrium. Nevertheless, overall, such an equilibrium was not consistent with the experimental data, as it did not satisfy inter-residue distance restraints (nor the corresponding NOE intensities) very well. Indeed, there is no ground to expect that a potential S-N equilibrium of sugar puckers involves a cooperative "all-or-nothing" transition between A- and B-forms of DNA. More probably, such an equilibrium can be achieved by local sugar repuckerings in individual residues. To test this hypothesis, we had to calculate potential conformers with N-type sugar puckers in all individual residues.

Calculation of the "local" N-conformers

Previous conformational calculations have revealed that the S-to-N repuckering is stereochemically possible for a single deoxyribose within the framework of the B form (Gorin et al., 1990); similar local repuckerings have been observed during MD-tar simulations of the DNA octamer (Schmitz et al., 1993). To calculate local N-conformers for our DNA octamer, we constructed several subsets of distance restraints, one for each nonterminal residue. Each subset was designed similar to the global one described above, but the changes in restraints affected only one residue at a time. Namely, each subset included base proton-to-H3′ distance restraints for one particular residue and excluded restraints involving H1′,

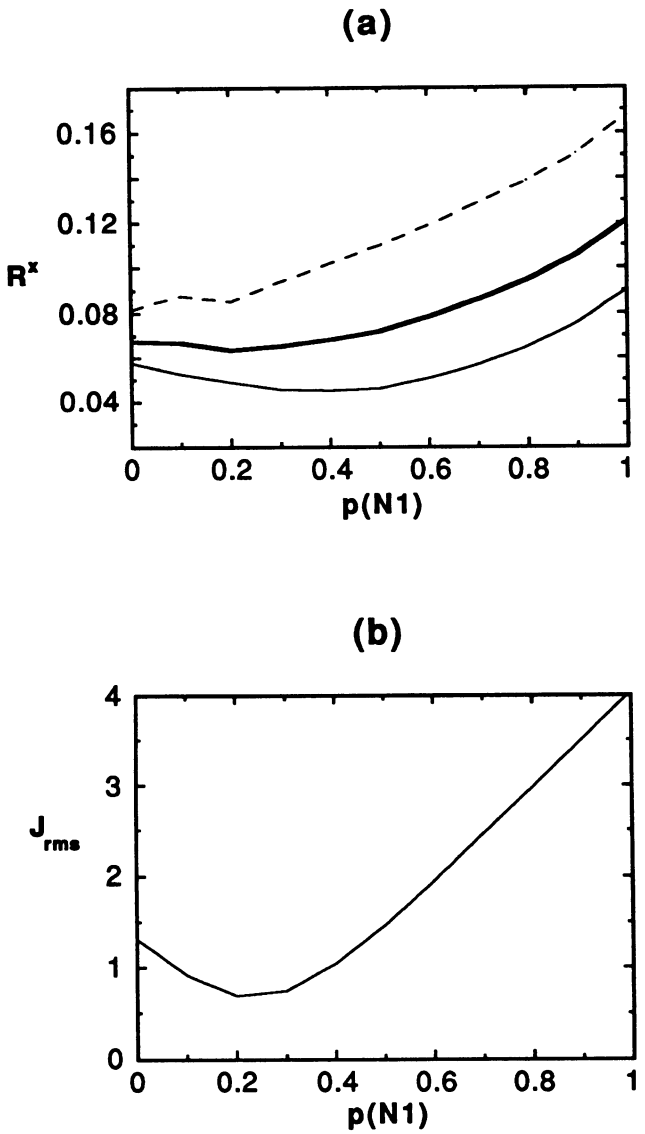

FIGURE 2 Figures of merit for a two-state equilibrium of the two conformations, $A2_{fin}$ and N1, as a function of the probability of the N1 conformation, p(N1). (*a*) The sixth-root-weighted 2D NOE residual index R^x calculated for the 100-ms dataset. The total value of the R^x and its inter- and intra-residue components are shown in thick, dashed, and thin lines, respectively. (*b*) The RMS deviation between experimental and calculated scalar coupling constants, J_{rms} (Hz).

H2′, and H2″ protons, whereas the rest of the restraints for the molecule were the same as in the main set.

Each subset of restraints was used for refinement via the rMC-simulated annealing procedure, starting from the $A1_{fin}$ conformation. (The $A1_{fin}$ conformation was calculated by the rMC refinement (Ulyanov et al., 1993), and it is very close to the $A2_{fin}$ with an atomic RMS deviation of 0.3 Å.) During the subsequent rMC simulations, only the local geometry of a complementary trimer was allowed to change, the one that included the affected residue; the rest of the DNA octamer was maintained, because the distance restraints were not changed there. As a result of these refinements, we obtained 12 different structures. For each nonterminal residue, except A5, the corresponding conformer had the residue with altered restraints in an N-type sugar conformation (C4′-*exo*/O1′-*endo*) with pseudorotation angle P in the range of 35° to 106°, whereas the remainder of deoxyriboses had S-type conformations (C1′-*exo*/C2′-*endo*). For the A5 residue, the resultant conformer had a C1′-*exo* sugar pucker with P = 135° (compared with P = 187° in the starting conformation). A superposition of the starting $A1_{fin}$ conformation with a typical local N-conformer, one calculated for the T4 residue, is shown in Fig. 3. Compared with the $A1_{fin}$ structure, the pseudorotation phase angle in the T4 residue decreased by 77° and the corresponding glycosidic angle decreased by 36°. Other structural changes are relatively small, and they are mainly localized in the T4-A5 step.

Calculation of the rest of the conformational pool

The remainder of the potential conformers were calculated in the three following ways.(i) Similar to the restraint subsets that led to the switching of individual sugar rings into the N-type conformation, H3′ proton-containing subsets were constructed for various combinations of residues. Refinements against each of these subsets led to distinct conformations of the DNA octamer with the corresponding deoxyriboses adopting an N-type conformation. However, not all possible combinations of residues were explored to contain the extent of calculations. Indeed, the total number of combinations of 12 nonterminal residues, 2^{12} = 4096, and even the number of combinations involving only various pairs of residues, (12 × 11)/2 = 66, is too large for a complete investigation.(ii) Several subsets of restraints were designed by excluding various combinations of inter-residue distances involving the A5 residue, which were poorly satisfied by the single best conformation refined (*vide supra*). This was done both in the context of the "S-type restraints", i.e., those containing base-to-H1′/H2′/H2″ and lacking base-to-H3′ restraints, and of the "N-type restraints" (including H3′- and excluding H1′/H2′/H2″-involving base-to-sugar distances). Refinements against each subset led to a series of

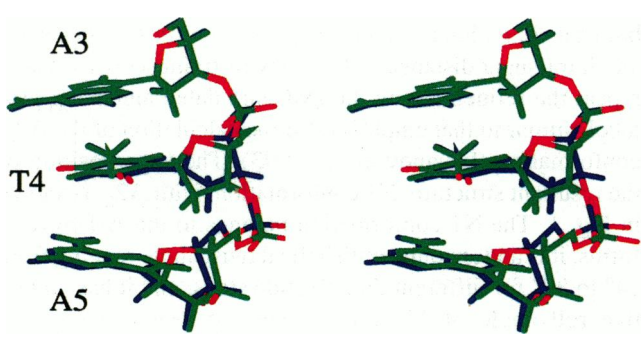

FIGURE 3 A superposition of conformation $A1_{fin}$ with a local N-conformer for residue T4 (see text), with the trinucleotide A3-T4-A5 shown. The S-type conformation $A1_{fin}$ is shown in blue, and the N-conformation is shown in green; all oxygens are colored red, and protons H5′ and H5″ are omitted for clarity. Sugar pseudo-rotation angle in residue T4 is 154° in $A1_{fin}$ and 77° in the N-type conformation.

different conformations (not shown).(iii) The biggest set of conformations was obtained via energy minimization, both restrained and unrestrained, which allowed control of the values of selected conformational parameters. Sugar pseudorotation angle P was systematically varied for each nonterminal residue by a "scanning procedure." Namely, for each residue involved, a series of conformations was calculated with P varying from 0° to 180° with a step of 10°, each conformation being optimized with respect to all degrees of freedom except pseudorotation angle P in the selected residue, which was held fixed. In total, the pool of conformations consisted of more than 500 potential conformers.

Assessment of the conformational equilibrium

The dipolar relaxation rates t_{ki} were calculated via a complete relaxation matrix for each generated conformation. Then the probabilities p_i minimizing the relaxation rate-based residual index Q^r were found as described in Theory and Methods. Two separate series of calculations were carried out with two different sets of experimental rates E_k generated from the 2D NOE datasets at 100 and 150 ms. Remarkably, the two datasets yielded two conformational equilibria consisting of almost the same sets of structures (Table 1): EQ_{100} (8 structures) and EQ_{150} (6 structures). Most of the potential conformers were eliminated at this stage and, surprisingly, the single best conformation $A2_{fin}$ was among the eliminated structures. Below we characterize briefly each structure from the equilibria EQ_{100} and EQ_{150} and the restraint subsets used for their derivation. For simplicity, we will use numbers to denote the conformers, and the rMC-refined $A2_{fin}$ will be called conformer No. 1 (Table 1).

Conformer No. 4 is a local N conformation for the A15 residue obtained by the rMC-simulated annealing protocol using restraints that included H3′ protons in the A15 (*vide supra*). The pseudorotation angle P is 35° for A15 in this conformation, compared with 146° in the $A2_{fin}$. Conformer No. 7 is a local N conformation for the T14 residue with $P = 10°$; however, in contrast to No. 4, it was obtained by a restrained minimization with fixed pseudorotation angle. Conformers No. 8 and 9 were obtained starting from global N-conformation N1 (*vide supra*) by 10,000 iterations of rMC at 300 K and a consecutive structural averaging; the helical parameter-based averaging procedure was described in detail elsewhere (Ulyanov et al., 1993). In both cases, only helical parameters of pairs A5·T12, A6·T11, and of steps T4-A5·T12-A13, A5-A6·T11-T12, A6-T7·A10-T11 were allowed to change during the simulations, whereas the rest of the octamer was kept fixed. The restraint sets were similar to that used for the calculation of the N1 conformation (*vide supra*); however, some modifications were made. For conformer No. 8, a single distance restraint, H2″T4-H8A5, was added to the set. In the case of conformer No. 9, all restraints pertinent to the A5-A6·T11-T12 dinucleotide set were removed, except two distances, H8A5-H8A6 and H2A5-H2A6. Among other changes, this caused a conversion of the A5 residue back to C2′-*endo* conformation with $P = 146°$.

The three structures, No. 3, 5, and 6, were obtained by a rMC-simulated annealing starting from a B-type conformation; various subsets of inter-residue distance restraints were applied in all three cases, whereas all intra-residue restraints were removed. In the two first cases, the restraint sets included H2′/H2″-H6/H8/CH_3 and the distances between base protons. In the case of No. 6, the set consisted of H3′-H6/H8/CH_3 and the distances between base protons. The only difference between the sets employed for the calculation of the No. 3 and 5 conformers involved the H8A5-H8A6 and H2A5-H2A6 distances. For the former structure, the H8A5-H8A6 restraint was removed from the set, and for the latter, the H2A5-H2A6 restraint was removed. Despite such small differences in the restraint sets, the resultant structures were quite distinct: No. 3 was bent into the minor groove in the A5-A6·T11-T12 step, whereas No. 5 was bent into the major groove at the same site; both conformers belong to the B-family of forms. The No. 6 conformer is mostly A-form with C3′-*endo* sugar puckers, although residues A5, G8, A10, A15, and C16 exhibit the C2′-*endo* sugar conformation, and residues A6 and C9 have C4′-*exo*/O1′-*endo* sugar puckers.

And finally, conformer No. 2 was obtained by a restrained energy minimization of the $A2_{fin}$ conformation using a small set of restraints, those that were satisfied the worst by $A2_{fin}$. The sugar pseudorotation angles of all nine conformers including $A2_{fin}$ are given in Table 2; the pair-wise atomic RMS deviations between these conformers are listed in Table 3.

Conformational equilibria EQ_{100} and EQ_{150} (Table 1) have been found by minimizing the relaxation rate-based residual index Q^r-(*vide supra*). Because of the exponential relationship between relaxation rates and NOE values, the minimizer of the Q^r-index does not necessarily minimize the NOE-based figures of merit. In our case, the crystallographic-type R-factors decreased for the distributions EQ_{100} and EQ_{150} compared with the single best structure $A2_{fin}$, whereas the sixth-root weighted R^x-factors did not (Table 4). To check, if the NOE-based R-factors can be further improved, we attempted to minimize them by modifying the conformers' distribution. We added the conformation $A2_{fin}$ to the equilibrium EQ_{100} and systematically varied the probabilities of conformers. For this purpose, two conformers were selected

TABLE 1 Probabilities of conformers in three equilibrium distributions

Conformer	EQ_{100}	EQ_{150}	EQ_{fin}
1	0	0	0.325
2	0.389	0.462	0.078
3	0.151	0.218	0.145
4	0.022	0	0.171
5	0.337	0.131	0.144
6	0.001	0	0.027
7	0.014	0.104	0.090
8	0.066	0.067	0.010
9	0.021	0.018	0.010

All conformers are listed, which have a nonzero probability in one of the two equilibria EQ_{100} or EQ_{150}, together with $A2_{fin}$ structure (conformer No. 1).

TABLE 2 Pseudorotation phase angles in nine conformers

Residue	1	2	3	4	5	6	7	8	9	Aver.*	% S‡
G1	168	168	165	155	172	9	167	23	23	159	95.3
T2	133	139	112	128	96	21	125	39	39	119	95.3
A3	149	162	146	141	153	14	147	55	55	143	95.3
T4	161	159	93	154	78	29	147	64	61	131	80.9
A5	187	162	151	182	167	164	170	45	146	172	99.0
A6	125	115	127	128	119	73	132	57	109	123	96.3
T7	152	118	103	161	110	34	144	44	44	132	95.3
G8	147	149	110	150	136	155	142	22	22	138	98.0
C9	78	82	159	144	118	72	90	62	62	107	46.0
A10	144	146	115	156	121	146	155	21	21	137	98.0
T11	145	138	149	138	120	43	137	36	28	135	95.3
T12	152	151	87	149	59	39	136	50	97	123	67.4
A13	153	140	171	154	173	16	150	59	60	152	95.3
T14	165	125	94	166	125	30	10	54	54	126	86.3
A15	146	80	150	35	144	159	160	45	45	122	73.2
C16	80	46	91	122	99	141	94	48	48	91	57.7

Pseudorotation angles are given in degrees; conformer numbers are defined in Table 1.
*Average pseudorotation angles for the equilibrium EQ_{fin}.
‡Percentage of S-conformers for the equilibrium EQ_{fin}, defined as conformers with pseudorotation angle in the range of 90° to 180°.

TABLE 3 Atomic RMS deviations (Å) between nine conformers

*	1	2	3	4	5	6	7	8	9
1	*	1.84	1.45	1.08	1.76	2.89	0.66	3.02	2.78
2	1.55	*	2.06	1.86	1.23	2.19	1.80	2.22	1.94
3	1.26	1.78	*	1.57	2.03	3.12	1.16	3.19	2.97
4	0.61	1.47	1.17	*	1.86	2.80	1.12	2.86	2.62
5	1.55	0.96	1.66	1.47	*	1.56	1.75	1.76	1.52
6	2.67	2.00	2.78	2.59	1.50	*	2.83	1.08	1.16
7	0.62	1.46	1.00	0.69	1.46	2.52	*	2.93	2.70
8	2.67	1.97	2.72	2.53	1.59	1.01	2.46	*	0.62
9	2.38	1.64	2.47	2.23	1.29	1.05	2.18	0.65	*

*Identification of conformers is given in Table 1. The upper right triangle of the table shows RMS deviations calculated for all atoms in the octamer; the values calculated for the heavy atoms in nonterminal residues are given in the lower left triangle.

TABLE 4 Indices of agreement for the equilibria of conformers

Equili brium	N*	J_{rms} (Hz)	$R^x \times 100$ (100 ms)			$R^x \times 100$ (150 ms)			R (100 ms)			R (150 ms)		
			Total	Intra	*Inter*	**Total**	Intra	*Inter*	**Total**	Intra	*Inter*	**Total**	Intra	*Inter*
$A2_{fin}$	1	1.32	**6.78**	5.69	*8.41*	**7.06**	6.57	*7.77*	**0.343**	0.303	*0.446*	**0.400**	0.383	*0.438*
EQ_{100}	8	0.68	**6.60**	4.76	*9.36*	**7.49**	5.85	*9.93*	**0.277**	0.202	*0.468*	**0.333**	0.272	*0.474*
EQ_{150}	6	0.71	**6.68**	4.78	*9.53*	**7.50**	5.82	*9.99*	**0.278**	0.200	*0.479*	**0.330**	0.265	*0.478*
EQ_{fin}	9	0.74	**6.00**	5.02	*7.48*	**6.93**	5.97	*8.36*	**0.270**	0.219	*0.399*	**0.321**	0.280	*0.415*
MD-tar-subs	400	0.77	**5.92**	4.57	*7.96*	**6.17**	4.89	*8.06*	**0.318**	0.244	*0.510*	**0.322**	0.251	*0.485*

Index J_{rms} gives RMS deviation between experimental and calculated J couplings; 2D NOE indices R^x and R are calculated separately for the datasets acquired at 100 and 150 ms; the total value of the index and its intra- and inter-residue components are printed in bold, normal, and italic fonts, respectively.
*Number of conformers in an equilibrium.

out of the total nine, and their probabilities were changed in small increments such that the sum of all nine probabilities remained unit. After the R-factors were minimized with respect to the probabilities of these two conformers, another pair of conformers was selected; the process was continuously repeated until R-factors ceased to decrease. This procedure was performed semi-automatically; it was possible to carry out such a minimization with the relatively small number of conformers involved. The resulting distribution of conformers is listed as EQ_{fin} in Tables 1 and 4; the superposition of all nine structures is shown in Fig. 4. Interestingly, this minimization led to an improvement of the sixth-root weighted R^x factors, whereas R-factors remained almost the same. For all NOE-based indices, a redistribution of intra- and inter-residue components was observed, the former component increasing and the latter decreasing (Table 4).

DISCUSSION

Making use of internal inconsistencies in NMR restraints, we were able to calculate a number of distinct conformations for the DNA octamer, which better satisfy various subsets of distance restraints. Minimizing a relaxation rate-based residual index by a quadratic programming algorithm and sub-

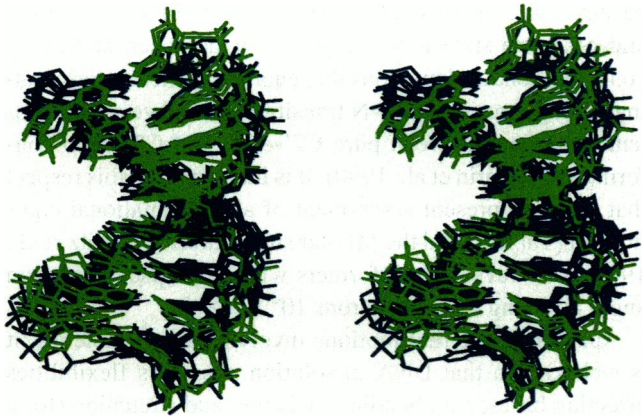

FIGURE 4 A superposition of the nine final conformers. Residues with S-type sugar pucker are shown in blue (pseudo-rotation angle > 90°), and those with N-type pucker are shown in green (pseudo-rotation angle < 90°, see Table 2). All protons are omitted; residue G1 is located in the upper left corner.

sequently optimizing the 2D NOE-based figures of merit, we found a distribution of conformers which fits experimental NOE intensities better as an ensemble than any single structure (compare the equilibrium EQ_{fin} with the single best conformation $A2_{fin}$ in Table 4). How unique is this ensemble description of DNA in solution?

Recently, a different ensemble for the same DNA octamer has been calculated via MD-tar simulations (Schmitz et al., 1993). The MD-tar does not enforce simultaneous satisfaction of all distance restraints in each structure along the MD trajectory; rather it generates an ensemble of structures satisfying restraints on average. Four hundred structures have been produced by the MD-tar calculations with explicit water molecules; 2D NOE intensities simulated for this ensemble fit experimental data better than any single structure (Schmitz et al., 1993). *R*-factors for the MD-tar ensemble are listed in Table 4 as $MD\text{-}tar_{subs}$; the subscript "subs" refers to the fact that the protons in all structures were substituted with idealized bond length and bond angle geometries before NOE simulations. Such a substitution has been made for a better compatibility of the AMBER- and DNAminiCarlo-generated structures; the details and rationale of the procedure are described elsewhere (Ulyanov et al., 1993). It is seen that the *R*-factors are roughly comparable for the equilibria EQ_{fin} and $MD\text{-}tar_{subs}$ (EQ_{fin} having somewhat lower crystallographic-type indices *R* and higher sixth-root-weighted indices R^x, Table 4), despite the fact that the two ensembles are quite different. In short, the MD-tar-generated ensemble is much "fuzzier" than EQ_{fin}: it has a very broad distribution of backbone torsion angles with representatives from the nonstandard B_{II} and "*trans*" families of forms, the average percentage of the major S sugar conformers in the 12 nonterminal residues is 78% compared with 90% in EQ_{fin}, etc. This is a rather disappointing conclusion, because it shows that, at the present stage, we were not able to identify *uniquely* the structural ensemble for the DNA octamer in solution, unlike the refinement of the average solution structure (*vide supra*). Moreover, the very fact that two quite different structural ensembles agree to a similar extent with the experimental data, probably, indicates that neither ensemble is correct in detail. This is further corroborated by a close inspection of the agreement indices (Table 4). Indeed, both EQ_{fin} and MD-tar ensembles have lower *R*-factors compared with any single structure, but this improvement was not very dramatic. In addition, it is interesting to address the difference between ensembles EQ_{100} and EQ_{150} on the one hand, which are optimal with respect to the relaxation rate-based Q^r-index, and EQ_{fin} on the other hand, which has been refined against NOE intensity-based indices. In an ideal situation, if we had noise-free and error-free data, and if we had the "true" ensemble, both type of indices should be zero. However, in a real situation, because of the exponential relationship between dipolar relaxation rates and NOE magnitudes, refinement against different kinds of indices leads to different ensembles, which is another indication that the "true" ensemble is yet to be found. EQ_{100} and EQ_{150} have slightly lower intra-residue components of *R*- and R^x-factors, but their inter-residue components are higher than in EQ_{fin}; we found that it was necessary to include $A2_{fin}$ in equilibria EQ_{100} and EQ_{150} to improve the inter-residue components of the NOE-based indices (Tables 1 and 4).

As evident from the above, the calculated ensembles do not represent at the present stage the "true" Boltzmann distribution of the DNA octamer structures in solution. Is there any useful and reliable information at all in the multi-conformational description of the DNA structure (generated by PDQPRO- and/or MD-tar)? Yes, there is. First, there are a number of common features in both descriptions of a conformational equilibrium, the most notable being an S-N equilibrium of sugar conformations. How do we know that it is not a mere coincidence? The existence of a dynamic repuckering of sugar rings is confirmed by a decrease in J_{rms}, the RMS deviation between calculated and experimental proton-proton *J* coupling constants, for the ensembles compared with single structures (Table 4). We remind that experimental scalar coupling was not involved in calculation of the conformational equilibria, but was used as an independent test (*vide supra*). This approach is similar in philosophy to the cross-validation method introduced by Brünger (Brünger, 1992; Brünger et al., 1993), scalar coupling constants representing the test set of data in our case. Using subsets of distance restraints (or NOE intensities) as test sets requires a sufficiently high redundancy of experimental data (Weisz et al., 1994), especially in the case of flexible molecules in solution, when refinement against different subsets of restraints leads to distinct conformers (*vide supra*). The distribution of S and N sugar puckers was observed for all equilibria calculated in the present work and by the MD-tar simulations (Schmitz et al., 1993), accompanied by a decrease in J_{rms}-index (Table 4). We regard this decrease to be significant, especially considering that *J* couplings were excluded from refinement. The accuracy of parametrization of the Karplus equations was estimated to be 0.5 Hz (Haasnoot et al., 1980), and the uncertainty in scalar coupling constant

determination from 2DF-COSY spectra simulations varies from ±0.2 to ±0.4 Hz (Schmitz et al., 1992b), which suggests that the J_{rms}-index, currently being of ~0.7 Hz, could be further improved. It has been discussed recently that observed three-bond couplings may be underestimated because of the influence of dipolar cross-relaxation on the apparent coupling, the effect being more pronounced for smaller J couplings (Harbison, 1993). If so, the increase in the smallest observed coupling, J(H3′,H2″), currently estimated as 2–3 Hz (Schmitz et al., 1992b), would probably require a somewhat higher percentage of the minor N conformers for a better agreement with experimental data, because J(H3′,H2″) > 3 Hz only for pseudorotation angle $P < 90°$ and $P > 270°$.

In addition, we have identified distances that are most sensitive to the presence of minor N-conformers of deoxyriboses, namely, the H3′-to-base proton intra-residue distances. Such a distance decreases when a single sugar "pseudo-rotates" from S to N conformation within the overall B-type duplex DNA (Ulyanov et al., 1992). These distances are overestimated on average by 1.06 Å in the major $A2_{fin}$ conformer. Either an intermediate conformation with "eastern"-type deoxyriboses, or an equilibrium of S and N conformers is necessary to explain this set of distance restraints; however, the intermediate ("eastern") single conformation contradicts the rest of the restraints. The average deviation for the sixth-root-averaged H6/H8-H3′ intra-residue distances was reduced to 0.65 Å in the EQ_{fin} distribution. This deviation is still very high, which suggests that the fraction of S sugar conformers (calculated as a percentage of structures with pseudo-rotation phase angle in the range of 90° to 180°, for a given residue) is still overestimated in the EQ_{fin} equilibrium (Table 2). Both scalar coupling constants and the intra-residue component of the R^x-factor calculated for the $A2_{fin}$-N1 mixture, suggest a somewhat larger percentage of minor conformers (Fig. 2), although the inter-residue restraints are satisfied better with a smaller fraction of N sugar puckers. Also note that optimization of NOE-based R-factors was accompanied by a redistribution of inter- and intra-residue components to the advantage of the former (Table 4). This may provide an explanation of what is incorrect in the calculated ensembles: both intra-residue distance restraints and scalar couplings suggest a higher fraction of the minor N conformers, but we were unable to generate proper minor conformers accounting for all inter-residue restraints. It is plausible that to fit better both intra- and inter-residue distances, a larger pool of potential conformers must be generated with various combinations of residues in the N-conformation; however, the total number of potential combinations involved is unrealistically large (*vide supra*). A future approach to this problem may involve multiconformational characterization of local geometries, e.g., at the level of duplex dinucleotide steps. In addition, we note that the percentage of C3′-*endo* sugar puckers estimated from J-coupling using the two-state model (Schmitz et al., 1992b) is higher than the N-percentage calculated for the EQ_{fin} equilibrium (Table 2). However, one must be cautious when comparing results of structure calculations at atomic resolution and of a two-state model analysis. It has been shown by conformational calculations that energy profiles corresponding to single residue S-to-N transitions can have several potential minima between pure C2′-*endo* and C3′-*endo* conformations (Gorin et al., 1990). It is interesting in this respect that both the present assessment of a conformational equilibrium (Table 2) and the MD-tar simulations (Schmitz et al., 1993) have revealed conformers with sugar pseudorotation angle covering the range from 10° to 180°.

Apart from internal motions involving sugar moieties, it is well known that DNA in solution possesses flexibilities affecting bases, e.g., bending, twisting, and stretching (for a review, see Hagerman, 1988). However, the question is still open concerning possible contributions from such motions to NMR signals. Nevertheless, at least one sequence motif, 5′-TA_n, has been identified where some kind of internal motion on the sub-millisecond time-scale was manifest in line broadening and $T_{1\rho}$ values for protons at the TpA junction (Schmitz et al., 1992b; Kennedy et al., 1993). In our DNA octamer, the affected residue is A5; this may explain the fact that some inter-residue restraints involving this residue are poorly satisfied in model structures, most notably, distances H2A5-H2A6 and H8A5-H8A6 (*vide supra*). Although these two distance restraints are satisfied better in some of the nine calculated conformers (data not shown), sixth-root averaging over the EQ_{fin} ensemble exhibits very little improvement over the single $A2_{fin}$ conformation. One possible explanation for this is that we may fail to generate proper potential conformations that accurately represent internal motions involving the A5 residue. Another source of errors may be due to possible differences in the internal motions of sugars and bases that, in its turn, may require different types of averaging for the relaxation rates. Clearly, these matters deserve additional investigation, probably including heteronuclear relaxation measurements and calculations of rotating frame relaxation times $T_{1\rho}$ for potential equilibria, similar to those performed recently for a cyclic decapeptide antamanide (Blackledge et al., 1993).

It is important to note that probabilities of minor N conformers are relatively low for this DNA octamer, estimated both by the present analysis (Tables 1 and 2) and by the two-state analysis of vicinal scalar couplings in deoxyriboses (Schmitz et al., 1992b). Conformers No. 6, 8, and 9, which differ the most from the single best conformation $A2_{fin}$ (see Table 3), have very low probability in all calculated equilibria (Table 1). The average pseudorotation angles for the EQ_{fin} equilibrium are very close to the pseudorotation angles in $A2_{fin}$ (Table 2); such averages could be meaningless if the probabilities of the minor conformers were much higher. All things considered, the best rigid structure obtained by a conventional restrained refinement remains a good approximation of the DNA octamer in solution. A similar conclusion has been recently drawn based on a completely different methodology for the generation of solution conformational ensembles, MD-tar calculations (Schmitz et al., 1993).

Our approach to assess conformational equilibria for DNA in solution resembles the MEDUSA method developed by Ernst and co-authors (Brüschweiler et al., 1991; Blackledge et al., 1993). The MEDUSA algorithm has the advantage of exhaustively testing various combinations of distance restraints. Thus, in an application to antamanide (Blackledge et al., 1993), the calculation of each potential conformer consisted of randomly ordered consecutive addition of distance restraints, one at a time, of the total list of 23 restraints, with restrained minimization of the structure at each addition. After each addition and minimization, the conformer was accepted or rejected based on an energy criterion. In our case, however, the restraint set consisted of 184 interproton distances, which practically precluded construction of all possible subsets; instead, "educated guesses" were used to design subsets of distances. On the other hand, our approach has the advantage of rigorous calculation of the conformer's probabilities minimizing the relaxation rate-based residual index after the pool of potential conformations was constructed. This is done by the PDQPRO program using a constrained global minimization by a quadratic programming algorithm described by Fletcher (1981). This method can be applied to the analysis of a multi-conformational equilibrium and is not limited to the consideration of small groups of potential conformers, which is especially important for molecules with larger degrees of flexibility, such as DNA.

CONCLUSIONS

We have demonstrated a method, termed PARSE, for the determination of probabilities of potential conformers in the solution ensemble via global optimization of the relaxation rate-based residual index Q'. The method involves interaction of the optimization program PDQPRO and relaxation matrix analysis programs MARDIGRAS and CORMA, which calculate dipolar relaxation rates from experimental 2D NOE data and for theoretical structures, respectively. This method is independent of the approach used for generation of conformer pool; therefore, it can be applied to any molecule once the initial pool of conformers is constructed.

At this stage, we cannot unambiguously determine the structural ensemble for the Pribnow box DNA octamer in solution. Indeed, two different ensembles emerged from the present work and from our previous MD-tar simulations. Besides, even though the agreement with the experimental data is better for each of the two descriptions than for any single structure, some of the experimental restraints are still not completely satisfied. It shows that characterization of the solution ensemble can and must be further improved. However, we have obtained some reliable knowledge about the conformational flexibility of the DNA octamer in solution. Most importantly, the S-N equilibrium of sugar ring conformations must be taken into account to explain internal inconsistencies in intra-residue distance restraints involving H3′ and H2′/H2″ protons. Both MD-tar simulations and the present work produced a distribution of DNA conformers that included S- and N-type sugar conformations for each residue. A similar equilibrium of sugar puckers follows also from analysis of independent experimental data, inter-proton scalar coupling constants in sugar rings. However, we consider our new approach as a significant step forward from the traditional two-state analysis of vicinal coupling constants in sugars, because we are developing the tools for incorporating contributions from the structures of minor conformers. At the same time, because of a relatively small population of minor N-conformers, a rigid structure, obtained via conventional restrained optimization with 2D NOE-derived restraints, remains a good approximation of the DNA octamer in solution.

This work was supported by National Institutes of Health Grants GM41639, GM39247, and RR01695. We thank Drs. He Liu and Klaus Weisz for many helpful discussions. We gratefully acknowledge use of the UCSF Computer Graphics Laboratory (supported by NIH Grant RR01081). We also acknowledge use of the Cray C90 supercomputer, supported by a grant from the Pittsburgh Supercomputing Center through the NIH National Center for Research Resources cooperative agreement U41RR6009, and a grant from the National Science Foundation Cooperative Agreement ASC-8500650.

REFERENCES

Baikalov, I., K. Grzeskowiak, K. Yanagi, J. Quintana, and R. E. Dickerson. 1993. The crystal structure of the trigonal decamer C-G-A-T-C-G-6meA-T-C-G: A B-DNA helix with 10.6 base-pairs per turn. *J. Mol. Biol.* 231: 768–784.

Bishop, K. D., F. J. H. Blocker, W. Egan, and T. L. James. 1994. Hepatitis B virus direct repeat sequence: imino proton exchange rates and distance and torsion angle restraints from NMR. *Biochemistry.* 33:427–438.

Blackledge, M. J., R. Brüschweiler, C. Griesinger, J. M. Schmidt, P. Xu, and R. R. Ernst. 1993. Conformational backbone dynamics of the cyclic decapeptide antamanide. Application of a new multiconformational search algorithm based on NMR data. *Biochemistry.* 32:10960–10974.

Boelens, R., T. M. G. Koning, G. A. van der Marel, J. H. van Boom, and R. Kaptein. 1989. Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach. Application to a DNA octamer. *J. Magn. Reson.* 82:290–308.

Borgias, B. A., and T. L. James. 1989. Two-dimensional nuclear Overhauser effect: complete relaxation matrix analysis. *Methods Enzymol.* 176: 169–183.

Borgias, B. A., and T. L. James. 1990. MARDIGRAS: A Procedure for Matrix Analysis of Relaxation for Discerning GeometRy of an Aqueous Structure. *J. Magn. Reson.* 87:475–487.

Borgias, B. A., P. D. Thomas, H. Liu, A. Kumar, and T. L. James. 1993a. CORMA (COmplete Relaxation Matrix Analysis), Version 4.0, University of California, San Francisco.

Borgias, B. A., P. D. Thomas, H. Liu, A. Kumar, and T. L. James. 1993b. MARDIGRAS (Matrix Analysis of Relaxation for DIscerning GeometRy of an Aqueous Structure), Version 2.0, University of California, San Francisco.

Brünger, A. T. 1992. Free R-value—a novel statistical quantity for assessing the accuracy of crystal structures. *Nature.* 355:472–475.

Brünger, A. T., G. M. Clore, A. M., Gronenborn, R. Saffrich, and M. Nilges. 1993. Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. *Science.* 261:328–331.

Brüschweiler, R., M. Blackledge, and R. R. Ernst. 1991. Multi-conformational peptide dynamics derived from NMR data: a new search algorithm and its application to antamanide. *J. Biomol. NMR.* 1:3–11.

Celda, B., H. Widmer, W. Leupin, W. J. Chazin, W. A. Denny, and K. Wüthrich. 1989. Conformational studies of d-$(AAAAATTTTT)_2$ using constraints from nuclear Overhauser effects and from quantitative analysis of the cross-peak fine structures in two-dimensional ^{1}H nuclear magnetic resonance spectra. *Biochemistry.* 28:1462–1471.

Davies, D. B., and S. S. Danyluk. 1975. Nuclear magnetic resonance studies of 2′- and 3′-ribonucleotide structures in solution. *Biochemistry.* 14: 543–554.

Fesik, S. W., T. J. O'Donnell, R. T. Gampe, Jr., and E. T. Olejniczak. 1986. Determining the structure of a glycopeptide Ac_2-Lys-D-Ala-D-Ala complex using NMR parameters and molecular modelling. *J. Am. Chem. Soc.* 108:3165–3170.

Fletcher, R. 1981. Practical Methods of Optimization. Vol. 2. Constrained Optimization. Wiley & Sons, New York. 220 pp.

Gorin, A. A., N. B. Ulyanov, and V. B. Zhurkin. 1990. S-N transition of the sugar ring in B-form DNA. *Mol. Biol. (Eng. transl.)* 24:1036–1047.

Haasnoot, C. A. G., F. A. A. M. de Leeuw, and C. Altona. 1980. The relationship between proton-proton NMR coupling constants and substituent electronegativities-I. *Tetrahedron.* 36:2783–2792.

Hagerman, P. J. 1988. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* 17:265–286.

Hagerman, P. J. 1990. Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* 59:755–781.

Harbison, G. S. 1993. Interference between J-couplings and cross-relaxation in solution NMR spectroscopy—consequences for macromolecular structure determination. *J. Am. Chem. Soc.* 115:3026–3027.

James, T. L. 1991. Relaxation matrix analysis of 2D NOE spectra to obtain accurate biomolecular structural constraints and to assess structural quality. *Curr. Opin. Struct. Biol.* 1:1042–1053.

James, T. L., M. Gochin, D. J. Kerwood, D. A. Pearlman, U. Schmitz, and P. D. Thomas. 1991. Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy. J. C. Hoch, F. M. Poulsen, and C. Redfield, editors. Plenum Press, New York. 331–348.

Jardetsky, O. 1980. On the nature of molecular conformations inferred from high-resolution NMR. *Biochim. Biophys. Acta.* 621:227–232.

Keepers, J. W., and T. L. James. 1984. A theoretical study of distance determinations from NMR. Two-dimensional nuclear Overhauser effect spectra. *J. Magn. Reson.* 57:404–426.

Kennedy, M. A., S. T. Nuutero, J. T. Davis, G. P. Drobny, and B. R. Reid. 1993. Mobility at the TpA cleavage site in the T_3A_3-containing AhaIII and PmeI restriction sequences. *Biochemistry.* 32:8022–8035.

Kessler, H., C. Griesinger, J. Lautz, A. Müller, W. F. van Gunsteren, and H. J. C. Berendsen. 1988. Conformational dynamics detected by nuclear magnetic resonance NOE values and *J* coupling constants. *J. Am. Chem. Soc.* 110:3393–3396.

Kim, Y., and J. H. Prestegard. 1989. A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry.* 28:8792–8797.

de Leeuw, F. A. A. M., A. A. van Beuzekom, and C. Altona. 1983. Through-space effects on vicinal proton spin-spin coupling constants mediated via hetero atoms: nonequivalence of cis couplings in five-membered rings. *J. Comput. Chem.* 4:438–448.

Liu, H., P. D. Thomas, and T. L. James. 1992. Averaging of cross-relaxation rates and distances for methyl, methylene and aromatic ring protons due to motion or overlap: extraction of accurate distances iteratively via relaxation matrix analysis of 2D NOE spectra. *J. Magn. Reson.* 98:163–175.

Macaya, R., E. Wang, P. Schultze, V. Sklenar, and J. Feigon. 1992. Proton nuclear magnetic resonance assignments and structural characterization of an intramolecular DNA triplex. *J. Mol. Biol.* 225:755–773.

Madrid, M., E. Llinas, and M. Llinas. 1991. Model-independent refinement of interproton distances generated from ^{1}H-NMR Overhauser intensities. *J. Magn. Reson.* 93:329–346.

Mujeeb, A., S. M. Kerwin, W. Egan, G. L. Kenyon, and T. L. James. 1992. A potential gene target in HIV-1: Rationale, selection of a conserved sequence, and determination of NMR distance and torsion angle constraints. *Biochemistry.* 31:9325–9338.

Mujeeb, A., S. M. Kerwin, W. M. Egan, G. L. Kenyon, and T. L. James. 1993. Solution structure of a conserved DNA sequence from the HIV-1 genome: restrained molecular dynamics simulation with distance and torsion angle restraints derived from 2D NMR spectra. *Biochemistry.* 32: 13419–13431.

Pearlman, D. A., and P. A. Kollman. 1991. Are time-averaged restraints necessary for NMR refinement: a model study for DNA. *J. Mol. Biol.* 220:429–457.

Poltev, V. I., and N. V. Shulyupina. 1986. Simulation of interactions between nucleic acid bases by refined atom-atom potential functions. *J. Biomol. Struct. Dyn.* 3:739–765.

Post, C. B., R. P. Meadows, and D. G. Gorenstein. 1990. On the evaluation of interproton distances for three-dimensional structures determination by NMR using a relaxation rate matrix analysis. *J. Am. Chem. Soc.* 112: 6796–6803.

Rinkel, L. J., and C. Altona. 1987. Conformational analysis of the deoxyribose ring in DNA by means of sums of proton-proton coupling constants: A graphical methods. *J. Biomol. Struct. Dyn.* 4:621–649.

Schmitz, U., G. Zon, and T. L. James. 1990. Deoxyribose conformation in $[d(GTATATAC)]_2$: evaluation of sugar pucker by simulation of double-quantum-filtered COSY cross-peaks. *Biochemistry.* 29:2357–2368.

Schmitz, U., D. A. Pearlman, and T. L. James. 1991. Solution structure of $[d(GTATATAC)]_2$ via restrained molecular dynamics simulations with NMR constraints derived from relaxation matrix analysis of 2D NOE experiments. *J. Mol. Biol.* 221:271–192.

Schmitz, U., A. Kumar, and T. L. James. 1992a. Dynamic interpretation of NMR data: molecular dynamics with weighted time-averaged restraints and ensemble R-factor. *J. Am. Chem. Soc.* 114:10654–10556.

Schmitz, U., I. Sethson, W. M. Egan, and T. L. James. 1992b. Solution structure of a DNA octamer containing the Pribnow box via restrained molecular dynamics simulation with distance and torsion angle constraints derived from 2D NMR spectral fitting. *J. Mol. Biol.* 227:510–531.

Schmitz, U., N. B. Ulyanov, A. Kumar, and T. L. James. 1993. Molecular dynamics with weighted time-averaged restraints for a DNA octamer: dynamic interpretation of NMR data. *J. Mol. Biol.* 234:373–89.

Steitz, T. A. 1990. Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophys.* 23:205–280.

Stolarski, R., W. M. Egan, and T. L. James. 1992. Solution structure of the *Eco*RI DNA octamer containing 5-fluorouracil via restrained molecular dynamics using distance and torsion angle constraints extracted from NMR spectral simulations. *Biochemistry.* 31:7027–7042.

Torda, A. E., R. M. Scheek, and W. F. van Gunsteren. 1990. Time-average nuclear Overhauser effect distance restraints applied to tendamistat. *J. Mol. Biol.* 214:223–235.

Trifonov, E. N. 1985. Curved DNA. *CRC Crit. Rev. Biochem.* 19:89–106.

Ulyanov, N. B., A. A. Gorin, and V. B. Zhurkin. 1989. Conformational mechanics of the DNA double helix. A combined Monte Carlo and energy minimization approach. *In* Proceedings Supercomputing '89: Supercomputer Applications. L. P. Kartashev, and S. I. Kartashev, editors. Int. Supercomputing Inst., Inc., St. Petersburg, FL. 368–370.

Ulyanov, N. B., A. A. Gorin, V. B. Zhurkin, B.-C. Chen, M. H. Sarma, and R. H. Sarma. 1992. Systematic study of nuclear Overhauser effects vis-á-vis local helical parameters, sugar puckers, and glycosidic torsions in B DNA: insensitivity of NOE to local transitions in B DNA oligonucleotides due to internal structural compensations. *Biochemistry.* 31: 3918–3930.

Ulyanov, N. B., U. Schmitz, and T. L. James. 1993. Metropolis Monte Carlo calculations of DNA structure using internal coordinates and NMR distance restraints: an alternative method for generating high-resolution solution structure. *J. Biomol. NMR.* 3:547–568.

Ulyanov, N. B., and T. L. James. 1994. Statistical analysis of DNA duplex structures in solution derived by high resolution NMR. *Appl. Magn. Reson.* 7:21–42.

Weisz, K., R. H. Shafer, W. Egan, and T. L. James. 1992. The octamer motif in immunoglobulin genes: Extraction of structural constraints from two-dimensional NMR studies. *Biochemistry.* 31:7477–7487.

Weisz, K., R. H. Shafer, W. M. Egan, and T. L. James. 1994. Solution structure of the octamer motif in immunoglobulin genes via restrained molecular dynamics calculations. *Biochemistry.* 33:354–366.

Wüthrich, K. 1986. NMR of Proteins and Nucleic Acids. Wiley & Sons, New York. 292 pp.

Zhurkin, V. B., V. I. Poltev, and V. L. Florentiev. 1981. Atom-atom potential functions for conformational calculations of nucleic acids. *Mol. Biol. (Eng. transl.)* 14:882–895.

Zhurkin, V. B., Yu. P. Lysov, V. L. Florentiev, and V. I. Ivanov. 1982. Torsional flexibility of B-DNA as revealed by conformational analysis. *Nucleic Acids Res.* 10:1811–1830.

Zhurkin, V. B., N. B. Ulyanov, A. A. Gorin, R. L. Jernigan. 1991. Static and statistical bending of DNA evaluated by Monte Carlo simulations. *Proc. Natl. Acad. Sci. USA.* 88:7046–7050.