Journal of Biomolecular NMR, **24:** 317–328, 2002.
KLUWER/ESCOM

Residual dipolar coupling constants and structure determination of large DNA duplexes

Olivier Mauffret, Georges Tevanian & Serge Fermandjian*
Département de Biologie et Pharmacologie Structurales, UMR 8532 CNRS, Institut Gustave-Roussy, 94800 Villejuif and ENS Cachan, 94223 Cachan, France

Received 16 June 2002; Accepted 2 October 2002

Key words: DNA, NMR, residual dipolar couplings, simulations, structures

Abstract

Several NMR works have shown that long-range information provided by residual dipolar couplings (RDCs) significantly improve the global structure definition of RNAs and DNAs. Most of these are based on the use of a large set of RDCs, the collect of which requires samples labeled with ^{13}C, ^{15}N, and sometimes, ^{2}H. Here, we carried out torsion-angle dynamics simulations on a non-self complementary DNA fragment of 17 base-pairs, d(GGAAAATATCTAGCAGT).(ACTGCTAGAGATTTTCC). This reproduces the U5 LTR distal end of the HIV-1 cDNA that contains the enzyme integrase binding site. Simulations aimed at evaluating the impact of RDCs on the structure definition of long oligonucleotides, were performed in incorporating (i) nOe-distances at both < 4.5 Å and < 5 Å; (ii) a small set of ^{13}C-^{1}H RDCs, easily detectable at the natural abundance, and (iii) a larger set of RDCs only accessible through the ^{13}C labeling of DNAs. Agreement between a target structure and a simulated structure was measured in terms of precision and accuracy. Results allowed to define conditions in which accurate DNA structures can be determined. We confirmed the strong impact of RDCs on the structure determination, and, above all, we found that a small set of RDC constraints (ca. 50) detectable at the natural abundance is sufficient to accurately derive the global and local DNA duplex structures when used in conjunction with nOe-distances < 5 Å.

Introduction

Influence of residual dipolar couplings (RDCs) data on precision and accuracy of NMR nucleic acid structures and particularly of global foldings has been the objective of several studies (Vermeulen et al., 2000; Tjandra et al., 2000; Warren and Moore,2001; Mollova et al., 2000; Kuszewski et al., 2001; Bayer et al., 1999, Trantirek et al., 2000). RDCs add long-range information which is not accessible from nOes and dihedral coupling constants (Tjandra and Bax, 1997; Tolman et al., 2001; Clore and Gronenborn, 1998; Bax et al., 2001; MacDonald and Lu, 2002a). RDCs can be measured in DNA spectra using either phage or bicelle methodology (Hansen et al., 2000; Tjandra et al., 2000). Several types of RDCs, ^{1}H-^{1}H, ^{13}C-^{13}C, ^{1}H-^{13}C, ^{15}N-^{15}N and ^{1}H-^{15}N, can be used, but the collect of most of them needs preparation of ^{13}C and ^{15}N labeled molecules (Tjandra et al., 2000; Mac Donald et al., 2001; MacDonald and Lu, 2002b, Zimmer and Crothers, 1995) and sometimes of selectively deuterated molecules (Tjandra et al., 2000; MacDonald et al., 2002b). In some cases also, the information provided by ^{31}P NMR RDCs can also be very useful for structure determination (Wu et al., 2001a, b).

Here, we used the oligonucleotide, d(5′ GGAAAA TCTCTAGCA**GT** 3′). 5′ ACTGCTAGAGATTTT CC 3′), of 17 bp long and non-self complementary at the natural isotopic abundance to address the influence of RDCs on DNA structure determination. This oligonucleotide is biologically relevant as it reproduces the extremity of the HIV-1 cDNA U5 LTR (Long Terminal Repeat). It contains the attachment site of the retroviral enzyme, integrase, and can be used as DNA substrate and target in *in vitro* integration

*To whom correspondence should be addressed. E-mail: sfermand@igr.fr

assays. Integrase excises the consensus dinucleotide GT (underlined in the above sequence) at each extremity of HIV-1 cDNA before catalyzing its transfer into the cellular DNA. Specific recognition of the LTR DNA site by integrase is therefore a critical event, initiating the viral infection; to understand its mechanism requires a good knowledge of the fine structure and dynamics of the DNA site in solution.

Despite its length and complexity, our 17 bp oligonucleotide displays a NMR spectrum with a favorable resonance spreading in both carbon and proton dimensions. Nearly all the sugar H1′-C1′ and aromatic H6/8/2/5-C6/8/2/5 RDCs could be measured without help of ^{13}C-labeling or selective deuteration (not shown) and we wondered whether this small set of RDCs, when combined to conventional nOe distance and torsion angle restraints, could be enough to improve the structure determination. This is a quite general question that concerns the best applicability of the current NMR methods to large molecules.

To learn on the influence of RDCs on accuracy of structure determination required an approach based on the confrontation of the so-called 'true' DNA structure to the different DNA structures, obtained by varying the amount and types of constraints. The reference 'true' structure was generated by computer, and simulations consisting in recovering this reference structure were performed in incorporating different set of restraints determined from the reference structure. Note that no high-resolution crystal structure for an oligonucleotide of 17 bp length which could be used as reference is available at the present time. Anyhow, the use of such a structure as reference is risky as crystal packing may affect both the local and global DNA conformations.

Influence of the number and precision of nOe data along with different sets of RDCs was assessed on both the accuracy (evaluated as the root mean square deviation (rmsd) of the calculated structures versus the reference structure) and the precision (evaluated as the rmsd of the calculated structures versus an average of the calculated structures) in each different condition. The results confirm the improvement of the global fold determination of nucleic acid molecules by RDCs (Vermeulen et al., 2000; Warren and Moore, 2001; Tjandra et al., 2000; Mollova et al., 2000). Yet, the most significant feature is that a small set of RDC constraints, measurable under natural isotopic abundance conditions, may substantially improve the quality of the structure when added to the torsion angle and nOe-distance data. Results provide practical rules to determine an accurate DNA duplex structure from the combined use of RDC and classical NMR constraints (nOes and J couplings).

Materials and methods

Generating the target structure

The 17 bp oligonucleotide d (GGAAAATCTC-TAGCAGT)(CACTGTAGAGATTTTCC) was constructed under B-form in InsightII (Accelrys) and minimized through 2000 steps of Powell minimization using the CNS program and its parameter file dna_rna_allatom.param and the topology file dna_rna_allatom.top (Brünger et al., 1998; Rife et al., 1999).

Structures were minimized with hydrogen bond restraints and weak (10 kcal mol^{-1}) planarity restraints. Planarity restraints did not hamper propeller twisting of base-pairs, which could attain $-20\,^{\circ}$C. Hydrogen bond and planarity restraints were similar to those used commonly in calculations.

Distance constraints

Distance constraints were generated from the target structure to simulate an experimental data set. Several sets of data were generated with proton-proton distances either less than 4.5 Å or less than 5 Å. To obtain realistic sets of distances we discarded those distances implicating the not easily assignable H5′, H5″ protons of ribose and amino protons of guanine and adenine. We discarded also the intraresidue sugar-sugar interactions, as such interactions are difficult to assign. Distance implicating methyl protons were averaged.

In some calculations, and for sake of comparison with the study of Vermeulen et al. (2000), we used an uncertainty of $\pm$ 0.5 Å for all the distances, except those involving a methyl group for which an uncertainty of 0.5 Å was added. In other calculations we made a difference between the distances implicating exchangeable protons and those which do not implicate exchangeable protons, this leading to more realistic restraints. For the exchangeable protons we applied an uniform large uncertainty of $\pm$ 0.8 Å, as such distances are difficult to determine precisely. For the other distances the precision was varied from 5% to 50% (with 70% of the measured distances), in order to evaluate its impact on structure determination. We generated also several files with 50 to 100% of the measured distances (with a precision of 15%) in order

to test the impact of the amount of data. The error of 15% on distances seems reasonable according previous works (Wijmenga and Van Buuren, 1998; Tjandra et al., 2000).

Hydrogen bond constraints for base-pairs were taken as in Tjandra et al. (2000), that is between the proton and the hydrogen bond acceptor and between the two heavy atoms of each hydrogen bond. We used an uncertainty of $\pm$ 0.2 Å.

Torsion angle restraints

The program CURVES 5.2 (Lavery and Sklenar, 1988) was used to determine the torsion angles of the target structures which includes the following uncertainties: $\alpha \pm 50°$, $\beta \pm 50°$, $\gamma \pm 30°$, $\varepsilon \pm 50°$, $\zeta \pm 50°$, $\delta \pm 25°$, $\chi \pm 60°$.

Note that the sugars were constrained only through restraints on δ angles and the range of values chosen for the different torsion angles is rather large. To simulate realistic torsion angle restraints, we used as constraint, a sample of 70% of the angles randomly selected.

Residual dipolar coupling constraints

^{1}H-^{13}C RDC constraints for the target structure were calculated using equations described in other publications (Clore et al., 1998a; Tjandra et al., 1997; Vermeulen et al., 2001). RDCs were generated with $D_a = -20$ Hz (axial component of the molecular alignment) and $D_r = 0$ Hz (rhombic component). Random errors were added to the RDCs values, standard deviations from exact values were $\pm$ 1 Hz.

To simulate a realistic set of RDC constraints for DNA, only a certain percentage of all the possible one-bond ^{1}H-^{13}C RDCs was selected. Set A contained only 70% of the C1′-H1′ and aromatic C-H RDCs and corresponded to a realistic set of RDCs measurable in our conditions. Set B contained 45% of the C2′-H2′, C2′H2″ RDCs, 70 % of the C1′-H1′, C3′-H3′, C4′-H4′ RDCs and 70% of aromatic RDCs. This corresponds to an ensemble which could be reasonably obtained with labeled compounds.

Statistics of the restraints used are listed in Table 1.

Calculations using nOe-distance and torsion angle constraints

Structure simulations were carried out using the CNS 0.9 algorithm (Brunger et al., 1998) which includes a refinement with RDCs. The refinement uses torsion-angle dynamics in the initial stages of the calculations (Stein et al., 1997; Rice and Brunger, 1994), followed by a Cartesian molecular dynamics. Starting structures were generated by randomizing the backbone torsion angles. Twenty structures were obtained. Calculations were made using distance files with the lower bound on distances set to a value of 1.8 Å. Torsion-angle dynamics was used at 20 000 K during 80 ps (timestep of 15 fs) with only a repulsive term for van der Waals (vdW) energy and set to 1/10 of its full value. The nOe force constant was 50 kcal mol^{-1} Å^{-2} and the dihedral force constant 5 kcal mol^{-1} deg^{-2}. During another 80 ps period, the vdW repulsive term was scaled to its full value and the temperature cooled from 20 000 K to 1000 K . These two periods of torsion-angle dynamics with a simplified force field were necessary to obtain satisfactory structures from the initial randomized starting structures. From this point, Cartesian molecular dynamics were applied during 15 ps (timestep 1 fs) with a more complete force field including Lennard–Jones functions for the vdW term, an electrostatic term, as well as a planarity term for the base-pair. During this period the temperature was cooled from 1000 K to 300 K and 2000 steps of Powell minimization completed the first stage. At this time the structure generally adopted a correct folding, from which the second stage, i.e. the refinement, might be initiated. New distance files were used during this stage identical to the preceding, except that the lower bound of 1.8 Å was replaced by an uncertainty identical to that of the upper bound. Eight ps of Cartesian molecular dynamics were run at 1000 K with a nOe force constant of 50 kcal mol^{-1} Å^{2} for the old distance files and of 0.2 (initial value) scaled slowly to 50 kcal mol^{-1} Å^{2} for the new distance file. At the same time the dihedral force constant was scaled from the weak value of 5 kcal mol^{-1} deg^{-2} to 100 kcal mol^{-1} deg^{-2}. The temperature was then cooled from 1000 to 300 K in 14 ps, and an additional 8 ps dynamics was run at 300 K. During all these periods a complete force field (with the LJ and electrostatics terms) was used. A final minimization of 2000 steps was performed on the resulting structure.

Refinement with dipolar couplings

Determining the axial and rhombic components of the alignment tensor

A grid search was performed to determine, if with the different sets of used dipolar couplings it is possible

Table 1. Number and types of restraints used in the calculations (case of 70% of nOe-distances < 5 Å)

	Number	Number per residue
Intraresidue nOes	178	5
Interresidue nOes (non exchangeable protons)	321	9.4
Interresidue nOes (exchangeable protons)	80	2.3
All nOes	579	16
Dihedral	168	4.9
Planarity	17	0.5
Hydrogen bond	82	2.4
Residual dipolar couplings, set A	50	1.4
Residual dipolar couplings, set B	124	3.6

to obtain the correct values of the axial and rhombic components of the alignment tensor. The input structures for these calculations are those obtained with the classical constraints. For the grid search Da was varied from −26 to −16 (2 Hz steps) and R from 0 to 0.4 (0.1 step size). For each Da and R combination, 1 ps of torsion-angle dynamics at 400 K was used then the temperature was lowered at 100 K in 2.4 ps before 2000 steps of Powell minimisation were performed. Energies of minimised structures were computed.

A protocol refinement using RDCs was applied to folded structures calculated from distance and torsion angle constraints. Calculations were made at low temperature (200 K) using torsion-angle molecular dynamics. In a first step, we used a simplified repulsive vdW term (to its full value), and the dipolar coupling force constant was set to a low value of 0.01 kcal/mol/Hz. In a second step of 20 ps , the dipolar coupling force constant was ramped to 10 kcal/mol/Hz. Then an additional torsion-angle dynamics period of 7 ps was run at 100 K with a complete force field followed by 2000 steps of Powell minimization. During these calculations, all the other constraints were maintained with force constants set to their final values. At the end a new 2000 steps of Powell minimization was performed, with the dipolar coupling force constant lowered to 0.1 kcal/mol/Hz. This entailed a very small structure displacement (rmsd of $\sim$ 0.02 Å).

Strategy of calculations and selected molecules

Twenty structures with randomized torsion angles were generated from which calculations with classical distance and torsion angle constraints were undertaken. As detailed in Results and Discussion, different sets of constraints files were used. The results are presented for both the 10 and 15 lowest energy structures from the totality of the calculated structures (i.e., 20 structures).

Calculations using the RDC constraints were then conducted on the 10 best structures provided by the classical constraints. The final results concerned these 10 structures and not a selection of the best molecules.

Results and discussion

Determination of DNA structures using torsion angle and distance constraints

While the NMR analysis of protein structures generally follows a consensual strategy, that treating nucleic acid structures rests on various approaches which exhibit large differences among them. A parameter that widely varies, is the precision on distances. For proteins, distance restraints are often used with wide bounds, while for nucleic acids, and notably DNA, both the lower proton density and the lack of folding makes necessary a better quantification of nOes, which is achieved through the use of matrix relaxation calculations (Wijmenga and Van Buuren, 1998).

We first analyzed the impact of both the nOe-distance precision and the number of restraints on determination of DNA structures. To make the study as realistic as possible we assumed that the precision on distances varies with the distance length and is not the same for exchangeable and non-exchangeable protons. Impact of the restraint number on structure accuracy and precision, was also assessed. Torsion

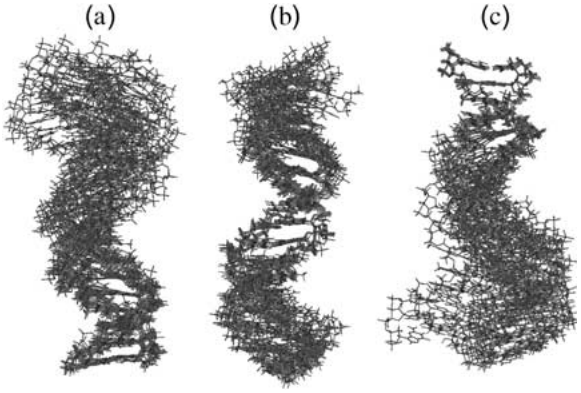

Figure 1. The ten best structures of the 17-mer duplex determined with distances < 5 Å. Superpositions are based on the three: (a) Lower base-pairs; (b) central base-pairs; (c) upper base-pairs.

angle and nOe-distance restraints were systematically used in combination as this has an impact on the structure determination (Vermeulen et al., 2000). Two nOe-distances were selected: < 4.5 Å and < 5 Å.

Calculations with nOe-distances < 4.5 Å and < 5 Å (70% of the possible distance restraints)

Calculations were performed with various lower and upper bounds. First, when lower bounds were set uniformly to 1.8 Å and upper bounds to +0.5 Å (non-exchangeable protons) and to +0.8 Å (exchangeable protons) the mean rmsd value (mean of the ten best structures versus the target structure) was found equal to 3.53 Å. The rmsds of 1.67 Å for a 10-mer and of 2.73 Å for a 14-mer obtained by Vermeulen et al. (2000) under nearly the same conditions are therefore consistent with our 3.53 Å rmsd value considering the longer size of our DNA. With a lower bound, set to −0.5 Å for non-exchangeable protons and −0.8 Å for exchangeable protons, the rmsd of accuracy is reduced to 3.07 Å. Second, when the cutoff for distance selection was increased from 4.5 to 5 Å, the latter being quite well registered by nOes, the rmsd was lowered to 2.0 Å. The latter can be considered as a realistic value, considering the available nOe-distance information. However, the so obtained structures were well defined locally but presented poor global definition (Figure 1).

Influence of precision and number of constraints on structure definition

The next question was whether the structure definition could be improved by increasing either the precision on non-exchangeable proton distances or the number

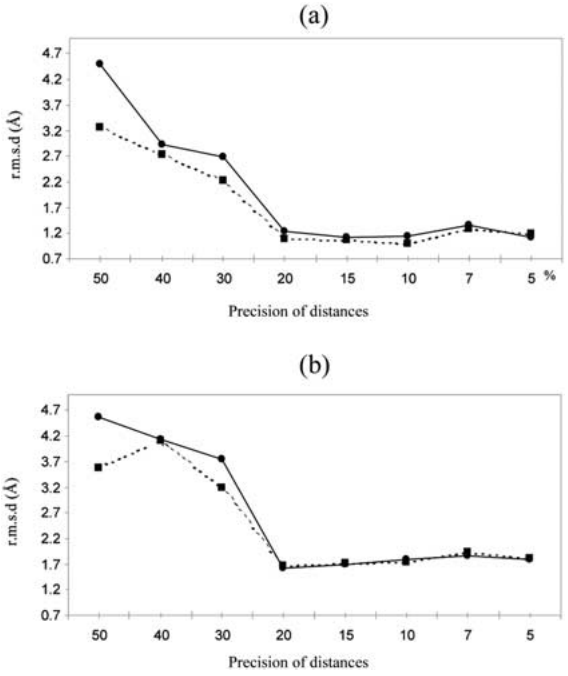

Figure 2. Curves reflecting the effects of bounds on distances on: (a) The precision and (b) the accuracy of the 17 mer duplex structure. The results are presented for the 15 best structures (—●—) and the 10 best structures (- -■- -).

of constraints. Obviously, distance precision can be improved from relaxation matrix calculations, as these allow to discard the spin diffusion effects (Borgias and James, 1990; James, 1991; Kaluarachchi et al., 1991; Boelens et al., 1989). Matrix relaxation calculations can also be used in conjunction with 3D-NMR, that helps to resolve the signal overlap problem (Zhang et al., 1995; Donne et al., 1995; Thiviyanathan et al., 1999). Rmsds were calculated versus an average structure to get the precision, and versus the target structure to get the accuracy. The local precision was also assessed calculating rmsds upon 6 base-pair segments. The derived helical parameters can be found below, in the section dealing with RDCs.

Impact on the precision and the accuracy can be seen in Figures 2a and 2b, respectively. Calculations involved 20 starting structures with randomized torsion angles, and only the 15 and 10 best-energy structures were retained for estimations. One observes an effect for bounds on distance constraints between ± 50% and ± 20%, but not from 20% to 5% in the part better corresponding to real experimental values, and where the rmsd of accuracy for the 10 best structures shows a mean value of 1.76 Å. At the same time the rmsd for precision is found equal to 1.12 Å, which

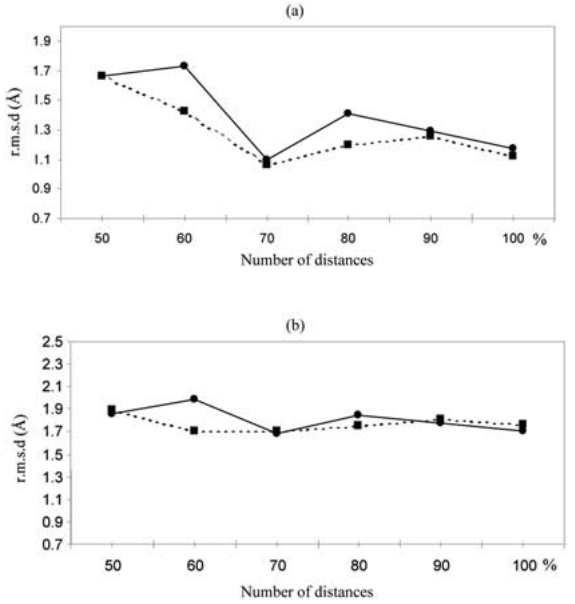

Figure 3. Curves reflecting the impact of the distance number (with bounds of 15%) on: (a) Precision and (b) accuracy of the 17 mer duplex structure. The r.m.s.ds are presented for the 15 best structures (—●—) and the 10 best structures (- -■- -).

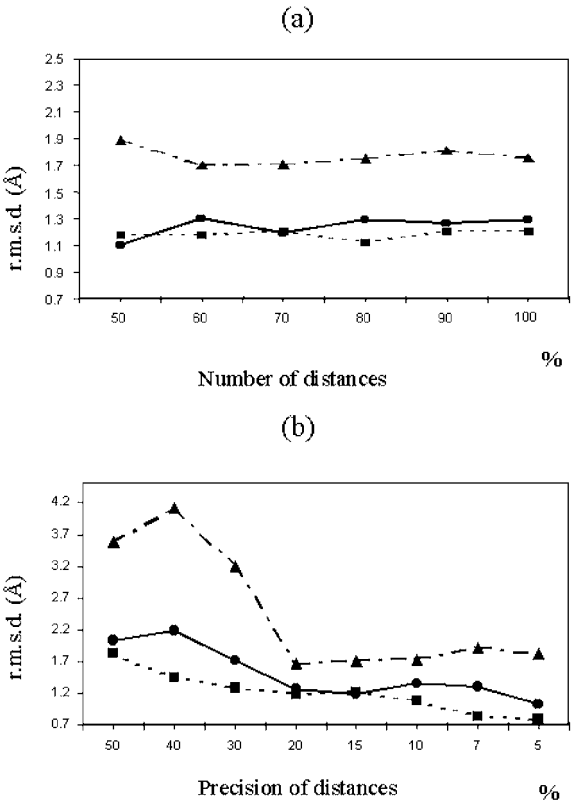

Figure 4. Effect of RDC restraints on the structure determination: (a) rmsds as a function of the the number of distance restraints (%), and (b) rmsds as a function of the precision of distances (%). The r.m.s.ds correspond to the aveage of the 10 best structures relatively to the target structure. Curves are obtained with: nOes for distances < 5 Å (-▪-▲-▪-); set A of RDCs combined to nOes (—●—); and, set B of RDCs combined to nOes (- -■- -) (see also text for explanations).

could a priori reflect a good structure definition. Yet, the difference of 0.6 Å between the two rmsds reveals a misleading feature of precision, which is further confirmed by the rather poor definition of the generated structures (Figure 1). Results registering the impact of the restraint number are presented in Figure 3. The number of nOe constraints weakly affects accuracy, while a small gain of precision is observed upon variation of constraints from 60% to 70%. The number of correct structures, on the twenty calculated for each point, grows in parallel. For the local precision based on a six base-pair segment, one finds results similar to those provided by the matrix relaxation refinements in other studies.

All together, global structures of long oligonucleotides generated from distance and torsion angle restraints are poorly defined. Precision is a misleading factor and has to be used with caution: good values of rmsd, reflecting low dispersion relative to the average structure, are artificial. Definition of global structures cannot be improved beyond 1.7–1.8 Å by increasing the distance precision on non-exchangeable protons or the restraint number. In their previous studies Allain and Varani (1997) have shown that the number of constraints affects more significantly the structure determination than the precision on nOe distance constraints, while the present work indicates that these two factors are equally inefficient for obtention of an accurate structure. This difference could be explained by the rather globular nature of the molecule studied by Allain and Varani, that is a hammerhead ribozyme, compared with the linear nature of our DNA fragment. Note that in their paper, the latter authors further mention that long-range features cannot be determined reliably by nOe-distances, which is also proved in our study.

Yet, better-defined global structures (< 1.4 Å) can be obtained when a good distance precision ($\pm$ 7%) is imposed on both non-exchangeable and exchangeable protons (not shown). Actually, the exchange properties of latter protons prevent the obtention of such a high precision.

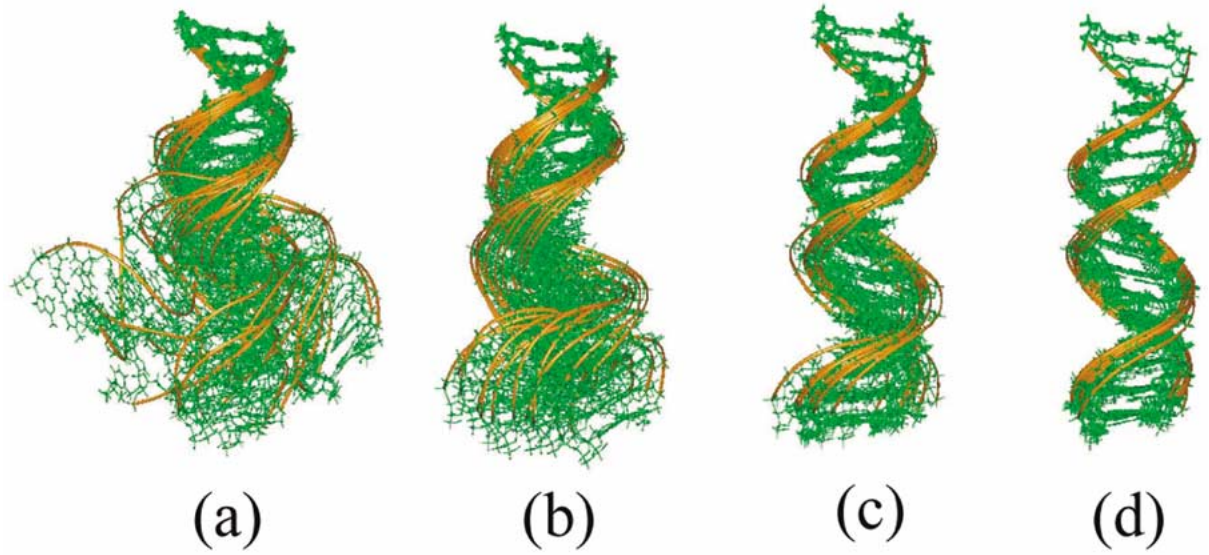

Figure 5. Effect of the different sets of restraints on the accuracy of the 17-mer structure. The 10 best structures and the target structure are shown for each set of restraints. The three upper base-pairs serve as basis for the superposition of the molecules. Pictures corrspond to: (a) Calculations with the distance set d < 4.5 Å (70% of the possible distances and a precision of 15% on each distance); (b) the same as in (a) but with distances set < 5 Å; (c) the same as in (b) with inclusion of the set A of RDCs; and the same as in (b) with inclusion of set B of RDCs.

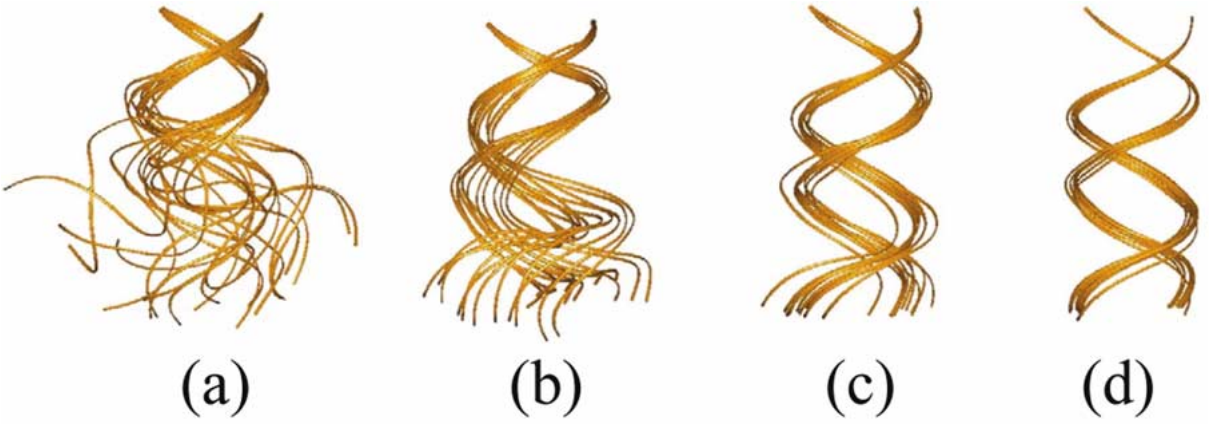

Figure 6. Effect of the different sets of restraints on the accuracy of the 17-mer structure. The same as Figure 5 but the molecules are presented with only their backbone to underline the molecular definition.

Calculations including RDC constraints

Collect of RDCs

Use of only traditional NMR data provides global structures which often lack good definition, while addition of RDC constraints during the refinement procedure can significantly improve the results (Tjandra et al., 2000; Vermeulen et al., 2000; MacDonald et al., 2001). For our 17-mer duplex which is neither palindromic nor isotopically labeled, RDC measurements were restricted to the only weakly crowded C1'-H1' and aromatic regions of ^{1}H-^{13}C spectra. In these two regions, virtually all peaks could be assigned thanks to the use of TROSY-based experiments (Brutscher et al., 1998). Assignment in the other regions of interest, i.e., C3′-H3′, C4′-H4′, C2′-H2′/H2″ require isotope labeling, sometime selective, even with small and self-complementary oligonucleotides (Tjandra et al., 2000; Mc Donald et al., 2001).

Impact of the RDC number and precision on the global structure determination

Impact of RDC constraints on the global structure determination of our 17-mer DNA was tested using two

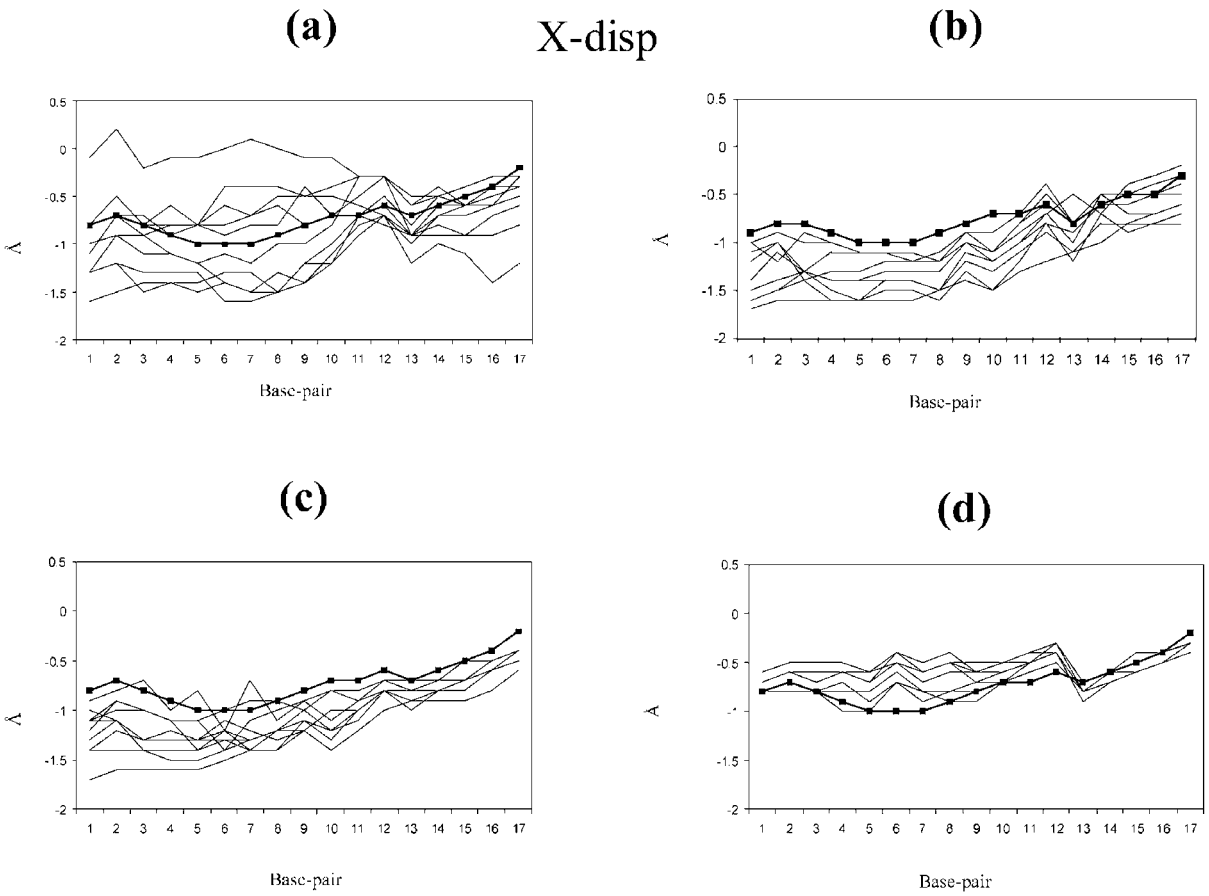

Figure 7. The x-disp helical parameter for each base-pair (noted from 1 to 17) of the 10 best molecules. Conditions correspond to calculations with: (a) The distances set d < 4.5 Å (80% of the possible distances and a precision of 15 % on each distance); (b) the same calculations as in (a) but with the distance set < 5 Å; (c) the same as in (b) with inclusion of the set A of RDCs; and (d) the same as in (b) with inclusion of the set B of RDCs. The curve corresponding to the target structure is presented as a thick dotted line.

sets of RDC values: A set A, gathering these values easily measurable from unlabeled DNA; and a set B, composed of values accessible from ^{13}C labeled and selectively deuterated compounds. First, we examined if the derivation of the Da and R values of the axial and rhombic components of the alignment tensor could be achieved by a grid search performed as described in Materials and Methods. Obtention of these values with a clear cut is a prerequisite (Warren and Moore, 2001). It appears that the reduced set of dipolar couplings (set A) is sufficient to obtain the correct values of Da and R , the selection being made from lower energy structures. There is a clear cut in the energies, and the accuracy of obtention of these values is quasi-similar to that obtained with the extended set B. This validates the strategy using a small set of dipolar couplings, since this set is sufficient to provide the parameters characterizing the alignment tensor.

Data were first compared to those obtained using distances < 5 Å (Figure 4a and Table 2). With distances < 5 Å, whatever the set, A or B, the global DNA structure was improved: the mean rmsd drops from 1.77 Å to 1.24 Å with set A, and to 1.18 Å with set B. Improvement is already visible when only 50% of the distance constraints are used, which is very encouraging for experimental studies.

Effect of combining RDCs with distances < 5 Å determined with precision of 5 to 50% is shown in Figure 4b. For distance precision chosen better than 30%, structure accuracy was found similar to that shown in Figure 4a. Remarkably, bounds of ± 20% on distances were sufficient to provide very good definition.

When RDCs are used in conjunction with distances < 4.5 Å (not shown), inclusion of set A did not improve resolution, even with 100% of the distances. On another hand, the use of set B led to high accuracy.

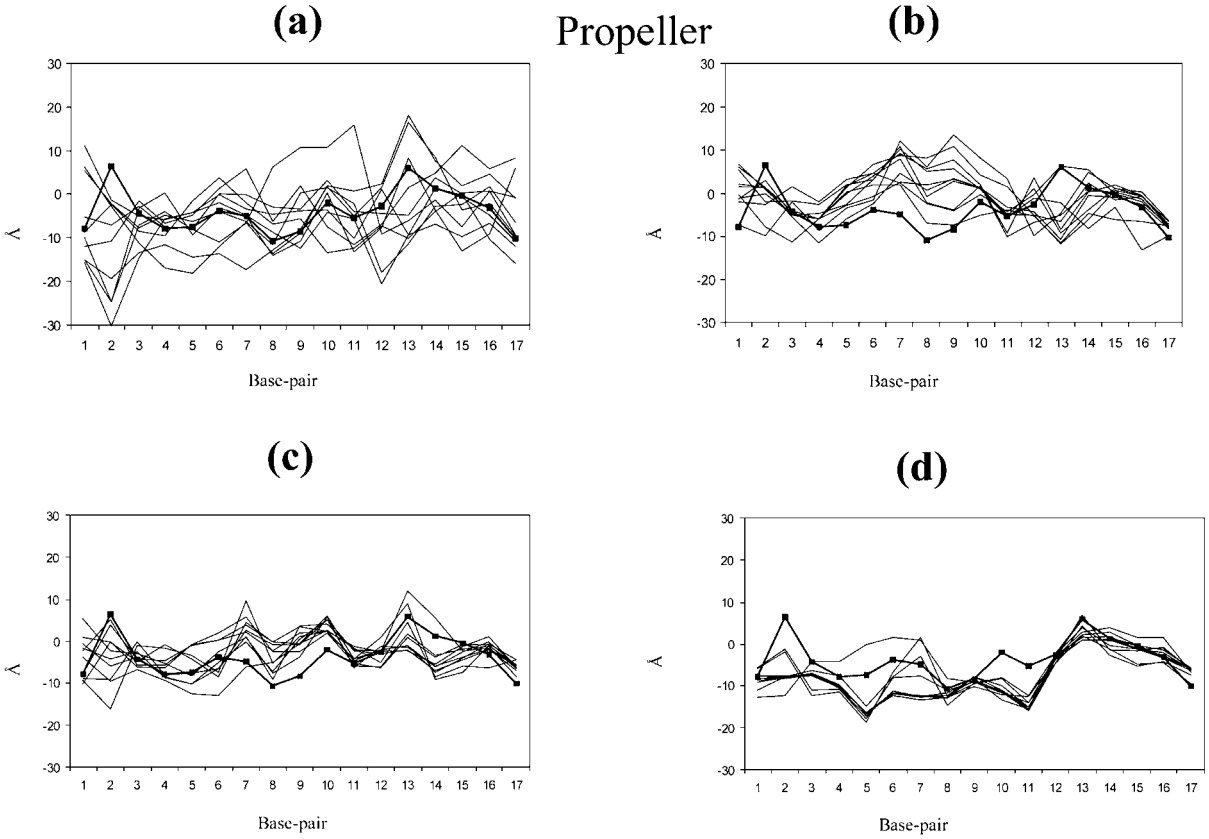

Figure 8. The propeller twist parameter analyzed as for as the x-disp in Figure 7.

Table 2 gathers the molecule energies obtained in different conditions. The refinement using RDCs in the presented protocol did not entail any increase in the holonomic (impropers, angles and bonds) energy terms, proving in turn that the agreement with RDCs is not related to a poor local geometry.

An error of ±1 Hz on the RDC coupling values seems reasonable, considering the type of experiments generally used in such studies. Note that Sibille et al. (2001) give a RDC uncertainty of ± 0.6 Hz with their TROSY experiments. Errors in the RDC constraints could also originate from local dynamics.

Actually, we performed calculations with different errors, ± 1 Hz, ± 1.5 Hz, ± 2 Hz, and we found the corresponding global rmsds: 1.24 Å, 1.39 Å and 1.49 Å, respectively. A similar evolution is observed with the set B of RDCs. Thus, the method seems rather robust.

The ten best structures provided by calculations involving both RDCs and nOes, are presented in Figures 5 and 6. These visualize the successive improvements produced by: (i) Replacement of distances <5 Å by distances < 5 Å (compare Figures 5a, 6a with Figures 5b, 6b); (ii) combination of distances < 5 Å with set A (compare Figures 5b, 6b with Figures 5c, 6c); and (iii) combination of distances < 5 Å with set B (compare Figures 5c, 6c with Figures 5d, 6d). The structure improvement produced by inclusion of set A in most calculations is highly significant.

Impact of RDCs on local structures

We also examined the effects of RDCs on helical parameters. Results relative to Xdisp, propeller twist, twist and roll are given in Figures 7–10. They are deduced from the ten best structures provided by both distance sets combined to RDC sets. Comparison of structures determined with distances d < 4.5 Å and d < 5 Å (Figures 7a–10a, respectively), shows that the longer distances with an appropriate treatment of distance (i.e., ± 15%), improves resolution. One can note that extreme values adopted by the twist and roll are accessible to distance data, in particular with those d < 5 Å. The gain of resolution produced by RDCs

Table 2. Energies (kcal mol^{-1}) of the structures refined in different conditions

	nOes only[a]	nOes and setA[b]	nOes and set B[c]
Total	−1044	−1014	−999
Bonds	11	10	10
Angles	47	46	50
Impropers	10	10	10
wdW	−497	−517	−513
nOe	0.7	1.2	1.6
Dipolar couplings		4.3	9.9

[a]The conditions of the calculations are 70% of distances with a precision of 15% on distances.
[b]Same as a but with set A as dipolar coupling restraints.
[c]Same as a but with set B as dipolar couplings restraints.

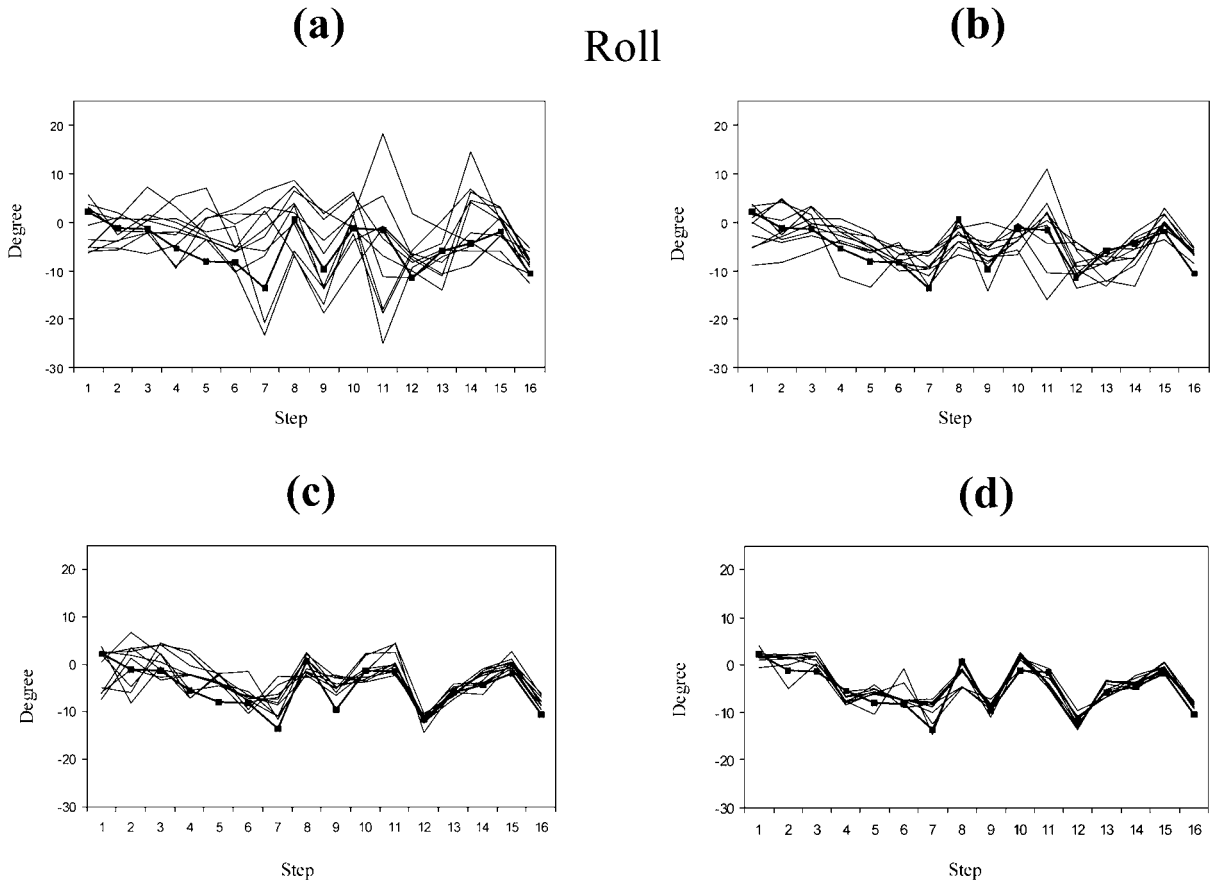

Figure 9. The roll parameter. Steps are numbered from 1 to 16. Conditions of calculations are the same as in Figure 7.

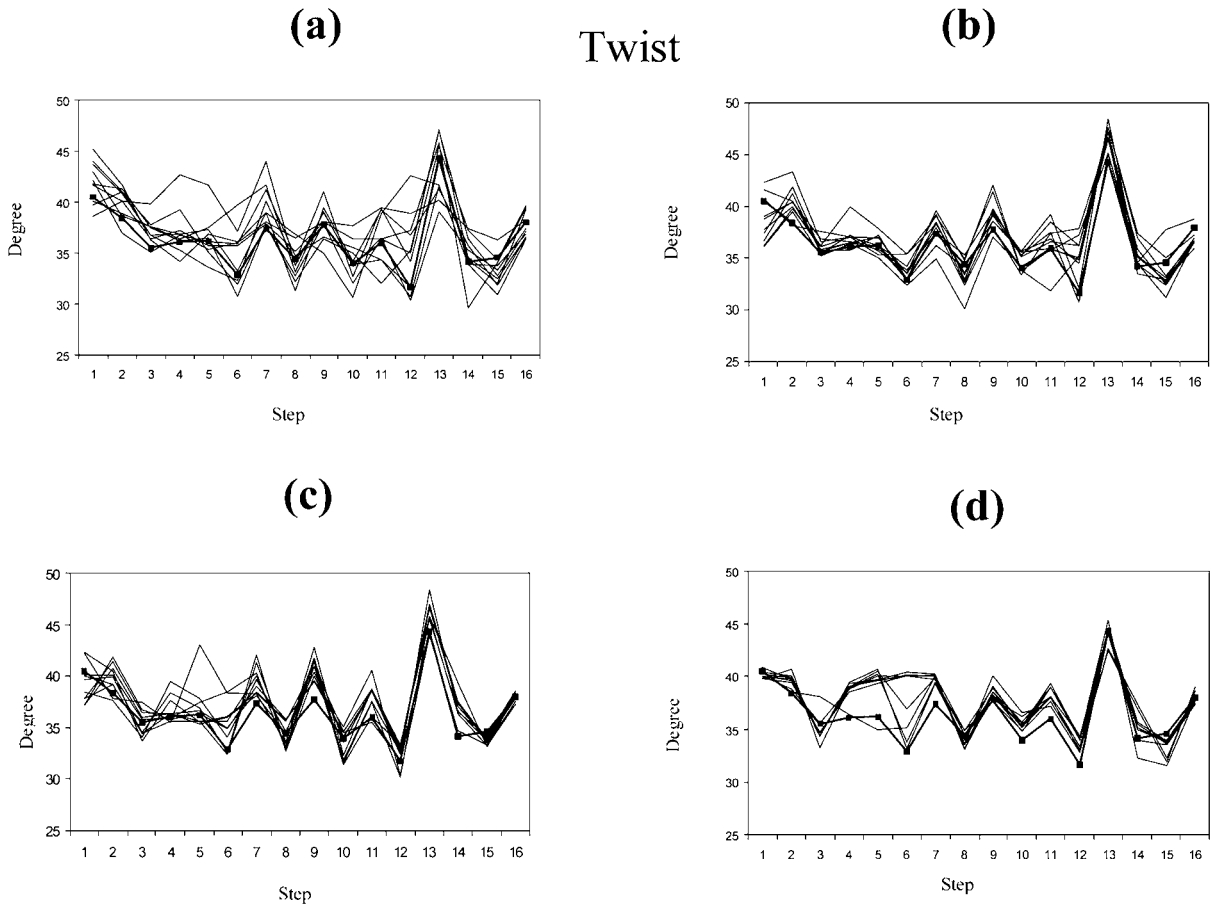

Figure 10. The twist parameter. Steps are numbered from 1 to 16. Conditions of the calculations are the same as in Figure 7.

can be appreciated using Figures 7c–10c. For instance, confrontation of Figures 7b and 8b to Figures 7c and 8c reveals the positive effect of set A on the resolution of xdisp and of propeller twist, repectively. Confrontation of Figures 7d and 8d to Figures 7c and 8c shows that resolution of the above parameters, is even more significantly improved by the larger set B.

The above observations prove that both global precision and local precision can be improved with the use of only a small number of RDC constraints (set A). For local parameters, the gain in resolution produced by RDC constraints appeared better for the xdisp and propeller twist compared with roll and twist. In fact, the roll and twist, which are linked parameters, were already well defined by the 'distance only' sets of constraints, which was not the case for the X-disp and propeller twist. Another point was the low precision observed on rolls within the AAA/TTT tract (steps 3 to 5). This is explained by the fact that thymines in contrast to cytosines, lack base H5 protons from which numerous nOe distances and RDCs useful for structure determination could be measured.

Concluding remarks

In this work we evaluated the impact of various sets of RDC constraints on the global and local structures of a large size nucleic acid fragment. Structure accuracy was analyzed using computational protocols, namely torsion-angle molecular dynamics, which are protocols similar to those applied to proteins. The approach avoiding simulations initiated with standard A- or B-DNA geometries provided an extremely high level of convergence.

That a small set of RDCs measurable in non-labeled DNA spectra, is able to improve significantly the structure resolution, is rather exciting. This set, when added to distances < 5 Å strongly improves the accuracy as illustrated by the rmsd which is reduced

from ∼1.8 Å to ∼1.2 Å. The synergistic effect of parameter combination is already observed with 50% of distance restraints and nOe-distance bounds of ± 15%.

Actually, the use of distances < 5 Å which are within range of nOe measurements, dramatically improves the accuracy of simulated structures. In that case, additional improvement can be obtained only when the larger set of RDC constraints is incorporated in the calculations. However, obtention of this set is very expensive and time consuming as it requires uniform ^{13}C labeling and selective deuteration, making this strategy not accessible to every laboratory.

Globally, our results confirm the very strong impact of RDCs on the structure accuracy and precision of long DNA segments. From the foregoing, one can conclude that a good quality of structure determination is possible using a small set of NMR data measurable on DNA under natural abundance conditions. Also, our results provide guidelines for the efficient use of RDC restraints, together with classical NMR restraints (nOes and J couplings), to improve the accuracy of DNA structures. We found that a relatively small set of RDC constraints (ca. 50), which can be detected in spectra of samples at the natural isotopic abundance, significantly increases the accuracy of global and local structures of DNA helices, when it is used in conjunction with nOe-distances up to 5 Å. This type of combination could be particularly useful for determining the degree of DNA deformation by proteins or peptides. The biologically important DNA bending either intrinsic or induced by proteins could be certainly detected using this combination.

References

Allain, F.-H.T. and Varani, G. (1997) *J. Mol. Biol.*, **267**, 338–351.

Bax, A., Kontaxis, G. and Tjandra, N. (2001) *Meth. Enzymol.*, **339**, 127–174.

Bayer, P., Varani, L. and Varani, G. (1999) *J. Biomol. NMR*, **14**, 149–155.

Boelens, R., Koning, T.M.G., van der Marel, G.A., van Boom, J.H. and Kaptein, R. (1989) *J. Magn. Reson.*, **82**, 290–308.

Borgias, B.A. and James, T.L. (1990) *J. Magn. Reson.*, **87**, 475–487.

Brünger, A.T. et al. (1998) *Acta Cryst.*, **D54**, 905–921.

Brutscher, B., Boisbouvier, J., Pardi, A., Marion, D. and Simorre, J.-P. (1998) *J. Am. Chem. Soc.*, **120**, 11485–11851.

Clore, G.M. and Gronenborn, A.M. (1998) *Proc. Natl. Acad. Sci. USA.*, **95**, 5891–5898.

Clore, G.M., Gronenborn, A.M. and Bax, A. (1998a) *J. Magn. Reson.*, **133**, 216–221.

Clore, G.M., Gronenborn, A.M. and Tjandra, N. (1998b) *J. Magn. Reson.*, **131**, 159–162.

Donne, D.G., Gozansky, E.K. and Gorenstein, D.G. (1995). *J. Magn. Reson.*, **B106**, 156–163.

Hansen, M.R., Hanson, P. and Pardi, A. (2000). *Meth. Enzymol.*, **317**, 220–240.

Kaluarachchi, K., Meadows, R.P. and Gorenstein, D.G. (1991) *Biochemistry*, **30**, 8785–8797.

Kuszewski, J., Schwieters, C. and Clore, G.M. (2001) *J. Am. Chem. Soc.*, **123**, 3903–3918.

Lavery, R. and Sklenar, H. (1988) *J. Biomol. Struct. Dyn.*, **6**, 63–91.

MacDonald, D. and Lu, P. (2002a) *Curr. Opin. Struct. Biol.*, **12**, 337–343.

MacDonald, D. and Lu, P. (2002b) *J. Am. Chem. Soc.*, **124**, 9722.

MacDonald, D., Herber, K., Zhang, X., Polgruto, T. and Lu, P. (2001) *J. Mol. Biol.*, **306**, 1081–1098.

Mollova, E.T., Hansen, M.R. and Pardi, A. (2000) *J. Am. Chem. Soc.*, **122**, 11561–11562.

Rice, L.M. and Brunger, A.T. (1994) *Proteins*, **19**, 277–290.

Rife, J.P., Stallings, S.C., Corell, C.C., Dallas, A., Steitz, T.A. and Moore, P.B. (1999) *Biophys. J.*, **76**, 65–75.

Sibille, N., Pardi, A., Simorre, J.P. and Blackledge, M. (2001) *J. Am. Chem. Soc.*, **123**, 12135–12146.

Stein, E.G., Rice, L. and Brunger, A.T. (1997) *J. Magn. Reson.*, **124**, 154–164.

Thiviyanathan, V., Luxon, B.A., Leontis, N.B., Illangasekare, N., Donne, D.G. and Gorenstein, D.G. (1999) *J. Biomol. NMR*, **14**, 209–221.

Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.

Tjandra, N., Tate, S-I., Ono, A., Kainosho, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 6190–6200.

Tolman, J.R., Al-Hashimi, H., Kay, L.E. and Prestegard, J.H. (2001) *J. Am. Chem. Soc.*, **123**, 1416–424.

Trantirek, L., Urbasek, M., Stefl, R., Feigon, J. and Sklenar, V. (2000) *J. Am. Chem. Soc.*, **122**, 10454–10455.

Vermeulen, A., Zhou, H. and Pardi, A. (2000) *J. Am. Chem. Soc.*, **122**, 9638–9647.

Warren, J.J. and Moore, P.B. (2001) *J. Biomol. NMR*, **20**, 311–323.

Wijmenga, S. and Van Buuren, B.N.M. (1998) *Progr. NMR Spectrosc.*, **32**, 287–387.

Wu, Z., Tjandra, N. and Bax, A. (2001a) *J. Biomol. NMR*, **19**, 367–370.

Wu, Z., Tjandra, N. and Bax, A. (2001b) *J. Am. Chem. Soc.*, **123**, 3617–3618.

Zhang, Q., Chen, J., Gozansky, E.K., Zhu, F., Jackson, P.L. and Gorenstein, D.G. (1995) *J. Magn. Reson.*, **B106**, 164–169.

Zimmer, D.P. and Crothers, D.M. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 3091–3095.