Structure and Dynamics of a DNA·RNA Hybrid Duplex with a Chiral Phosphorothioate Moiety: NMR and Molecular Dynamics with Conventional and Time-Averaged Restraints†

Carlos González,‡,§ Wojciech Stec,|| Mark A. Reynolds,⊥ and Thomas L. James*,‡

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, Department of Bioorganic Chemistry, Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Lódz, Poland, and Genta Inc., 3550 General Atomics Court, San Diego, California 92121

Received December 8, 1994; Revised Manuscript Received February 6, 1995®

ABSTRACT: The three-dimensional structure of two thiophosphate-modified DNA·RNA hybrid duplexes d(GCTATAA$_{ps}$TGG)·r(CCAUUAUAGC), one with *R*-thiophosphate chirality and one with *S*-thiophosphate chirality, have been determined by restrained molecular dynamics simulations (rMD). As the two yielded almost identical results, a description of results can be presented in the singular. The conformational flexibility of this hybrid has been investigated by employing time-averaged constraints during the molecular dynamics simulations (MD-tar). A set of structural restraints, comprising 322 precise interproton distance constraints obtained by a complete relaxation matrix analysis of the 2D NOE intensities as well as *J* coupling constants obtained from quantitative simulations of DQF-COSY cross-peaks in deoxyriboses, was reported in our previous paper [González, C., Stec, W., Kobylanska, A., Hogrefe, R. I., Reynolds, M., & James, T. L. (1994) *Biochemistry 33*, 11062−11072]. Multiple conformations of the deoxyribose moieties were evident from the scalar coupling constant analysis. Accurate distance constraints, obtained from complete relaxation matrix analysis, yielded a time-averaged solution structure via conventional restrained molecular dynamics which is not compatible with the experimental *J* coupling constants (root-mean-square deviation in *J* value ~2 Hz). However, vicinal coupling constant information can be reproduced when time-averaged constraints are used during the molecular dynamics calculations instead of the conventional restraints (J_{rms} ~0.6 Hz). MD-tar simulations also improve the NMR *R* factors. This improvement is more evident in the DNA than in the RNA strand, where no indication of conformational flexibility had been obtained. Analysis of the MD-tar trajectories confirms that deoxyriboses undergo pucker transitions between the S and N domain, with the major conformer in the S domain. The ribose moieties in the RNA strand, however, remain in the N domain during the entire simulation. Conformations of deoxyriboses in the intermediate domain near O4′-endo are obtained when the average structure is calculated with conventional NMR restraints. Since these conformations cannot account for the experimental *J* coupling information, and they only appear in a very low population in the MD-tar ensemble, we conclude that intermediate E sugar puckers are artifacts produced by the attempt to fit all the structural constraints simultaneously when in reality more than one conformer is present. Most structural features of the duplex remain the same in the average structure and in the MD-tar ensemble, e.g., the minor groove width, exhibiting an intermediate value compared with those of canonical A- and B-like structures. Calculations with the two duplexes of differing chirality in the modified phosphate have been carried out. Only minor effects in the backbone close to the thiophosphate have been detected.

Over the last few years, the structure of DNA·RNA hybrid duplexes has attracted considerable attention because of their importance in many biological processes such as transcription, DNA replication, or reverse transcription from RNA into DNA during retroviral replication. Especially interesting is the interaction between hybrid duplexes and ribonuclease H. This enzyme cleaves the RNA strand in DNA·RNA duplexes, but it is inactive toward double-stranded RNA. Ribonuclease H is also part of the reverse transcriptase and is involved in the elimination of the viral RNA from the DNA·RNA duplex during the reverse transcription process (Varmus, 1988). Several structural studies of DNA·RNA hybrids have been carried out in the solid state (Arnott et al., 1986; Egli et al., 1992) and in solution (Salazar et al., 1993a; Hall, 1993). Very recently, detailed solution structures of two duplexes have been obtained by NMR (Fedoroff et al., 1993; Lane et al., 1993). It is now well established that the sugar conformations in the RNA and DNA strands are different in solution. Riboses adopt N-type (near C3′-endo with pseudorotation phase angle ~0°) conformations typical for A-form structures, whereas deoxyriboses in the

† This work was supported primarily by NIH Grant No. GM39247, with use of the UCSF Computer Graphics Laboratory supported by NIH Grant RR 01081, and computations at the Pittsburgh Supercomputing Center through the NIH National Center for Research Resources Cooperative Agreement U41RR04154 and a grant from the National Science Foundation Cooperative Agreement ASC-8500650.

* Author to whom correspondence should be addressed. Telephone: (415) 476-1569. Fax: (415) 476-0688.

‡ University of California, San Francisco.

§ Present address: Instituto de Estructura de la Materia (CSIC). C/. Serrano 119, 29006 Madrid, Spain.

|| Polish Academy of Sciences.

⊥ Genta, Inc.

® Abstract published in *Advance ACS Abstracts,* April 1, 1995.

0006-2960/95/0434-4969$09.00/0

DNA strand are closer to the S domain (near C2′-endo with pseudorotation phase angle ~160°) (Chou et al., 1989; Katahira et al., 1990; Salazar et al., 1993a,b; Lane et al., 1993). However, some controversy has arisen about the structure of the sugars in the DNA strands of hybrids and chimeric duplexes. Salazar et al. (1993a,b) conclude that deoxyriboses assume an unusual conformation, approximately O4′-endo, with pseudorotation phase angle around 90°. On the other hand, Lane et al. (1993) found that their NMR data could be explained better by assuming a two-state approximation with a major population of sugars in S-type conformations. In our preliminary work (González et al., 1994), strong evidence of multiple conformations in the DNA strand of our hybrid duplexes was presented.

In addition to the attention paid to hybrid duplexes, modified oligonucleotides have also attracted great interest due to their potential use as chemotherapeutic agents. Phosphorothioates are especially interesting because of their enhanced nuclease resistance and ability to form hybrid duplexes with the target RNA which can be attacked by ribonuclease H, improving the inhibitory activity of the antisense molecule (Zon, 1988; Uhlmann & Peyman, 1990). In spite of their considerable interest, not much is known about the structural effects of these kind of modifications. The few structural studies in modified oligonucleotides carried out so far focus on DNA·DNA duplexes (Heinemann et al., 1991; Gao et al., 1992; Stolarski et al., 1992). In most cases, the modification appears to fit well in the general B-type structure of the duplex, but in the case of phosphorodithioates significant distortions from the B-family of structures have been detected (Cho et al., 1993). Although mRNA is the most attractive target for the antisense chemotherapeutical strategy, virtually nothing is known about the structure of oligonucleotide analogs in DNA·RNA duplexes.

The aim of this work is to determine the structure of the fragment d(GCTATAA$_s$TGG)·r(CCAUUAUAGC) in solution by NMR. The central part of this sequence is the Pribnow box, whose double-stranded DNA form has been studied previously in our laboratory (Schmitz et al., 1992a). The phosphate between A7 and T8 is modified by substituting sulfur for one oxygen atom, leading to two different stereoisomers (*R* and *S*). Two samples with pure *R* or pure *S* stereochemistry in the modified phosphate have been studied.

In our first paper, a careful extraction of structural constraints was described. More than 300 distance restraints were obtained from 2D NOE[1] data acquired in H_2O and D_2O with the program MARDIGRAS, which makes use of the complete relaxation matrix approach (Borgias & James, 1988, 1990). In addition to interproton distances, precise vicinal proton−proton coupling constants were obtained by comparing experimental DQF-COSY cross-peaks with quantitative simulations carried out with the programs SPHINX and LINSHA (Widmer & Wüthrich, 1986, 1987). From these structural constraints the heteronomous structure of the hybrid could be confirmed. Riboses in the RNA strand were found in a N-type conformation typical of the A-form family, but deoxyriboses in the DNA strand adopted a different conformation with sugar puckers partially in the S-type domain. Most importantly, all coupling constants in the deoxyriboses were not simultaneously consistent with any single sugar conformation. A two-state dynamic equilibrium between N- and *S*-type conformers had to be assumed to fit the experimental *J* coupling constants. In general, the population of major *S* conformer was lower than in double-stranded DNA duplexes, indicating that hybrid duplexes might be more flexible than pure DNA or RNA.

This enhanced flexibility causes an additional difficulty for determination of the structure of the hybrid duplex. Structure determination of nucleic acids in solution relies on extracting accurate interproton distance constraints from NOE intensities and torsion angle constraints from homonuclear proton−proton coupling constants. These two types of structural constraints are used subsequently along with an assessment of the experimental errors to calculate the three-dimensional structure of the molecule by restrained molecular dynamics or Monte Carlo simulations (Ulyanov et al., 1993). Both methods search for one single structure which satisfies all the constraints simultaneously, but, in cases where multiple conformations are present, the experimental quantities reflect an average over all the exchanging conformations. In these cases, the structural constraints extracted from the experiment may not be compatible with any single rigid structure. Furthermore, due to the different kind of averaging of NOE intensities and proton−proton coupling constants, both types of constraints may represent contradictory information. The attempt to fit all this information simultaneously may lead to artifacts that do not represent real structural features. Although these flexibility effects should be considered in all structure determination, the analysis of *J* coupling constants carried out in our previous paper (González et al., 1994) indicates that dynamic effects are particularly important for DNA·RNA hybrids.

Several approaches to the problem of multiple conformations in NMR structure determination have been developed (Torda et al., 1989; Brüschweiler et al., 1991; Ulyanov et al., 1995). Among them, we have employed in this study a molecular dynamics approach which uses time-averaged restraints as suggested by Torda et al., (1989, 1990). This method has been implemented in the AMBER program suite (Pearlman et al., 1991). In this approach, the penalty term for the experimental restraints does not depend on the instantaneous constraint violation but on the average violation during some period of time. An exponential weighting factor is added to emphasize the most recent snapshots in the MD trajectory. In this way, the constraints do not need to be satisfied in each single snapshot but the values along the trajectory as a whole yield agreement with the experimental data. This approach was first used with NOE derived distance constraints (Torda et al, 1990; Pearlman & Kollman, 1991) and more recently with vicinal proton−proton *J* coupling restraints (Torda et al, 1993; Mierke et al, 1994; Pearlman, 1994b). Direct NOE refinement against NOE intensities using a time-average strategy has been also explored (Bonvin et al., 1994). Most of the applications have

[1] Abbreviations: 2D NOE, two-dimensional nuclear Overhauser effect; DQF-COSY, double-quantum filtered COSY; E_{NOE}, NOE distance constraint violation energy; E_J, *J* coupling violation energy; J_{rms}, average root mean square deviation for coupling constants; MD-tar, time-average molecular dynamics; rMD, restrained molecular dynamics; RMSD, root mean square deviation; *R*-duplex, hybrid duplex d(GCTATAA$_{ps}$TGG)·r(CCAUUAUAGC) where the modified phosphate has an *R*-type chirality; *S*-duplex, hybrid duplex where the modified phosphate has an *S*-type chirality.

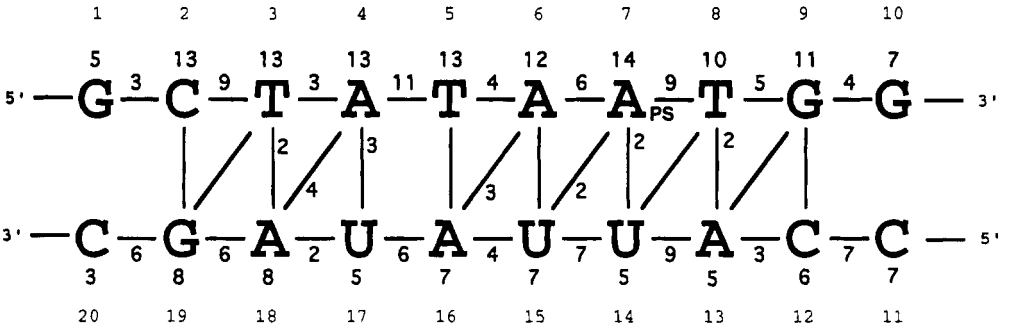

FIGURE 1: Distribution of NOE-derived distance restraints for [d(GCTATAA$_{ps}$TGG)·r(CCAUUAUAGC)]. Bold numbers indicate number of intraresidue, sequential, and interstrand constraints with placement implying attribution. Plain numbers indicate the sequence order.

focused on cyclic peptides and small proteins, but studies in nucleic acids have been also carried out (Schmitz et al., 1992b, 1993; Schmitz & James, 1994).

MATERIALS AND METHODS

Generation of NMR Restraints. The extraction of J coupling constants as well as distance restraints has been described in our previous paper (González et al., 1994). A total of 322 accurate distance constraints have been obtained from 2D NOE cross-peak intensities using a complete relaxation matrix analysis with the program MARDIGRAS (Borgias & James 1988, 1990). In addition to dipolar interaction, exchange effects were taken into account to obtain distance constraints involving labile protons (Liu et al., 1993). No internal motions were considered in MARDIGRAS calculation except for methyl groups where a three-state jump model was employed (Liu et al., 1992a).

To avoid any possible bias coming from initial conditions and to estimate the error bounds in the interprotonic distances properly, different MARDIGRAS calculations were carried out with different initial models, mixing times, and correlation times. Three initial models were used, i.e., standard A- and B-form duplexes and a heteronomous duplex (vide infra). Correlation times of 1.5, 2.0, and 2.5 ns, in an interval around the experimentally obtained τ_c, were employed. Experimental intensities were recorded at three different mixing times (100, 200, and 300 ms) for nonexchangeable protons, and at a single mixing time (200 ms) for labile protons. Final distances were obtained by averaging a total of 27 MARDIGRAS outputs (3 mixing times × 3 starting structures × 3 correlation times). MARDIGRAS calculates upper and lower bounds for each distance by performing an error propagation analysis from the estimated uncertainty in the corresponding intensity. Final distance errors were determined by averaging upper and lower bounds in all 27 individual runs. Additional error bounds in distances involving labile protons were determined by assuming an interval in exchange rates of $0-4\ s^{-1}$ for T and U imino protons and $0-10\ s^{-1}$ for imino C and G protons as well as amino protons. The total number of distance constraints was 322, including both exchangeable and nonexchangeable protons. This set was comprised of 182 intraresidue, 115 interresidue, and 25 interstrand constraints. The distribution of these constraints along the sequence is shown in Figure 1.

Distance constraints with their corresponding error bounds were incorporated into the AMBER potential energy by defining a flat-well potential term with the form:

$$E_{NOE} = 2k_{NOE}(r_2 - r_1)(r_1 - r) + k_{NOE}(r_2 - r_1)^2 \text{ if } r < r_1$$

$$= k_{NOE}(r_2 - r)^2 \text{ if } r_1 < r < r_2$$

$$= 0 \text{ if } r_2 < r < r_3$$

$$= k_{NOE}(r - r_3)^2 \text{ if } r_3 < r < r_4$$

$$= 2k_{NOE}(r_4 - r_3)(_r - r_4) + k_{NOE}(r_4 - r_3)^2 \text{ if } r_4 < r$$

where r_3 and r_2 are the calculated upper and lower distance bounds. In some cases, where the difference between r_3 and r_2 was lower than 0.4 Å, a minimum flat-well width of 0.4 Å was maintained by setting the target distance limits to the value of the average distance ±0.2 Å. The form of the penalty function is parabolic between r_3 and r_4 and between r_1 and r_2. When the distance is larger than r_4 or lower than r_2, the penalty function is linear. The width of both parabolic regions was set to 1 Å.

Proton coupling constant values for deoxyriboses were obtained by simulating DQF-COSY cross-peaks with the programs SPHINX and LINSHA (Widmer & Wüthrich, 1986) as described in González et al. (1994). In contrast to other studies where J coupling information is translated first into torsion angle constraints and then incorporated into the force field, we have used the experimental J coupling constants directly in the molecular dynamics refinement. In those calculations where coupling constants were considered, a J coupling penalty term, analogous to E_{NOE}, was included in the force field. Estimated error bounds in the experimental J coupling constants [see Table 3 in González et al., (1994)] determined the size of the flat-well in the E_J energy term. The width of the parabolic regions was set to 1 Hz. The theoretical J coupling constants were calculated by using the Karplus equation. The present implementation of the Karplus equation in the SANDER module of AMBER, employed in this study for the molecular dynamics calculations, makes use of the coefficients determined by Davies (1978). A more elaborate Karplus equation (Haasnoot et al., 1980; de Leeuw et al., 1983) has been used to calculate root mean deviations between theoretical and experimental J coupling constants during the analysis of the average structures and molecular dynamics trajectories.

In addition to these experimentally derived constraints, Watson–Crick hydrogen bond restraints were used. Target values for distances and angles related to hydrogen bonds were set as described by using crystallographic data (Saenger et al., 1984). No backbone angle constraints were employed.

Initial Model Building. Three different model structures were used for MARDIGRAS calculations and as initial models for molecular dynamics refinements. Two of these structures correspond to classical A- and B-form and the third one to a heteronomous duplex, where the RNA strand is in standard A-form and the DNA strand adopts an intermediate conformation with the sugar pseudorotation angle around 90°, in the E-type range. All the initial models were built with the program DNAminiCarlo (Ulyanov et al., 1993) which allows generation of structures by specifying generalized helical parameters. The heteronomous structure was created with glycosidic angles of −130° in the DNA strand and −150° in the RNA strand. Pseudorotation phase angles were 90° and 25°, respectively. To avoid any unfavorable atom contact, the structures were energy minimized by carrying out 1000 cycles of unrestrained minimization with the method of steepest descent. The potential energy was decreased to reasonable values without significant distortion of the initial structures.

Molecular Dynamics. All restrained energy minimizations and molecular dynamics calculations were carried out with the SANDER module of the program package AMBER Version 4 (Pearlman et al., 1991). Calculations were performed *in vacuo* as well as with explicit solvent. In both cases, an all-atom representation was used, and Na^+ counterions were added at a distance of 5 Å to the phosphorus atoms to neutralize the negative charge. For *in vacuo* simulations, the counterions were hexahydrated. The electrostatic term was calculated using a distance-dependent dielectric constant, and the cut-off value for nonbounded interactions was 10 Å. Usual force field parameters of the AMBER potential (Weiner et al., 1986) were used except for the modified phosphate, where parameters obtained from *ab initio* calculations by Jaroszewski et al. (1992) were employed. The temperature and the relative weights of the experimental constraints were varied during the simulations according to an annealing strategy explained below. The temperature was maintained at the required value by rescaling the velocities when fluctuations were bigger than 20 K. The SHAKE algorithm (Ryckaert et al., 1977) was used to keep bond lengths fixed, allowing a time step of 2 fs in the integration of the motion equations.

Two types of NMR restrained molecular dynamics simulations have been carried out in this study. The first calculations were designed to obtain the average structure that best resembles the NMR constraints. To properly assess that final results do not depend on the initial conditions of the simulations, three different starting structures (A, B, and heteronomous) and various initial velocities for each were used. Each structure was first restrained energy minimized and then submitted to 25 ps of molecular dynamics. During the first 5 ps the temperature was raised linearly from 100 to 900 K, maintained at this value for the next 5 ps, and decreased to 300 K over 3 ps. This final value was maintained for the remaining 12 ps. During the high temperature period some counterions were observed to drift away from the molecule. To prevent any distortion due to the lack of neutrality in the system, counterions were relocated at a distance of 5 Å of the phosphorus atoms after the first 15 ps. The force constants for the distance restraints was varied simultaneously with the temperature. K_{NOE} was raised from 20 kcal/(mol·Å^2) up to 100 kcal/(mol·Å^2), kept at this value during the high temperature period, and reduced to 40 kcal/(mol·Å^2). This value was maintained during the rest of the simulation and the subsequent restrained energy minimization. The last 5 ps of the trajectory were used for averaging, and the resulting structure was subjected to 1000 steps of restrained energy minimization using the steepest descent method. No constraints derived from J coupling constants were used in the calculation of the average structures.

To explore the accuracy of the different structural parameters and to gain more insight into the conformational flexibility of this hybrid duplex, longer trajectories were calculated starting from the average solution structure of the hybrid. Four runs of 120 ps of rMD and MD-tar were carried out *in vacuo* and using explicit water molecules. In all the cases, the temperature was maintained at 300 K, and K_{NOE} was set to 20 kcal/(mol·Å^2). The force constant for the J coupling penalty term was 5.0 kcal/mol, within the range of acceptable values determined by Pearlman (1994b). When explicit solvent was taken into account, the starting structure was solvated with a 6 Å shell of TIP3P water molecules (Jorgensen et al., 1983). An energy minimization and 10 ps of molecular dynamics, in which the atoms in the duplex were kept fixed and the water molecules were allowed to move, were carried out to accommodate the water shell around the solute molecule. Although explicit solvent is considered, a distance-dependent dielectric constant has been employed since its use has been shown very effective to maintain the water molecules close to the solute (Guenot & Kollman, 1992, 1993). In the MD-tar calculations, the E_{NOE} and E_J terms of the potential function are substituted by their time-averaged equivalents (Torda et al., 1990; Pearlman & Kollman, 1991; Pearlman, 1994a). Hydrogen bond distance constraints were not time-averaged in the MD-tar simulations. All calculations were carried out using a third-root averaging for distances and a damping constant of 10 ps. Most standard rMD calculations were carried out on a Hewlett-Packard 735 computer and the MD-tar calculations were performed on the Cray C90 computer at the Pittsburgh Supercomputing Center for convenience. However, for the same computer and the same number of steps in the simulation, running MD-tar compared with with standard rMD requires only a factor of about 20% more time, and using a shell of water compared with *in vacuo* calculations requires a simulation about three times as long. Our experience with a few nucleic acid sequences though indicates that, while a 25−30 ps simulation is adequate for rMD, 120 ps is required for MD-tar.

Analysis of Structures and Trajectories. Analysis of the average structures as well as the MD trajectories was carried out with the programs Curves (Lavery & Skelenar, 1988, 1989) and Dials and Windows (Ravishanker et al., 1989). Reliability factors (R factors) (Thomas et al., 1991; González et al., 1991) were used to compare experimental with theoretical NOE intensities calculated with the program CORMA by using a complete relaxation matrix analysis (Keepers & James, 1984; Borgias & James, 1988). Theoretical intensities were calculated for single structures or for complete trajectories, considered as an ensemble of rapidly interchanging structures (Liu et al., 1992b; Schmitz et al., 1992b), giving rise to the usual R factors or ensemble R factors, respectively. In all CORMA calculations, a single correlation time of 2.2 ns was used.

Comparison between experimentally determined deoxyribose coupling constants and theoretically calculated values

Table 1: Atomic rms Differences (Å) between Nine Structures Resulting from a High Temperature rMD Refinement[a]

	A1	A2	A3	B1	B2	B3	H1	H2	H3
A1	0.00	0.61	0.54	0.74	0.71	0.69	0.54	0.78	0.49
A2		0.00	0.50	0.79	0.47	0.70	0.49	0.63	0.56
A3			0.00	0.64	0.63	0.69	0.56	0.67	0.45
B1				0.00	0.72	0.70	0.74	0.89	0.72
B2					0.00	0.64	0.53	0.67	0.67
B3						0.00	0.56	0.72	0.76
H1							0.00	0.76	0.53
H2								0.00	0.68
H3									0.00

[a] See text for details. A1, A2, and A3 result from a starting structure of A-form duplex with three different initial trajectories. Likewise, those with the B prefix or H prefix arise from B- and H-form starting structures.

was carried out for all final average structures and ensembles using the J_{rms} indicator:

$$J_{rms} = \frac{1}{N}\sqrt{(J_{theo} - J_{exp})^2}$$

Theoretical J coupling constants were obtained with the modified Karplus equation utilizing the parametrization of Altona and co-workers (Hassnoot et al., 1980; de Leeuw et al., 1983). In all cases, ideal hydrogen atom positions were used (Schmitz et al., 1993).

RESULTS AND DISCUSSION

Calculation of the Average Structure. The two main sources of experimental information in NMR, NOE intensities and J coupling constants, reflect different types of time-averaging in the case of internal motions or conformational flexibility. In cases where more than one conformer is present, the incorporation of structural constraints derived from these two types of experimental data can lead to inconsistencies in the rMD refined structures (Stolarski et al., 1992; Schmitz et al., 1992b; Ulyanov et al., 1994). In our previous paper (González et al., 1994), strong evidence of multiple conformations in the deoxyriboses of the hybrid was found. Careful analysis of the J coupling constants indicate that repuckering in the DNA-strand is taking place with populations of minor conformers larger than for double stranded DNA (up to 40% in some cases). This enhanced flexibility of deoxyriboses prevented us from including torsion angle constraints directly derived from the experimentally obtained J coupling constants as usually done in high-resolution nucleic acid structure calculations (Kim et al., 1992; Mujeeb et al., 1993; Weisz et al., 1994). Therefore, the average NMR structure of this hybrid duplex was obtained only on the basis of the distance constraints derived from NOE intensities. The J coupling constants were used as an independent check of the quality of these final structures. A penalty term derived from the J coupling constants was included in subsequent calculations aimed to determine dynamic features of the duplex.

A total of nine structures for each stereoisomer were calculated following the high temperature restrained molecular dynamics protocol described under Materials and Methods. Three different sets of initial velocities for each of the three starting structures were used. Mutual rms deviations between the nine final structures are shown in Table 1 for the heavy atoms of nonterminal nucleotides. The average value of 0.65 Å indicates that good convergence has been reached. The final set of structures represents a well defined region of conformational space clearly different of the initial A, B, and heteronomous structures, as indicated by the average rms difference between the average final structure and initial ones (1.6, 2.6, and 2.0 Å for A, B, and H models). Atomic rms deviations between structures obtained from different starting velocities and identical coordinates are very similar to the values obtained from different starting coordinates and identical velocities, indicating that the specific protocol used here is sufficiently effective to overcome potential barriers and any trace of the starting structure has been removed. The nine resulting structures were averaged and submitted to a final restrained energy minimization, giving rise to the final average structure, rMD-FIN, shown in Figure 2a.

In contrast to other studies on NMR solution structures of double-stranded DNA sequences (Weisz et al., 1994; Mujeeb et al., 1993), torsion angle constraints are not necessary for obtaining good convergence from different starting structures. This may be related to the fact that the final structure is, in some way, intermediate between A- and B-form.

As shown in Table 2, final distance restraint violations are drastically improved in the final structures compared to the starting structures. The quality of the overall fit of the final structures to the experimental data can be assessed by a direct comparison between experimental and theoretically calculated NOE intensities using NMR R factors (González et al., 1991; Thomas et al., 1991). Both crystallographic

Table 2: Values of NMR R Factors, Residual Distance Constraint Violations, and J_{rms} for the Three Initial Models, Average Structures, and Different Ensembles Obtained with rMD and MD-tar in Vacuum and Using Explicit Solvent Conditions

	av violation (×100 Å)	av violation DNA	av violation RNA	R factor[a]	R factor[a] DNA	R factor[a] RNA	R^x factor[b] (×100)	R^x factor[b] DNA	R^x factor[b] RNA	J_{rms}[c]
A-form	31	41	23	0.50	0.74	0.37	10.0	11.0	9.3	4.33
B-form	48	29	89	0.87	0.89	0.86	20.0	12.0	26.0	1.60
H-form	21	23	24	0.61	1.00	0.40	11.0	13.0	10.0	2.00
av nine final structures	6.1	7.6	4.9	0.39	0.47	0.32	7.5	8.4	6.7	1.98
rMD-FIN	6.6	7.2	5.0	0.38	0.47	0.33	7.4	8.4	6.6	2.01
rMD vacuum	7.6	8.9	6.5	0.36	0.46	0.31	6.8	7.9	6.0	1.88
rMD water	7.8	9.1	6.4	0.35	0.48	0.28	6.8	8.0	5.8	1.68
MD-tar vacuum[d]	5.4	5.0	6.3	0.31	0.36	0.29	6.0	6.7	5.5	0.51
MD-tar water[d]	5.5	5.2	6.4	0.30	0.31	0.29	5.8	6.1	5.6	0.58

[a] Crystallographic-like NMR R factor. [b] Sixth-root weighted NMR R factor. [c] Average root mean square deviation for coupling constants. [d] Third-root average along a 120 ps trajectory for residual distance violations and R factors.

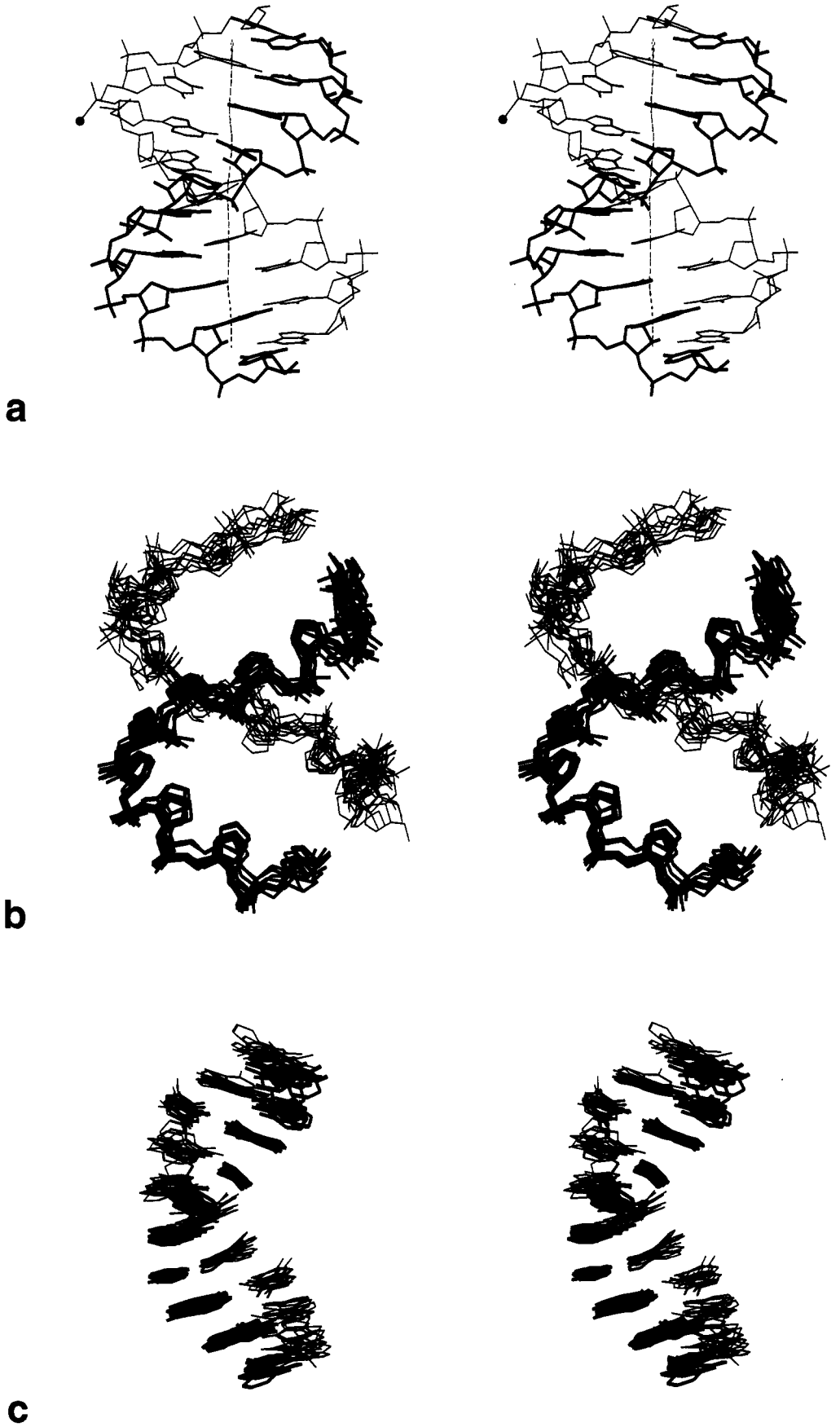

FIGURE 2: Stereoview of the solution structure of the *S*-stereoisomer of the hybrid duplex [d(GCTATAA$_{ps}$TGG)-r(CCAUUAUAGC)]. (a) Average solution structure, RMD-FIN, with helix axis indicated and black circle, indicating the sulfur atom position. (b) Best superposition of the backbone atoms for 10 snapshots of a 120 ps trajectory in water with time-averaged constraints. (c) Same as panel b but only base atoms shown. Bold lines indicate the RNA strand and plain lines the DNA strand.

and sixth-root R factors (Table 2) show a clear improvement compared with the starting structures. Especially interesting is the difference between R factors obtained for each individual strand in the final structures. The higher R factors in the DNA indicate that the fit with the experimental NOE data is better in the RNA than in the DNA strand.

Potential energy values (not including the NOE term) in the nine final structures are lower than in the initial structures, ranging from −2893 to −3290 kcal/mol, with an average value of −3100 kcal/mol. The final average and energy minimized structure presents an energy of −3282 kcal/mol. The E_{NOE} term does not show large fluctuations between structures, with an average value of 76 kcal/mol. The energies obtained indicate that all the resulting structures agree equally well with the AMBER force field and the distance constraints.

In addition to the comparison for NOE intensities, theoretical vicinal J coupling constants were calculated for each structure and compared with experimental values. J_{rms} for the initial A-form structure presents the highest value, demonstrating that the J coupling information is not in agreement with a DNA strand belonging to the A-family of structures. Contrary to the effect observed for NOE intensities, no significant improvement in the J_{rms} is observed in the final structures, which present values even higher than in the initial B-form. Hence, the structures that best fit the distance constraints do not represent the J coupling constants very well. This contradiction between distance restraints and J coupling information is consistent with the conclusion of our previous paper where, except in some nucleotides, no single sugar pucker for deoxyriboses could be found to reproduce all the J coupling constants observed (González et al., 1994).

Ensemble Calculations. The set of nine structures calculated in the previous section, whose averaged and energy-minimized coordinates yielded the final structure, RMD-FIN, shown in Figure 2a, represents the region of the conformational space consistent with all the distance constraints simultaneously. However, there is no reason to require that all the constraints must be accomplished at the same time. On the contrary, independent evidence arising from the J coupling information indicates that conformational averaging is taking place. To explore the dynamic effects in this duplex as well as to determine possible artifacts in the average structure coming from the attempt to fit simultaneously all the distance constraints, MD-tar calculations have been carried out.

Previous studies performed in our lab have demonstrated that inclusion of explicit solvent during the simulation serves to eliminate unusual high-energy conformations appearing in *in vacuo* MD-tar calculations (Schmitz et al., 1993). Thus, a long trajectory of 120 ps of MD-tar in water (MD-tar water) has been generated starting from the average solution structure rMD-FIN. To distinguish the effect of the explicit solvent from the effect of the time-averaged constraints, three additional calculations have been carried out: Similar calculations were carried out *in vacuo* (MD-tar vacuum) as well as using conventional restrained molecular dynamics, both *in vacuo* (rMD vacuum) and with explicit solvent (rMD water).

J couplings were explicitly included in all of these calculations by adding an extra energy term that penalizes the difference between the calculated and experimental J coupling. Systematic investigation of the influence of the weighting constant for the NMR terms in MD-tar calculations has been carried out by Schmitz et al. (1993) for the distance constraints and by Pearlman (1994b) for the coupling constant restraints. In both cases, exploration of the conformational space appears to be improved when lower constants are employed. According to this result, the value of K_{NOE} was decreased to 20 kcal/(mol·Å^2), lower than in the initial calculations.

General parameters indicating the quality of the different trajectories are shown in Table 2. The most remarkable feature is the enormous reduction in J_{rms} for MD-tar trajectories, indicating that experimental J coupling values can be fitted provided the assumption of one single average structure is relaxed. Although the explicit penalty term for J coupling constants is also utilized for the rMD vacuum and rMD-water calculations, J_{rms} values do not show any significant improvement compared to the values for the average structure, which was calculated without considering J coupling constraints. J coupling constants can only be satisfied by employing a time-average penalty term for the restraints.

R factors are also reduced in the MD-tar trajectories, and, most importantly, their values decrease more in the DNA than in the RNA strand. The improvement of the fit with the experimental data is enhanced for the DNA strand where more mobility was expected. Final R values in the MD-tar trajectories are similar for both strands, indicating that the higher R factors in the DNA strand of the average structure were largely due to the attempt to fit all distance constraints with one single conformer, which is apparently not entirely realistic. It must be pointed out that this effect does not appear in the RNA strand, and R factors do not decrease to the same extent.

Analogous to the R factors, average residual distance restraint violations are reduced in MD-tar compared to rMD trajectories. The improvement is bigger in the DNA than in the RNA strand. Violations presented in Table 2 correspond to distances obtained by calculating the third root-weighted average along the trajectory. The distance violations can be also calculated by carrying out an arithmetic average instead of a third root-weighted average. Both are very similar for rMD trajectories, but, for the MD-tar trajectory, the arithmetic average values increase up to 0.138 Å (vacuum) and 0.106 Å (water). Again, the difference between third root-averaged distance violations and arithmetically averaged distance violations is bigger in the DNA than in the RNA strand.

Although there is a slightly better agreement of the experimental data with *in vacuo* than with explicit water simulations, the small extent of the difference can not be considered significant. On the other hand, atomic rms deviations along the trajectories are smaller in water than in *in vacuo* conditions (1.1 and 1.6 Å, respectively). It has been shown in other studies (Schmitz et al., 1993; Pearlman, 1994a) that MD-tar can induce very fast conformational transitions with frequent visits to high energy states. As noticed before (Schmitz et al., 1993), these transitions occur less frequently when explicit solvent is considered, due to the damping effect of the water. Since the AMBER potential

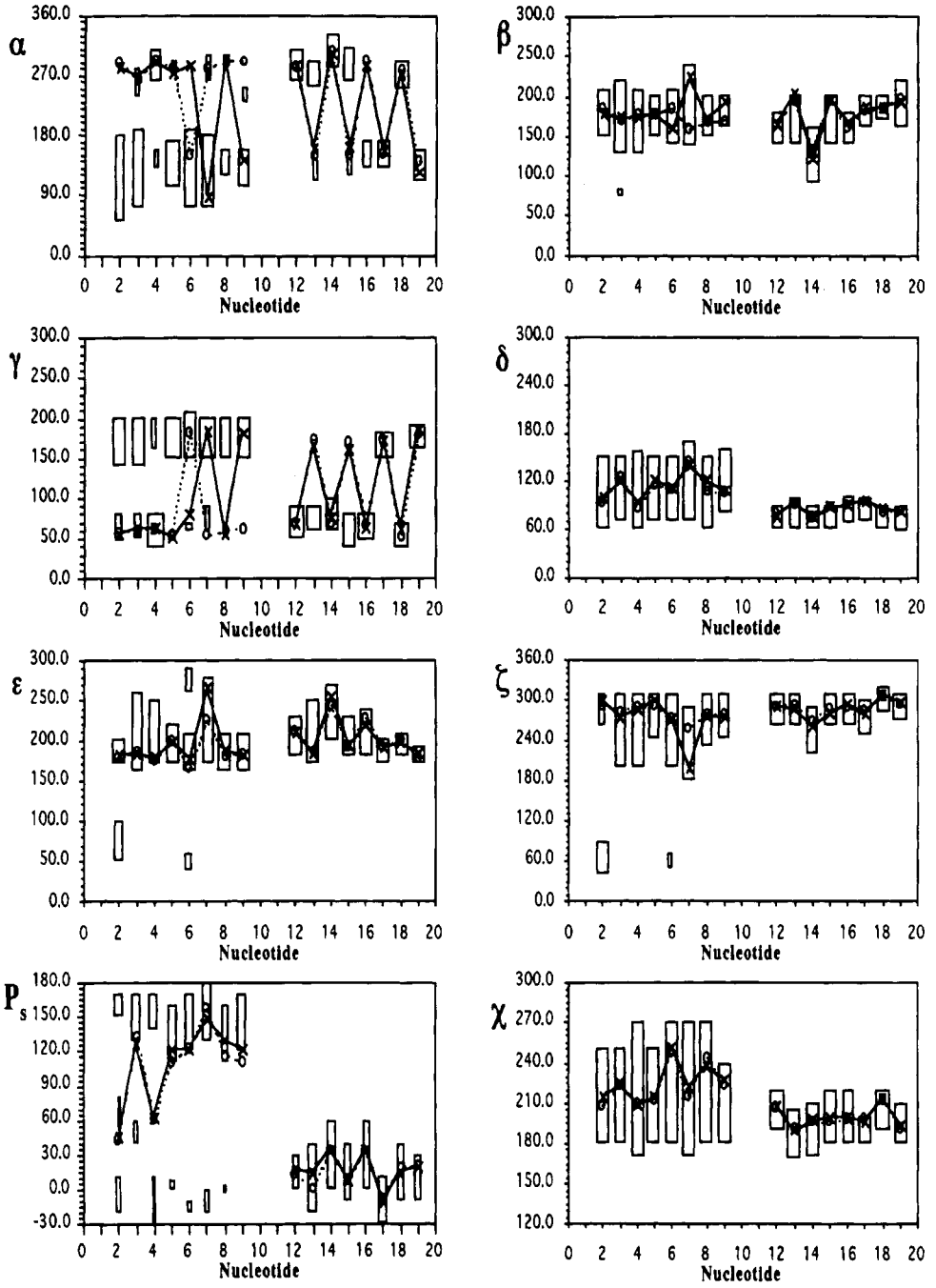

FIGURE 3: Backbone torsion angles, sugar puckers (P_s) and glycosidic torsion angles (χ) for the nonterminal nucleotides in the final average structures. Crosses and solid lines indicate *S*-duplex; circles and dashed lines, *R*-duplex. Boxes indicate the range of values for each angle with significant population in the MD-tar water ensemble. The total number of snapshots are classified in a 10°-step range of values and population larger than 3% are considered significant. Broad boxes indicate the major conformer, and narrow boxes mark minor conformers.

force field is not designed for this kind of unrealistically rapid transitions, the result of our MD-tar trajectory in water is considered a better ensemble of structures.

Structural Analysis. Sugar Pucker and Glycosidic Torsion Angles. Figure 3 shows backbone and glycosidic torsion angles as well as pseudorotation phase angles along the sequence for the average structure. The range of values with a significant population in the ensemble of structures obtained with MD-tar in water are also indicated.

As previously reported by a number of authors (Salazar et al., 1993a; Fedoroff et al., 1993; Lane et al., 1993; González et al., 1994), one of the most distinctive features of the DNA·RNA hybrid duplexes is the different sugar pucker in the DNA and the RNA strand with the average structure of our hybrid, pseudorotation phase angles shown in Figure 3 present values between −10° and 35° for the RNA strand and values ranging from 50° to 160° for the DNA strand. While riboses are clearly in an N conformation (near C3′-endo), classification of deoxyriboses is more complex. Some of them show pseudorotation phase angles characteristic of the S domain (near C2′-endo conformation), but others present smaller values, in the C1′-exo or even O4′-endo domain. Sugar puckers of deoxyriboses in an intermediate domain with P_s ~90° have the general E-type domain have been claimed to be a characteristic of hybrid duplexes (Salazar et al., 1993a; Fedoroff et al., 1993), but we have reported in our previous paper (González et al., 1994) that, at least in this duplex, such a conformation is

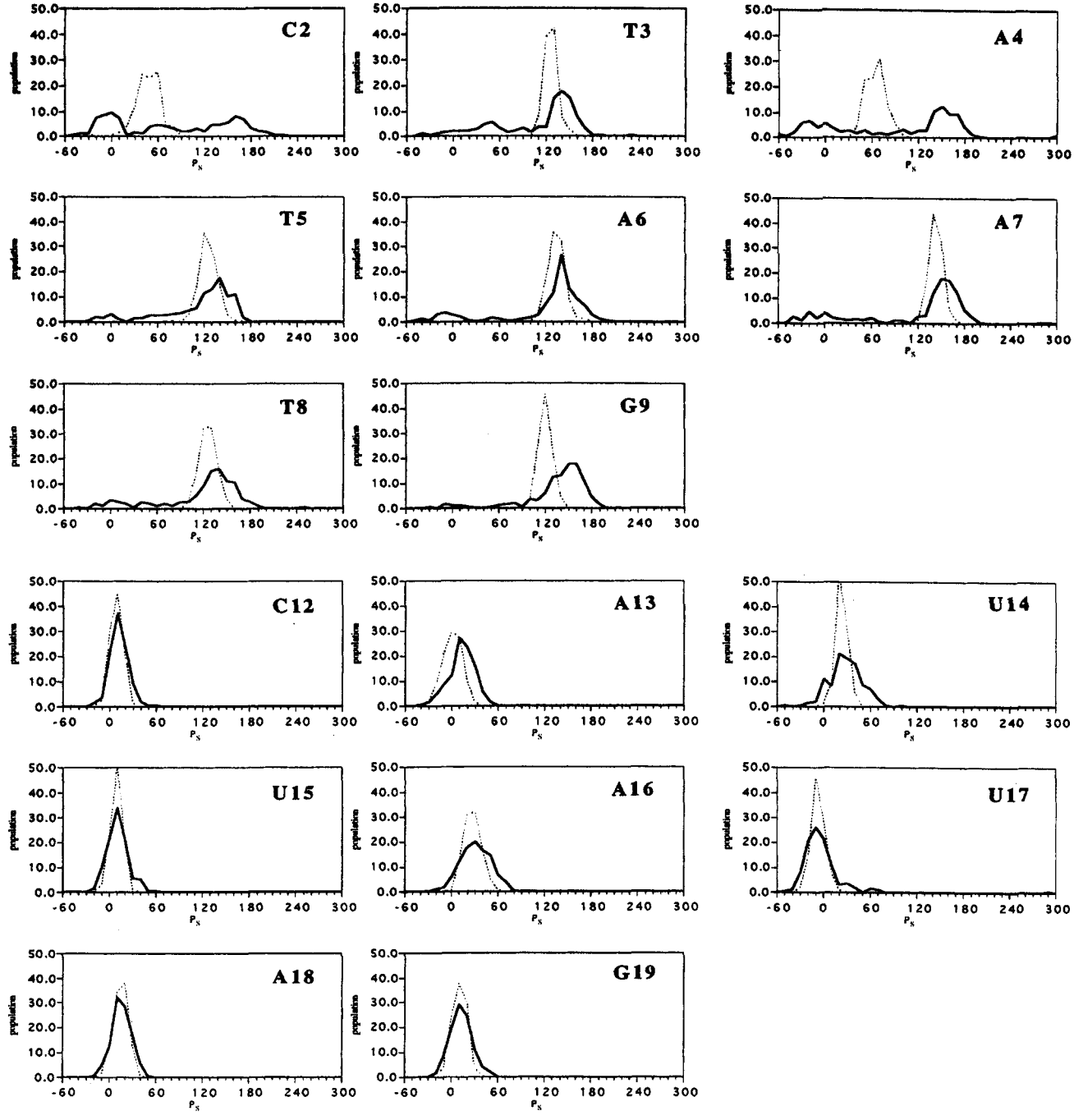

FIGURE 4: Distribution analysis of pseudorotation phase angles for nonterminal residues. (—) MD-tar in water; (- - -) rMD in water (200 snapshots over 120 ps were included).

not consistent with the experimental J coupling information. In fact, the low pseudorotation phase angles obtained for some residues in the DNA strand of the average structures are responsible of the large J_{rms} observed (see Table 2). Similar average values for P_S, the pseudorotation phase angle, are obtained in the rMD trajectory, with distributions of P_S always centered around the value obtained in the average structure (see Figure 4). This indicates that the inclusion of explicit water does not have an effect on the sugar pucker of the average structures.

The situation is completely different when the ensemble of structures obtained in the MD-tar trajectory is considered. Most of deoxyriboses manifest distributions with pseudorotation phase angles of the major conformer in the S-domain and a minor conformer in the N-domain. In many cases, a small population of intermediate pucker (P_S ~90°) is also present. The population of minor conformer is particularly large for those residues which exhibit lower P_S in the final structure RMD-FIN, C2 and A4, indicating that low pseudorotation phase angles, i.e., 90−120°, are artifacts of the average structure. In all deoxyriboses, the maximum of the distribution of P_S for the major conformer is shifted to higher values than in the average structure. Although the distributions shown in Figure 4 present a more complex pattern than a simple two-state model, it is interesting to compare the total population of major conformer with the occupancies of S and N conformers obtained from analysis of coupling constants assuming only two states (González et al., 1994). This comparison is shown in Table 3 together with the rms deviations between calculated and experimental J coupling

Table 3: Comparison of the Percentage of *S* Conformer and Average rms Deviation of Coupling Constants for Each Residue of the DNA Strand

	%*S* two-state model[a]	%*S* MD-tar water[b]	J_{rms} (Hz)[c]
C2	55−67	42	0.72
T3	57−69	63	0.78
A4	57−69	53	0.50
T5	78−86	79	0.32
A6	79−100	80	0.39
A7 *R*	56−67	76	0.75
A7 *S*	67−71	73	0.78
T8		74	
G9	79−100	90	0.41

[a] Percentage of *S*-conformer derived from a two-state model (González et al., 1994). [b] All conformations between 90° and 270° included. [c] Average rms deviation for closest of error bounds for coupling constants $J_{1'2'}$, $J_{1'2''}$, $J_{2'3'}$, and $J_{2''3'}$.

constants per residue. In general, the agreement is reasonably good except for C2, where the P_s distribution (Figure 4) presents a much more complicated behavior than a two-state model.

Bimodal distributions are not observed in the case of riboses. Populations obtained with rMD and MD-tar are very similar, with the maximum coincident with the value obtained in the average structure. The only difference affecting the RNA strand between the two trajectories is a broadening in the P_s distributions for some residues (see U14 and A16 in Figure 4) in the MD-tar ensemble. It is worth noting that MD-tar does not enhance the flexibility of those regions of the molecule where it is not required by the experimental restraints. As pointed out in our previous paper, all distance constraints indicate that riboses in the RNA strand are in the N-type domain. The lack of H1′−H2′ DQF-COSY cross-peaks in the RNA strand, indicating very low values for $J_{1'}$ coupling constants, is also in accord with a single N conformation for riboses, since small population of a minor confomer in the S domain would increase the $J_{1'}$ coupling constants to observable values.

In the RNA strand, glycosidic torsion angles exhibit very well defined values in the range between 190° and 210°, typical of A-form structures. Analysis of the MD-tar trajectories does not show significant population of χ values higher than 220°, and this only occurs in G12 and A18 (see boxes in Figure 3). In the DNA strand, χ values for the average structure are in the range between 210° and 250°. When the ensemble of structures is considered, a much bigger dispersion in χ values is observed in the DNA than in the RNA strand. This result is consistent with the presence of minor conformers only in deoxyriboses. In those residues where transitions in sugar puckers are detected, a correlation between the glycosidic torsion angle and the pseudorotation phase angle is clearly observed (data not shown).

Backbone Torsion Angles. Values of backbone torsion angles along the sequence are shown in Figure 3 together with the range of values appreciably populated in the ensemble of structures obtained with MD-tar in water. In general, backbone torsion angles, other than δ, are not well defined by the experimental constraints due to the limited number of protons in this region. Although 12 distance restraints involving H5′ and H5″ could be determined in our hybrid duplex, the lack of stereospecific assignments and the consequent pseudoatom correction in the target distance render these constraints not very informative.

Not much variation of the backbone torsion angles β, δ, ϵ, and ζ is observed along the sequence with the exception of the modified nucleotide (A7) and the cross-paired residue (U14), the latter only in torsion angle β and ϵ. All other nucleotides present torsion angles commonly found in double-stranded DNA. The interval of values with a significant population in the ensemble of structures is always centered at the value of the average structure. With the exception of C2, and to a smaller extent A6, minor conformers are not present. The range of values detected in the ensemble is significantly narrower in the RNA than in the DNA strand. It must be noticed that the largest difference for these torsion angles between A7 and the unmodified nucleotides correspond to the *S* stereoisomer.

Contrary to these backbone torsion angles, α and γ show a significant number of transitions between g^- and t for α and g^+ and t for γ, both in the DNA and the RNA strand. These transitions have been observed in other NMR studies (Schmitz et al., 1993; Weisz et al., 1994) and in theoretical calculations, both in double-stranded DNA (Keepers et al., 1982) and in DNA·RNA hybrids (Fritsch & Wolf, 1994; Sanghani & Lavery, 1994). In all those cases, a correlation between transitions in α and γ angles have been observed. The relationship between these two torsion angles is also apparent in crystal structures (Fratini et al., 1982). In the case of these two torsion angles, the values obtained in the average structure do not always correspond to the major conformer observed in the MD-tar trajectory. Conventional rMD trajectories present values centered in the initial torsion angles of the average structure, and these transitions are detected in only a few residues. It is noteworthy that the occurrence of transitions does not depend on the inclusion of explicit solvent, although its frequency is larger in MD-tar *in vacuo*. Inspection of α and γ torsion angles in the nine structures obtained in the calculation of the average structure show different values of α and γ torsion angles, with no significant preference for the value found in the final structure RMD-FIN. It should be noted that the angles α

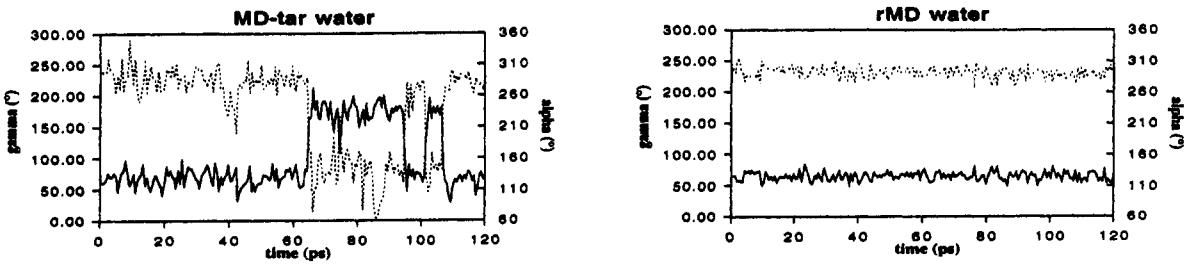

FIGURE 5: Time-dependent fluctuation of the α (- - -) and γ (—) angles in residue A4 for MD-tar and rMD trajectories run with a shell of explicit water molecules.

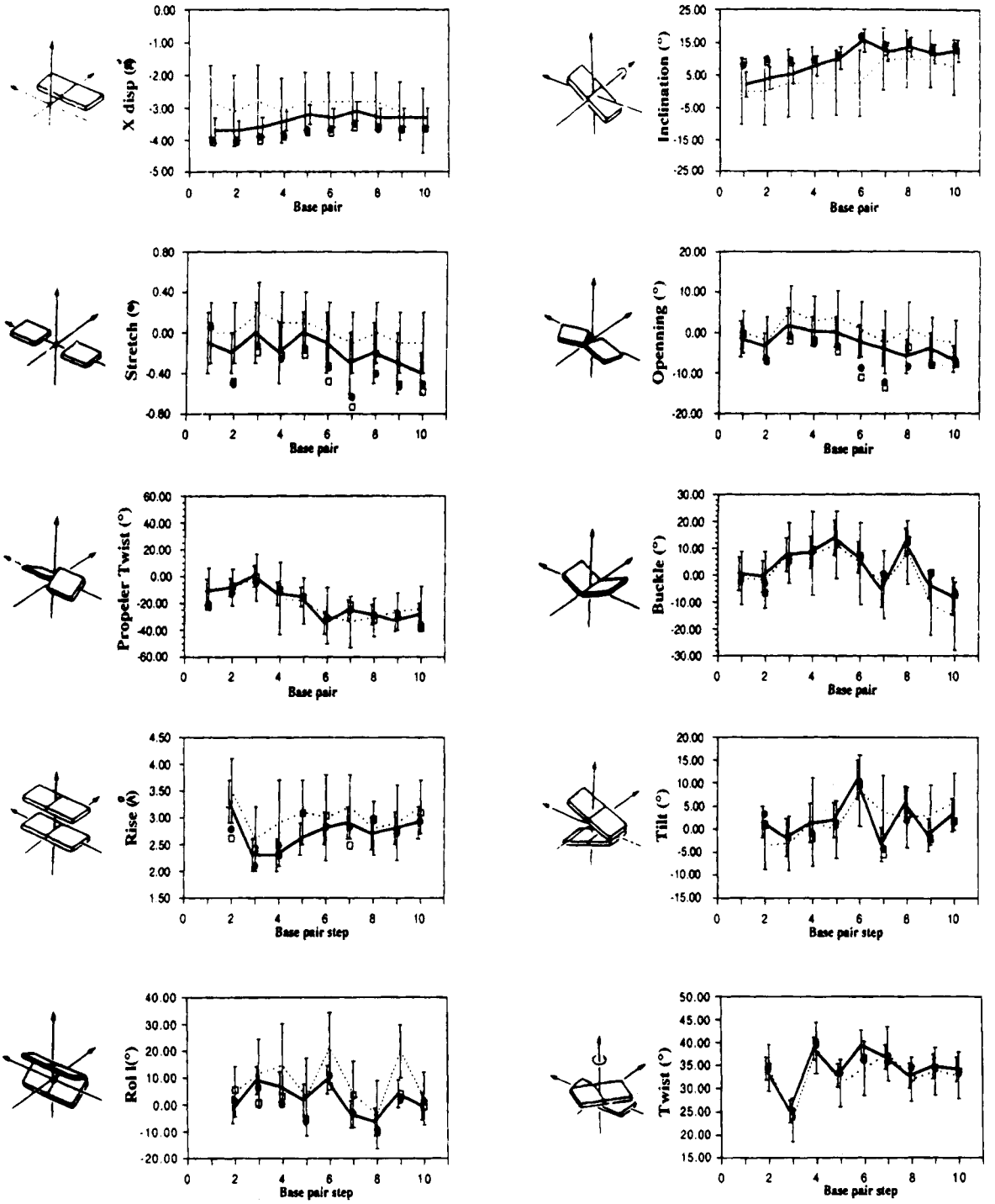

FIGURE 6: Selected helicoidal parameters for the final average structures: *S*-duplex (●); *R*-duplex (○). Solid lines indicate average values for the rMD trajectory in water and dashed line are for MD-tar in water. Error bars indicate the rms deviation of the different parameters in the two trajectories. The larger bounds correspond to the MD-tar ensemble and the smaller ones to the rMD ensemble.

and γ are ill-constrained by the experimental data, and the values reported here may arise largely from the force field potentials employed.

Helicoidal Parameters. Some helicoidal parameters calculated with the program Dials and Windows (Ravishanker et al., 1989) are shown in Figure 6. Values for the RMD-FIN structures in the S and R duplexes are presented and compared with the average values along the rMD and MD-tar trajectories in water. Error bounds obtained from the rms deviation for each parameter in the different ensembles are also included. Base−axis parameters are calculated with respect to the global helix axis.

The distribution of helix parameters becomes very broad when the whole ensemble of structures is considered. These broad distributions are represented by the large error bounds displayed in Figure 6. In general, all the distributions are reasonably symmetric, not revealing distinct major or minor conformers. Only the base−axis parameters, *X*-displacement and inclination, exhibit skewed distributions in the MD-tar trajectories. In both cases, the highest population for each parameter corresponds approximately with its value in the average structure. However, as a result of the bias toward higher values in the case of *X*-displacement or lower values in the case of inclination, the average of each distribution

differs from the corresponding value in the RMD-FIN structure. The difference between values for MD-tar and the average structure displayed in Figure 6 arises from the calculation of the mean for nonsymmetric distributions (Press et al., 1989) and not from a real structural feature. Strong correlation between *X*-displacement and inclination is observed, as reported in other theoretical studies (Poncin et al., 1992; Schmitz & James, 1994).

Translational intra-base-pair parameters, shear, stretch and stagger, are small in all residues, presenting values typical of regular helices. Some compression in the base pairs along the sequence is evident in the RMD-FIN structures, as indicated by the negative stretch observed. This effect is eliminated in the MD-tar ensemble. A similar tendency is observed in the opening parameter, where negative values are strongly reduced in rMD and MD-tar ensembles. It must be pointed out that this reduction is not an effect of considering the solvent explicitly, since both rMD and MD-tar *in vacuo* show the same tendency (data not shown).

Similar values for propeller twist and buckle are found for all the different calculations. In spite of the large error bounds associated with these parameters, sequence dependent trends are preserved for all of the different ensembles. In general, all base pairs exhibit negative propeller twist values, except for T3A18. This seems to be a characteristic common to AT base pairs in double-stranded DNA (Schmitz et al., 1992a; Arnott et al., 1983). On the other hand, values of the buckle parameter are close to zero or positive for the whole sequence, in contrast to the values obtained by Fedoroff et al. (1993) but more similar to fiber diffraction data (Arnott et al., 1986).

The base pair step parameters shift and slide are, in general, small and quite similar for all the trajectories. Rise presents low values between 2.1 and 3.0 Å for the average structures and slightly higher for MD-tar trajectories. Somewhat larger sequence dependent differences are observed for the three rotational inter-base-pair parameters. High roll values are observed in two out of three pyrimidine−purine steps, T5A16•A6U15 and T8U13•G9C12, and a high tilt value in T5A16•A6U15. The base pair step C2G19•T3A18 presents the lowest twist, 25°, that is compensated in the following step with the value of 40°. The three pyrimidine−purine steps exhibit the highest twist values.

Interphosphate Distance and Minor Groove Size. Two essential features of nucleic acid structures are the size of the minor groove and the interphosphate distance. As shown in Figure 7, the minor groove width of the hybrid duplex is intermediate between the typical values for A- and B-form structures. This seems to be a common characteristic of hybrid duplexes as observed in other hybrids studied in solution (Fedoroff et al., 1993; Lane et al., 1993). Remarkably, this result does not depend on the method used for the refinement of the structures. Similar minor groove widths are observed in the average structure and in the MD-tar ensemble. Error bars are larger in MD-tar trajectories than in rMD, but the uncertainty is much lower than the difference between the A- and B-forms. In addition, *in vacuo* and explicit water calculations manifest quite similar average values (data not shown).

A reduced minor groove width compared to A-form structures has been proposed as the main reason for the specific recognition of hybrid duplexes by ribonuclease H (Fedoroff et al., 1993). Enhanced flexibility has been also

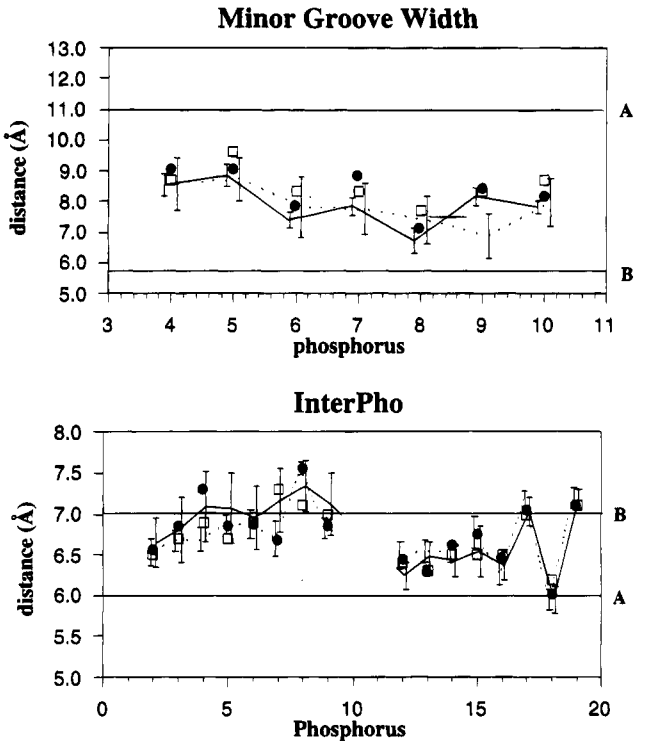

FIGURE 7: Minor groove width and inter-phosphate distance for the *S*-duplex (●), *R*-duplex (○), rMD trajectory (—), and MD-tar trajectory (- - -) in water. Values for canonical A- and B-form DNA are indicated for comparison.

alleged by other authors as a required feature for proper recognition (Nakamura et al., 1991). As shown in this study, backbone flexibility, with deoxyriboses in rapid exchange between S- and N-type puckers, is completely consistent with an intermediate minor groove width. Therefore, the two explanations for the specificity of ribonuclease H are not contradictory.

Interphosphate distances present a large variation along the sequence with the highest value in the step A7p−T8p, next to the modified phosphate, and the lowest values in the step U17p−A18p. Although the error bars are large, the average values are in general more similar to the B-family of structures, especially in the DNA strand.

Different Flexibility in DNA and RNA Strands. Strong evidence for intrinsically different flexibility between both strands in this hybrid duplex was obtained from the analysis of the experimental *J* coupling information carried out in our previous paper. This result is confirmed by the MD-tar calculations presented here. The different flexibility is reflected in the presence of at least two conformers in deoxyriboses (Figure 4) and in an augmented variation in the backbone torsion angles for the DNA strand (Figure 3). This can also be observed by calculating the atomic rms deviations along the MD-tar trajectory in water for different parts of the molecule. RMSDs are higher in the backbone than in the base region, but the difference is more significant when only atoms belonging to each of the strands are considered. The difference is especially important in the backbone, where the RMSD is 0.97 Å for the DNA and 0.68 Å in the RNA strand. The corresponding values in the rMD calculations are 0.49 Å for the DNA and 0.38 Å for the RNA, indicating that the time-averaged constraints induce more fluctuations in the DNA than in the RNA strand. This

enhanced flexibility in the DNA strand backbone is apparent in Figure 2b.

Differences between S and R Stereoisomers. All the NMR experiments and structural calculations have been carried out for both stereoisomers of the modified phosphate. Some differences in chemical shifts could be detected for protons in the proximity of the thiophosphate, but the difference in the intensity of the NOE cross-peaks was not significant, so identical distance constraints were used in the structural calculations. A slight variation in the J coupling constants of the preceding deoxyribose (A7) to the modified phosphate was observed (González et al., 1994). Consequently, different sets of J coupling constraints were used in the calculations for the S and R stereoisomers. Unfortunately, cross-peak overlapping prevented an estimation of the J coupling constants for the residue T8.

As expected, not much difference between the two stereoisomers was detected either in the average structure or in the ensemble calculations. Globally, both structures are identical with no significant variation in helical parameters. Most of the backbone torsion angles are the same (Figure 3). The largest difference was observed in the α and γ torsion angles of the modified nucleotide and the preceding nucleotide (Figure 3). Some slight variation was also detected in the β and ϵ torsions, but in this case the discrepancy is lower than the error bounds estimated from the MD-tar trajectories. The small difference in J coupling values for the nucleotide A7 is not reflected in a significant change in its pseudorotation phase angle. No variation is observed in any χ torsions.

Minor groove width in the S- and R-duplex are quite similar, indicating that the substitution does not have a significant effect on the size of the minor groove. Contrary to the case of B-form double-stranded DNA, where significant distortions produced by a diphosphorothioate substitution have been detected (Cho et al., 1993), no difference is observed between the modified and the unmodified phosphates. The increased minor groove width with respect to standard values in B-form structures allows for the sulfur atom to accommodate well into the two positions without need for any substantial distortion in the structure of the duplex.

CONCLUSIONS

The attempt to fit simultaneously all the experimental restraints obtained from NMR experiments can lead to artifacts when multiple conformers in dynamic exchange are present. The solution structure of this DNA·RNA hybrid duplex is especially illustrative, since different degrees of flexibility occur in each strand. The DNA strand is more flexible with most deoxyriboses undergoing rapid exchange between different conformations, whereas the riboses in the RNA strand remain in the N-type domain. The determination of the average structure of the duplex by using conventional restrained molecular dynamics leads to intermediate sugar puckers (approximately O4′-endo) for some deoxyriboses in the DNA strand. These intermediate puckers can be identified as structural artifacts by comparing the vicinal J coupling constants with the experimental values [see Figure 4 in González et al., (1994)]. A better description of the structure of this duplex can be achieved by using MD-tar calculations. The ensemble of dynamically exchanging structures resulting from the MD-tar trajectory reproduces the experimental J coupling constants, as indicated by the low J_{rms} obtained, and satisfies the distance constraints better than the structures obtained with the conventional rMD method. NMR R factors and residual distance violations decrease more in the DNA than in the RNA strand, suggesting that some internal contradiction between the NOE information in the DNA strand is also removed when dynamic effects are considered.

The sugar pucker distributions for the deoxyriboses are mainly bimodal, with the population of the major conformer (always in the general S-domain) in good agreement with the values obtained by a two-state interpretation of the J coupling constants. However, sugar pucker distributions are more complex than a simple two-state model, and a very small population of intermediate (P_S ~90°) conformers is also present. It must be underlined that no minor conformers are observed for the RNA strand. The effect of carrying out molecular dynamics calculations with time-averaged restraints, instead of conventional restraints leads only to a broadening of the distribution of sugar puckers in riboses. MD-tar does not produce extra flexibility in those places where is not necessary for fitting the experimental data.

Although sugar puckers in the unusual intermediate domain with P_S ~90° appear to be artifacts of the refinement method used, some other features remain in all the calculations. Most importantly, the minor groove width exhibits intermediate values between standard A- and B-forms in the rMD structure and in the MD-tar ensemble. The minor groove width is also independent of the inclusion of explicit water in the calculation. This is an important result, since a reduced minor groove has been proposed as responsible for the specificity of the ribonuclease H for hybrid duplexes.

A minor groove width larger than in the B-form family of structures may be responsible of the great similarity observed between the two stereoisomers studied. The sulfur atom can accommodate well into both positions without significant distortion in the global structure.

ACKNOWLEDGMENT

We gratefully acknowledge the help of Dr. Klaus Weisz during the initial phase of this project and Drs. Nikolai Ulyanov and Uli Schmitz for many useful discussions. C.G. acknowledges a postdoctoral fellowship from the Spanish Consejo Superior de Investigaciones Científicas. We acknowledge use of the Computer Graphics Laboratory at the University of California San Francisco and the Pittsburgh Supercomputing Center.

REFERENCES

Arnott, S., Chandrasekaran, R., Pigjaner, L. C., Walker, J. K., Hall, I. H., & Birdsall, D. L. (1983) *Nucleic Acids Res. 11*, 1457–1474.

Arnott, S., Chandrasekaran, R., Millane, R. P., & Park, H. S. (1986) *J. Mol. Biol. 188*, 631–640.

Bonvin, A. M. J. J., Boelens, R., & Kaptein, R. (1994) *J. Biomol. NMR 4*, 143–149.

Borgias, B. A., & James, T. L. (1988) *J. Magn. Reson. 79*, 493–512.

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

Brüschweiler, R., Blackledge, M., & Ernst, R. R. (1991) *J. Biomol. NMR 1*, 3−11.
Cho, Y., Zhu, F. C., Luxon, B. A., & Gorenstein, D. G. (1993) *J. Biomol. Struct. Dyn. 11*, 685−702.
Chou, S.-H., Flynn, P., & Reid, B. (1989) *Biochemistry 28*, 2435−2443.
Davies, D. B. (1978) *Prog. Nucl. Magn. Reson. Spectrsc. 12*, 135−225.
de Leeuw, F. A. A. M., van Beuzekom, A. A., & Altona, C. (1983) *J. Comput. Chem. 4*, 438−448.
Egli, M., Usman, N., Zhang, S., & Rich, A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 534−538.
Fedoroff, O. Y., Salazar, M., & Reid, B. R. (1993) *J. Mol. Biol., 233*, 509−523.
Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686−14707.
Fritsch, V., & Wolf, R. M. (1994) *J. Biomol. Struct. Dyn. 6*, 1161−1174.
Gao, X., Brown, F. K., Jeffs, P., Bischofberger, N., Lin, K., Pipe, A. J., Noble, S. A. (1992) *Biochemistry 31*, 6228−6235.
González, C., Rullmann, J. A. C., Bonvin, A. M. J. J., Boelens, R., & Kaptein, R. (1991) *J. Magn. Reson. 91*, 659−664.
González, C., Stec, W., Kobylanska, A., Hogrefe, R. I., Reynolds, M., & James, T. L. (1994) *Biochemistry 33*, 11062−11072.
Guenot, J., & Kollman, P. A. (1992) *Protein Sci. 1*, 1185−1205.
Guenot, J., & Kollman, P. A. (1993) *J. Comput. Chem. 14*, 295−311.
Haasnoot, C. A. G., de Leeuw, H. P. M., & Altona, C. (1980) *Tetrahedron 36*, 2783−2792.
Hall, K. B. (1993) *Curr. Opin. Struct. Biol. 3*, 336−339.
Heinemann, U., Rudolph, L. N., Claudis, A., Morr, M., Heikens, R. F., & Blocker, H. (1991) *Nucleic Acids Res. 19*, 427−433.
Jaroszewski, J. W., Syi, J.-L., Maizel, J., & Cohen, J. S. (1992) *Anti-Cancer Drug Des. 7*, 253−262.
Jorgensen, W. L., Chandraskhar, J., Madura, J., Impey, R. W., & Klein, M. L. (1983) *J. Chem. Phys. 79*, 926−935.
Katahira, M., Lee, S. J., Kobayashi, Y., Sugeta, H., Kyogoku, Y., Iwai, S., Ohtsuka, E., Benevides, J. M., & Thomas, G. J. (1990) *J. Am. Chem. Soc. 112*, 4508−4512.
Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404−426.
Keepers, J. W., Kollman, P. A., Weiner, P. K., & James, T. L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5537−5541.
Kim, S. G., Lin, L. J., & Reid, B. R. (1992) *Biochemistry 31*, 3564−3574.
Lane, A. N., Ebel, S., & Brown, T. (1993) *Eur. J. Biochem. 215*, 297−306.
Lavery, R., & Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63−91.
Lavery, R., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655−667.
Lavery, R., & Sklenar, H. (1990) *CURVES. Curves 3.0, helical analysis of irregular nucleic acids*, Laboratory of Theoretical Biochemistry CNRS, Paris.
Liu, H., Thomas, P. D., & James, T. L. (1992a) *J. Magn. Reson. 98*, 163−175.
Liu, H., Kumar, A., Borgias, B. A., & James, T. L. (1992b) *CORMA 3.0*, University of California, San Francisco.
Liu, H., Kumar, A., Weisz, K., Schmitz, U., Bishop, K., & James, T. L. (1993) *J. Am. Chem. Soc. 115*, 1590−1591.
Mierke, D. F., Huber, T. H., & Kessler, H. (1994) *J. Comput.-Aided Mol. Des. 8*, 29−40.
Mujeeb, A., Kerwin, S. M., Kenyon, G. L., & James, T. L. (1993) *Biochemistry 32*, 13419−13431.
Nakamura, H., Oda, Y., Iwai, S., Inoue, H., Ohtsuka, E., Kanaya, S., Kimura, S., Katsuda, C., Katayanagi, K., Morikawa, K., Miyashiro, H., & Ikehara, M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 11535−11539.
Pearlman, D. A. (1994a) *J. Biomol. NMR 4*, 1−16.
Pearlman, D. A. (1994b) *J. Biomol. NMR 4*, 279−299.
Pearlman, D. A., & Kollman, P. A. (1991) *J. Mol. Biol. 220*, 429−457.
Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibel, G. L., Singh, U. C., Weiner, P., & Kollman, P. A. (1991) *AMBER4.0*, University of California, San Francisco.
Poncin, M., Hartmann, B., & Lavery, R. (1992) *J. Mol. Biol. 226*, 775−794.
Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989) *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge.
Ravishanker, G., Swaminathan, D. L., Beveridge, D. L., Lavery, R., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 669−699.
Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Chem. 23*, 327−341.
Salazar, M., Champoux, J. J., & Reid, B. R. (1993a) *Biochemistry 32*, 739−744.
Salazar, M., Fedoroff, O. Y., Miler, J. M., Ribeiro, N. S., & Reid, B. R. (1993b) *Biochemistry 32*, 4207−4215.
Sanghani, S. R., & Lavery, R. (1994) *Nucleic Acids Res. 22*, 1444−1449.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Schmitz, U., & James, T. L. (1994) in *Structural Biology: The State of the Art* (Sarma, R. H., & Sarma, M. H., Eds.) pp 251−272, Adenine Press: New York.
Schmitz, U., Sethson, I., Egan, W. M., & James, T. L. (1992a) *J. Mol. Biol. 227*, 510−531.
Schmitz, U., Kumar, A., & James, T. L. (1992b) *J. Am. Chem. Soc. 114*, 10654−10656.
Schmitz, U., Ulyanov, N. B., Kumar, A., & James, T. L. (1993) *J. Mol. Biol. 234*, 373−389.
Stolarski, R., Egan, W., & James, T. L. (1992) *Biochemistry 31*, 7027−7042.
Thomas, P. D., Basus, V. J., & James, T. L. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1237−1241.
Torda, A. E., Scheek, R. M., & van Gunsteren, W. F. (1989) *Chem. Phys. Lett. 157*, 289−294.
Torda, A. E., Scheek, R. M., & van Gunsteren, W. F. (1990) *J. Mol. Biol. 214*, 223−235.
Torda, A. E., Brunne, R. M., Huber, T., Kessler, H., & van Gunsteren, W. F. (1993) *J. Biomol. NMR 3*, 55−66.
Uhlmann, E., & Peyman, A. (1990) *Chem. Rev. 90*, 543−584.
Ulyanov, N. B., Schmitz, U., & James, T. L. (1993) *J. Biomol. NMR 2*, 547−568.
Ulyanov, N. B., Schmitz, U., Kumar, A., & James, T. L. (1995) *Biophys. J. 68*, 13−24.
Varmus, H. (1988) *Science, 240*, 1427−1435.
Weiner, P. K., Kollman, P. A., Nguyen, D. T., & Case, D. (1986) *J. Comput. Chem. 7*, 230−252.
Weisz, K., Shafer, R. H., Egan, W., & James, T. L. (1994) *Biochemistry 33*, 354−366.
Widmer, H., & Wüthrich, K. (1986) *J. Magn. Reson. 70*, 270−279.
Widmer, H., & Wüthrich, K. (1987) *J. Magn. Reson. 74*, 331−336.
Zon, G. (1988) *Pharm. Res. 5*, 539−549.

BI942821K