Journal of Biomolecular NMR, 10 (1997) 337–350
KLUWER/ESCOM
© 1997 *Kluwer Academic Publishers. Printed in The Netherlands.*

J-Bio NMR 471

Analysis of ^{1}H chemical shifts in DNA: Assessment of the reliability of ^{1}H chemical shift calculations for use in structure refinement

Sybren S. Wijmenga[a,*], Martijn Kruithof[b] and Cees W. Hilbers[b]

[a]*Department of Medical Biochemistry and Biophysics, Umeå University, S-901 87 Umeå, Sweden*
[b]*Nijmegen Son Research Center, Department of Biophysical Chemistry, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands*

Received 3 April 1997
Accepted 20 June 1997

Keywords: Chemical shifts; Chemical shift calculation; DNA; RNA; Nucleic acids

Summary

The reliability of ^{1}H chemical shift calculations for DNA is assessed by comparing the experimentally and calculated chemical shifts of a reasonably large number of independently determined DNA structures. The calculated chemical shifts are based on semiempirical relations derived by Giessner-Prettre and Pullman [(1987) *Q. Rev. Biophys.*, **20**, 113–172]. The standard deviation between calculated and observed chemical shifts is found to be quite small, i.e. 0.17 ppm. This high accuracy, which is achieved without parameter adjustment, makes it possible to analyze the structural dependencies of chemical shifts in a reliable fashion. The conformation-dependent ^{1}H chemical shift is mainly determined by the ring current effect and the local magnetic anisotropy, while the third possible effect, that of the electric field, is surprisingly small. It was further found that for a double helical environment, the chemical shift of the sugar protons, H2' to H5", is mainly affected by the ring current and magnetic anisotropy of their own base. Consequently, the chemical shift of these sugar protons is determined by two factors, namely the type of base to which the sugar ring is attached, C, T, A, or G, and secondly by the χ-angle. In particular, the H2' shift varies strongly with the χ-angle, and strong upfield H2' shifts directly indicate that the χ-angle is in the *syn* domain. The H1' shift is not only strongly affected by its own base, but also by its 3'-neighboring base. On the other hand, base protons, in particular H5 of cytosine and methyl protons of thymine, are affected mainly by the 5'-neighboring bases, although some effect (0.2 ppm) stems from the 3'-neighboring base. The H2 protons are mainly affected by the 3'-neighboring base. As a result of these findings a simple scheme is proposed for sequential assignment of resonances from B-helices based on chemical shifts.

Introduction

The success of NMR spectroscopy as an analytical method in chemistry and biochemistry is based on the spectral dispersion of the resonance positions, the chemical shifts, of the individual nuclei. It has long been recognized that the chemical shift values depend on the electron densities around the nuclei, which can in subtle, but well measurable, ways be influenced by the nuclear surroundings and therefore may contain important structural information. Application of the chemical shift as a tool to derive three-dimensional characteristics of molecules has been attempted in the past, but has been overtaken by the use of J-couplings and NOE effects. Recently, however, a renewed interest in the chemical shift as a structural tool has emerged (see for recent reviews, e.g. Case et al., 1994; Wishart and Sykes, 1994; Szilagyi, 1996). For a large number of proteins the NMR spectra have been interpreted in detail, providing a large database of ^{1}H, ^{13}C, and ^{15}N chemical shifts. In addition, for a number of these proteins the solution structure has been determined with high accuracy from NOEs and J-couplings, and for some also X-ray structures are available. Combination of these data has paved the way for a reliable interpretation of the chemical shift in terms of structural parameters, either through a set of purely empirical rules or based on

*To whom correspondence should be addressed.

physical parameters. The latter can be derived via ab initio quantum mechanical or semiempirical calculations. As a result, so-called chemical shift indexes have become available which can be used as secondary structure identifiers in protein structure determination (Wishart and Sykes, 1994). Furthermore, for ^{1}H nuclei the reliability of the chemical shift calculations has been investigated and the main parametric dependencies have been established (Ösapay and Case, 1991,1994; Asakura et al., 1992,1995; Williamson and Asakura, 1993; Case et al., 1994; Williamson et al., 1995). For ^{13}C nuclei new and more extensive quantum mechanical calculations have shown that their chemical shift can be derived in a reliable fashion (de Dios et al., 1993; Laws et al., 1993; Oldfield, 1995). For ^{15}N chemical shifts similar attempts have also been made (Le and Oldfield, 1994). Finally, the first refinements of NMR protein structures against chemical shifts have been performed based on ^{1}H and/or ^{13}C chemical shifts (Ösapay et al., 1994; Celda et al., 1995; Kuszewski et al., 1995a,b).

In the meantime a number of nucleic acid NMR structures have been published, which makes it worthwhile to consider a similar approach for this class of molecules. Attempts to use ^{13}C-shifts to determine the glycosidic torsion angle have already yielded promising results (Ghose et al., 1994; Greene et al., 1995). Here, we concentrate on the use of ^{1}H chemical shifts. The reliability of ^{1}H chemical shift calculations is assessed by comparing the experimentally and calculated chemical shifts for a number of independently determined structures. The calculated chemical shifts are based on semiempirical relations derived by Giessner-Prettre and Pullman (1987). We find that the standard deviation between calculated and observed chemical shifts is quite small, i.e. 0.17 ppm. This high precision is achieved without adjusting parameters. This result allows a reliable analysis of the structural dependencies of chemical shifts. We find for a double helical environment the following. The position of the sugar proton resonances from H2' to H5" is mainly affected by the ring current and magnetic anisotropy of the base attached to this sugar. Consequently, the chemical shift of these sugar protons is determined by two factors, namely the type of base to which the sugar ring is attached, C, T, A, or G, and secondly, by the χ-angle. In particular, the chemical shifts of the H2' protons vary strongly with the χ-angle. Sequential effects are, in addition, only significant for the H1' sugar protons and stem mainly from the 3' neighboring base. Resonance positions of the base protons, in particular H5 of cytosine and the methyl of thymine are, on the other hand, affected mainly by the 5'-neighboring bases, although the 3'-neighboring base makes a contribution up to 0.2 ppm. The H2 shifts are mainly affected by the 3'-neighboring base. As a result of these findings we propose a simple scheme for sequential assignment of resonances based on chemical shifts.

Methods

The chemical shift of spin l in a molecule C depends on the electron density distribution around the nucleus. The electron density and the corresponding chemical shift could be obtained, in principle, by ab initio quantum mechanical calculations. However, such calculations have been prohibitively computer intensive, although recent progress in quantum mechanical computational procedures and computer hardware has made it possible to perform such calculations for molecular fragments large enough to reflect the essential features of the local environment (de Dios et al., 1993; Laws et al., 1993; Oldfield, 1995). For chemical shift calculations it is thus operationally expedient to divide a molecule into a number of fragments, i.e. into a fragment A, where the nucleus of interest resides, and which has a conformation defined with respect to a reference conformation, and a number, N, of fragments B interacting with A. The calculated shift contributions can then be divided into two categories, namely a conformation-independent part, δ_{intrin}, and a conformation-dependent part, δ_{conf}, respectively. The former represents the chemical shift of nucleus l in fragment A in its reference state and in the absence of other fragments. The latter represents the chemical shift changes of nucleus l with respect to the reference state. These shift changes can result (i) from changes in the local environment, $\delta_{l,lcA}$, that is, changes in fragment A with respect to its reference state, e.g. changes in torsion angle, bond length, bond angle etc.; or (ii) from changes in the interaction of nucleus l in A with the other molecular fragments B_j, $\delta_{l,j,B}$. For ^{1}H nuclei the conformational chemical shift of nucleus l in A does not depend strongly on the torsion angle changes in A, and the conformation-dependent shift can conveniently be attributed to interactions with other molecular fragments. As a result, δ_{intrin} of the ^{1}H nuclei in, for example, the sugar moieties of nucleic acids can be appropriately be defined as belonging to the individual C-H fragments. In contrast, for heteronuclei the chemical shift of a particular nucleus, l, may be quite strongly affected by torsion angle changes around this nucleus (see e.g. Szilagyi, 1996). For instance, for δ_{intrin} of a ^{13}C nucleus in ribofuranose it would be operationally more convenient to use as a reference state the S-puckered conformation of the sugar. The chemical shift, δ_l, of the resonance of nuclear spin l in molecule C can then formally be written as

$$\delta_l = \delta_{l,intrin} + \delta_{l,conf} = \delta_{l,intrin} + \delta_{l,lcA} + \sum_{j=1}^{N} \delta_{l,j,B} \qquad (1)$$

The conformation-dependent terms $\delta_{l,j,B}$ that result from interactions with other fragments can each be divided into a number of distinct contributions, δ_{rc}, δ_{ma}, δ_E, and δ_{CT}, so that

$$\delta_l = \delta_{l,intrin} + \delta_{l,lcA} + \sum_{j=1}^{N}\left[\delta_{l,j,rc} + \delta_{l,j,ma} + \delta_{l,j,E} + \delta_{l,j,CT}\right] \quad (2)$$

Together, δ_{rc} and δ_{ma} form the chemical shift variations resulting from magnetic anisotropy effects, with δ_{rc} being the chemical shift from ring current effects produced by aromatic rings, and δ_{ma} the chemical shift due to local magnetic anisotropy effects from, for example, an asymmetric electron distribution on atom B interacting with nucleus l of A. Analytical expressions containing adjustable parameters have been derived for the two magnetic anisotropy terms (they are discussed in more detail below). The parameters in the analytical expressions for δ_{rc} and δ_{ma} have been derived either from experimental shift data (e.g., Williamson and Asakura, 1993) or from fitting to quantum mechanically calculated isoshielding curves (Giessner-Prettre and Pullman, 1987). Note that the ab initio calculations do not allow separation of the two magnetic anisotropy terms, δ_{rc} and δ_{ma}, i.e. the quantum mechanically calculated isoshielding curves reflect the sum of the two terms δ_{rc} and δ_{ma} (Giessner-Prettre and Pullman, 1987). The analytical expressions for the ring current and magnetic anisotropy terms describe with good precision the isoshielding curves calculated from ab initio quantum mechanical calculations. They differ on average by not more than 0.07 ppm and at most 0.2 ppm (Giessner-Prettre and Pullman, 1987); the largest differences are found for the imino protons of guanine and the H2 protons of adenine. The polarization term, δ_E, is the chemical shift change resulting from polarization, by an electric field, of the electron density along the chemical bond extending from atom l. It can also be described by a simple analytical expression, the parameters being determined from comparison with experimental data (see below). The charge transfer term, δ_{CT}, has been introduced by Giessner-Prettre and Pullman (1987) to account for the effect of transfer of electron density on hydrogen bonding. The charge transfer term, δ_{CT}, is of relevance only when atom l is involved in hydrogen bonding.

In the chemical shift calculations the following considerations have been taken into account. Since only shifts are considered of non-exchangeable protons, i.e. protons not involved in hydrogen bonding, we have taken the terms $\delta_{l,lcA}$ and δ_{CT} to be zero. For δ_{rc} and δ_{ma} we have used, without adjustment, the ring and local magnetic anisotropy parameters derived by Giessner-Prettre and Pullman (1987). For δ_E the parameter values derived by Buckingham (1960) were used. For each proton, the calculated chemical shift is obtained by summing $\delta_{l,conf}$ over all surrounding rings, magnetically anisotropic groups, and all charges in the molecule. In case nucleus l concerns a base proton, the effect of the attached base is excluded. Furthermore, we did not include possible magnetic anisotropy terms resulting from the sugar ring atoms or from the phosphate backbone, since their effect has been shown by Giessner-Prettre and Pullman (1987) to be small. Finally, we note that the value of the term δ_{intrin} is not a priori known. It can be determined from experimental chemical shift data, as described below.

The best way to establish whether the calculated chemical shift value of a nucleus l, δ_l, is correct, is by comparing it with the corresponding observed experimental chemical shift value, δ_{exp}, of molecules of known conformation. A direct comparison is not possible, because the term δ_{intrin} is not known. However, since δ_{intrin} is conformation independent, it is possible to assess the validity of the calculations by comparing the conformation-dependent parts of the calculated and observed chemical shifts, that is, by comparing δ_{conf} and $\delta_{conf,exp}$, respectively. The term $\delta_{conf,exp}$ is given by

$$\delta_{conf,exp} = \delta_{exp} - \delta_{ref} \quad (3)$$

Here δ_{ref} is an experimentally determined reference value, which in fact is the experimental counterpart of δ_{intrin}. Ideally, one would like to obtain δ_{ref} from experimental data alone. At first glance, the random coil chemical shift, δ_{raco}, seems a good candidate. However, $\delta_{conf,exp}$ only reflects δ_{conf} correctly when the chemical shift resulting from interresidue contributions, $\delta_{conf} = \delta_{rc} + \delta_{ma} + \delta_E$, averages to zero for molecules in the random coil state. Otherwise a correction, δ_{const}, has to be introduced which accounts for the non-zero value of δ_{conf} in the random coil state,

$$\delta_{conf,exp} = \delta_{exp} - \delta_{raco} - \delta_{const} \quad (4)$$

This is the approach that has been used in studies on chemical shifts in proteins (Case et al., 1994). Ösapay and Case (1994) have indeed found that for proteins particular values of δ_{const} have to be used for different amino acid types in order to remove systematic errors. They subsequently demonstrated that these variations were due to residual δ_{conf} effects in the random coil state. Therefore, we have avoided the use of random coil chemical shifts, and derived instead the reference chemical shift from the experimental data by using it as an adjustable parameter, i.e., for each proton type one simply calculates δ_{ref} from

$$\delta_{ref} = \frac{1}{n}\sum \delta_{exp} - \delta_{conf} \quad (5)$$

where the sum is over all n protons of a certain type. In this way a set of reference values is obtained, one for each proton type. This approach has the added advantage that δ_{ref} thus obtained directly relates to the intrinsic chemical shift, δ_{intrin}.

The quality of the correspondence between calculated and observed chemical shifts is best expressed in the root-mean-square deviation, σ,

$$\sigma = (1/m)(\sum(\delta_{exp} - \delta_{ref} - \delta_{conf})^2)^{1/2} \quad (6)$$

where m is the total number of protons involved in the comparison. Henceforth, we will call the conformation-dependent chemical shift δ_{conf} and the calculated chemical shift δ_{calc}.

Ring-current shifts

Shifts introduced as a result of ring currents have been discussed extensively in the literature. The general form used for these shifts is

$$\delta_{rc} = iBG(\mathbf{r}) \quad (7)$$

where i is the ring-current intensity factor and B is a constant which has been adjusted empirically such that the ring-current intensity of benzene is unity. In this way the ring-current intensity factor, i, of the ring considered is equal to the ring-current intensity of this ring relative to that of benzene. G(**r**) is a geometric factor with **r** being the vector connecting the observed nucleus to the ring that generates the ring current. The two most popular methods to calculate this factor are those developed by Johnson and Bovey (1958) and by Haigh and Mallion (1980). The results generated by the two approaches are very similar. In the theory of Pullman and collaborators, the Johnson–Bovey method has been adopted. In this approach it is assumed that the ring-current shielding arises from two ring-current loops situated symmetrically at each side of and parallel to the plane of the aromatic molecule. If the electrons circulate in loops of radius a, the geometry factor is given by:

$$G(\rho,z) = \frac{1}{a}\frac{2}{\left[(1+\rho)^2+z_-^2\right]^{1/2}}\left(K(k)+\frac{1-\rho^2-z_-^2}{(1-\rho)^2+z_-^2}E(k)\right) + \frac{1}{a}\frac{2}{\left[(1+\rho)^2+z_+^2\right]^{1/2}}\left(K(k)+\frac{1-\rho^2-z_+^2}{(1-\rho)^2+z_+^2}E(k)\right) \quad (8)$$

Here, ρ and z are the dimensionless cylindrical coordinates relative to the ring center measured in units of a, $z_\pm = z \pm \langle z \rangle$, with $\langle z \rangle$ being the distance between the loop and the plane of the aromatic ring, and K and E are complete elliptic integrals of the first and second kind, respectively. The argument k is given by:

$$k = \left[\frac{4\rho}{(1+\rho)^2+z^2}\right] \quad (9)$$

We have used in the equations the parameters given by Ribas-Prado and Giessner-Prettre (1981). The ring-current loops are then taken to be above and below the aromatic plane at heights of 0.5770 and 0.5660 Å for cytosine/thymine and adenine/guanine, respectively. The radii of the current loops are 1.3675, 1.3790, 1.3610, 1.1540, 1.3430 and 1.1540 Å for cytosine, thymine, guanine-6, guanine-5, adenine-6, and adenine-5, respectively, with the numbers 5 and 6 indicating the five- and six-membered rings of the purine bases; the ring current intensities are taken to be 0.2750, 0.1120, 0.3, 0.6350, 0.9 and 0.66 for cytosine, thymine, guanine-6, guanine-5, adenine-6, and adenine-5, respectively. The constant B is an empirical parameter adjusted to a value of 2.13×10^{-6} Å, which is such that for benzene the ring-current shifts of its own protons equal −1.5 ppm when $\langle z \rangle = 0.61$ Å, and i = 1.

Local magnetic anisotropy

The chemical shift of atom B induced by the local magnetic anisotropy at atom A is given by:

$$\delta_{ma} = \frac{1}{3r^5}\sum_{\alpha\beta}\left(3r_\alpha r_\beta - r^2\delta_{\alpha\beta}\right)\left(1.967R_{\alpha\beta} - 5.368Q_{\alpha\beta}\right) \quad (10)$$

Here, r is the distance between atoms A and B, and α,β = x, y, z. Furthermore, $R_{\alpha\beta}$ and $Q_{\alpha\beta}$ are the diamagnetic and paramagnetic parts, respectively, of the αβ element of the magnetic susceptibility tensor of atom A. These contributions were calculated according to Ribas-Prado and Giessner-Prettre (1981).

Electric field effect

This effect has been introduced (Giessner-Prettre and Pullman, 1987) using the expression originally derived by Buckingham (1960):

$$\delta_E = AE_\parallel + BE^2 \quad (11)$$

Here, $E_\parallel$ is the projection of the electric field along the X-H bond at the proton considered. Application of different computational methods (Buckingham, 1960; Augspurger and Dijkstra, 1991) indicated that A may vary between 2.5 and 3.0×10^{-12} e.s.u.; for nucleic acids the value 2.9×10^{-12} e.s.u. has been chosen (Giessner-Prettre and Pullman, 1987). The value of $B = 0.74 \times 10^{-18}$ e.s.u. was taken from Giessner-Prettre and Pullman (1987). The electric field at the proton considered was derived, using Coulomb's law, as the vector sum of fields eminating from the electric monopoles (partial charges), q, predicted to be present at the different atoms:

$$\vec{E} = \sum_{j=1}^{N}\frac{q_j\vec{r}_j}{4\pi\varepsilon_0\varepsilon_r r_j^3} \quad (12)$$

The derivation of this electric field is complicated in two ways. A reliable calculation of the partial charges is required and an estimate of the dielectric constant over molecular distances is needed. To tackle the first problem, the derivation of the partial charges, we followed the approach taken in the construction of molecular force fields. We tried different sets of partial charges, but ulti-

mately we chose the partial charge set of the CHARMm force field, which is present as a parameter set in XPLOR (Brünger, 1992). In this set the net charge on the phosphate group (PO_4) is –0.32 e, which is distributed over the phosphorus atom (+1.2 e) and the oxygen atoms (–0.36 e for O3' and O5' and –0.4 e for O1 and O2). The choice of a net partial charge of –0.32 e on the phosphate group mimics the screening effect by counterion 'condensation'. This topic has been the subject of recent studies on the electrostatic properties of macro-ions, such as proteins or DNAs, in which modern methods to solve the Poisson–Boltzmann equation in the presence of a dielectric boundary were used. These Poisson–Boltzmann calculations demonstrate that in DNA the fraction of the phosphate charges neutralized by 'bound' counterions is 0.50 to 0.75 (Lamm and Pack, 1990; Pack et al., 1993). Thus, by choosing a partial charge of –0.32 e on the phosphates, as done in XPLOR, we do not overestimate the screening of backbone charges. A second parameter, which is not directly available in the calculation of the electric field, is the dielectric constant, ε_r. Following others we used a distance-dependent dielectric constant, $\varepsilon(r) = 4r$, where r is the distance. This approach is often used in molecular mechanics calculations to mimic solvent effects as well as counterion screening. In the equation a constant equal to 4 was introduced, because various studies have indicated that such a value is optimal (see e.g. Orozco et al., 1990; Kollman (1995) personal communication) and it has been used in an adaptation of the Weiner force field specifically designed for modelling nucleic acids (Weiner et al., 1986; Veal and Wilson, 1991; Kollman (1995) personal communication). Recent calculations of the variations of the dielectric constant around DNAs show that moving from the solvent closer to the DNA surface, the dielectric constant decreases from a value of 78.5 in bulk water to 40 near the DNA, while inside the minor and major grooves dielectric constants on the order of 10 to 25 may be found and within the DNA helix the dielectric constant is on the order of 2 to 10 (Pack et al., 1993; Lamm and Pack, 1997). We note that in their ^{1}H chemical shift calculations of proteins Williamson and Asakura (1993) have simply used a dielectric constant of 20 that is independent of distance. They find a rather small electric field effect. The present calculations on nucleic acids show (see below) that here the effect of charge (or electric field) on the ^{1}H chemical shift is also small compared to the changes generated by ring currents and magnetic anisotropy when the parameter set of XPLOR is used and a distance-dependent dielectric constant is introduced. Consequently, the calculated chemical shifts depend neither on the exact

TABLE 1
DATA USED IN THE CALCULATIONS[a]

Name	Sequence	Reference
Double Helix	**5'-3' one strand**	
Dickerson	CGCGAATTCGCG	[b]
CATGCATG	CATGCATG	Baleja et al., 1990a
GTACGTAC	GTACGTAC	Baleja et al., 1990a
GTATAATG	GTATAATC	Schmitz et al., 1992
Hiv	AGCTTGCCTTGAG	Mujeeb et al., 1992
HpaI	GCCGTTAACGGC	Kim and Reid, 1992
Lambda	TCTATCACCG	Baleja et al., 1990b
Hairpins	**5'-3'**	
ATCT	CGC-ATCT-GCG	Ippel, 1993
CACG	CG-CACG-CG	Ippel, 1993
CCCG	..T-CCCG-A..	Van Dongen et al., 1996
CTAG	CG-C^{ara}TAG-CG	Pieters et al., 1990
CTCG	CG-CTCG-GCG	Ippel, 1993
GTTA	ATCCTA-GTTA-TAGGAT	[c]
T3FA	GGATCG-TTT-CGATCC	Boulard et al., 1991
T3FX	TCTCTC-TTT-GAGAGA	Mooren et al., 1994
TTCA	CGC-TTCA-GCG	Ippel et al. 1993
TTTA	ATCCTA-TTTA-TAGGAT	Blommers et al., 1991
TTTT	ATCCTA-TTTT-TAGGAT	Hilbers et al., 1994
Other	**5'-3'**	
Thrombin[d]	GGTTGGTGTGGTTGG	Schulze et al., 1994
Cirt4[e]	-CGC-TT-GCG-TT-	Ippel et al., 1995

[a] Crystal structures or NMR structures were either taken from the Brookhaven Databank, or kindly provided by the authors.
[b] Dickerson and Drew (1981) and Wing et al. (1980): crystal structures from the Brookhaven Databank; NMR resonance. assignments from Hare et al. (1983).
[c] Van Dongen et al. (1997), see also Hilbers et al. (1994).
[d] Thrombin aptamer.
[e] Circular dumbbell.

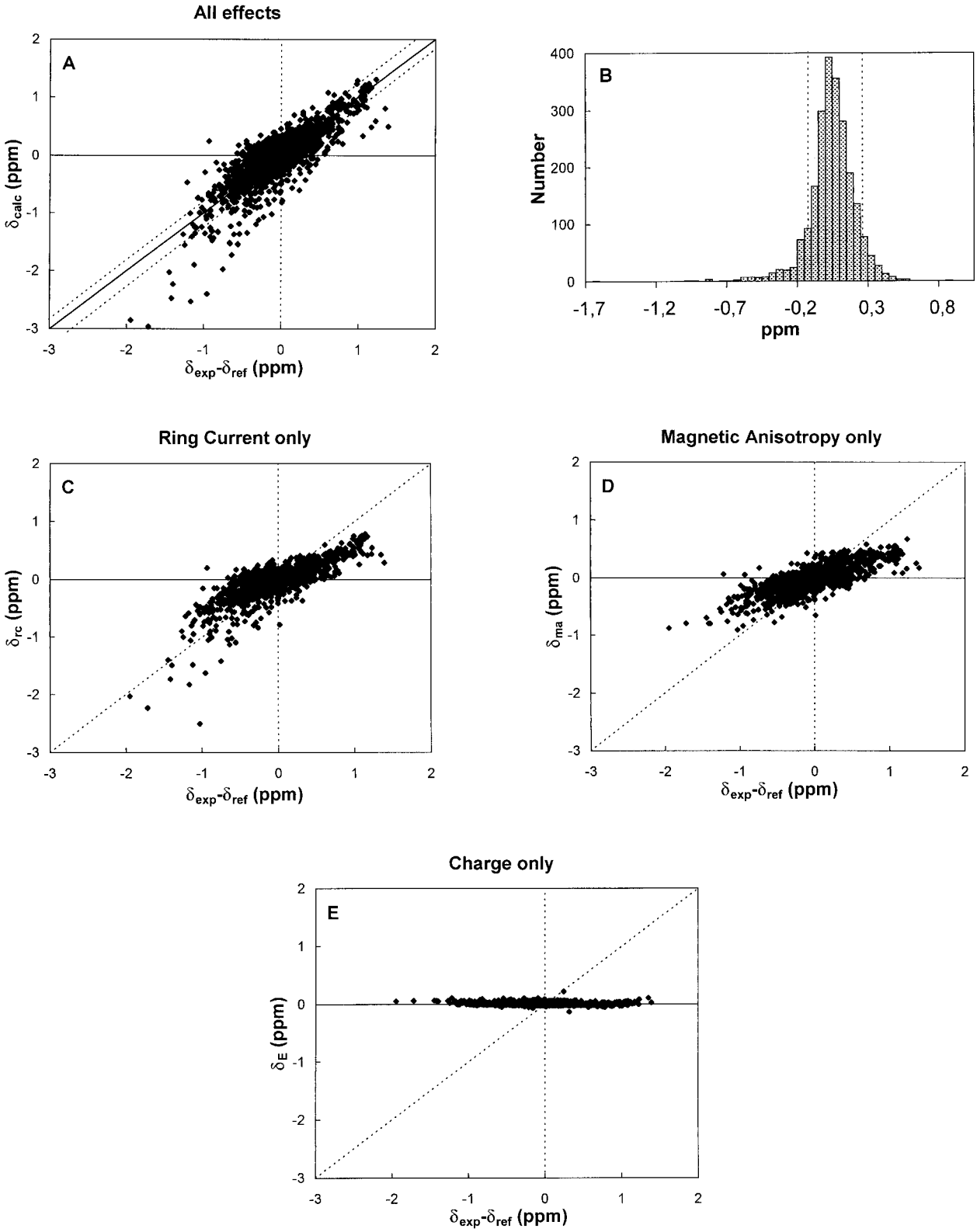

Fig. 1. Correlations between calculated shifts and observed structural shifts for all (2272) non-exchangeable protons. (A) 'All effects': calculated shift using the sum of the ring current, magnetic anisotropy, and electric field terms; the dashed lines run parallel to the diagonal at a distance of one standard deviation (0.17 ppm). (B) Distribution of the errors between calculated and observed structural shifts (see text); the vertical dashed lines indicate the standard deviation. (C) 'Ring-current only': calculated shift using the ring currents alone. (D) 'Magnetic anisotropy': calculated shift using magnetic anisotropy alone. (E) 'Charge only': calculated shift using electric field term alone (see text). For the ring current and magnetic anisotropy, the parameters given by Ribas Prado and Giessner-Prettre (1981) were used without adjustment; the same applies for the parameters for the electric field term (see text).

values of the partial charge nor on the exact value of the constant governing the distance dependence of the dielectric constant (see below).

Results and Discussion

Assessment of the reliability of the chemical shift calculations

The nucleic acid molecules used to generate the chemical shift database are summarized in Table 1. These structures were available at the start of this project either from the Brookhaven Databank, or provided by the authors, or from our own work. In view of the still limited number of structures available for RNA we chose to restrict the database to DNA molecules. The set of molecules considered consists of a number of double helices and hairpins, the thrombin-binding aptamer, which forms a unimolecular quadruplex in solution, and a circular dumbbell molecule. We have used the chemical shifts measured for all of the non-exchangeable protons of these molecules without making any corrections for differences in measuring conditions. The shifts were calculated, as described in the Materials and Methods section, on the basis of the structures available for the molecules listed in Table 1. When for a molecule more than one structure was available that fulfilled the NMR restraints, the average of the shifts calculated for a particular type of proton for each conformer was used. Also, for the thymine methyl protons one average of the shift value was used.

Plotted in Fig. 1A are the observed structural shifts, the experimental minus the reference shifts ($\delta_{exp} - \delta_{ref}$) as a function of the calculated chemical shift values, δ_{calc}. The data points scatter around the drawn diagonal line, which has slope 1; the dashed lines parallel to the diagonal represent the standard deviation of the shifts with respect to the diagonal. The horizontal drawn line indicates zero calculated shift and the vertical dashed line the situation that the experimental shift equals the reference shift.

In Fig. 1A, the data points strongly overlap and to obtain a better indication of their distribution all data were projected in a cross section through the point (0,0) perpendicular to the diagonal and plotted in the form of a histogram, see Fig. 1B. We obtain a close to Gaussian distribution with a standard deviation of 0.17 ppm, which is indicated in Fig. 1B by the vertical dashed lines. Thus, a quite narrow distribution is obtained, despite the fact that on the one hand the experimental shifts, δ_{exp}, were not corrected for differences in measuring conditions, such as differences in temperature or salt conditions, and on the other hand no parameters were optimized to derive the calculated shifts, δ_{calc}. The standard deviation of 0.17 ppm is quite small when compared with the region covered by the experimental data, i.e. $\delta_{exp} - \delta_{ref}$ covers a region of approximately 4 ppm (Fig. 1A). In fact, a better correspondence between calculated and observed chemical shifts is obtained than for the H^{α} chemical shifts in proteins, where a standard deviation of approximately 0.25 ppm is found (Williamson and Asakura, 1993; Case et al., 1994; Williamson et al., 1995). The contributions of the different terms to the total shift, i.e. the ring current effect, δ_{rc}, the magnetic anisotropy effect, δ_{ma}, and the charge effect, δ_{E}, are indicated in Figs. 1C, D and E, respectively. It turns out that on average the contributions of the ring-current and magnetic anisotropy effects to the total shift are about the same, while the contributions of the charge effects are practically negligible. Thus, in summary: (i) the observed chemical shift variations can be calculated with good accuracy, 0.17 ppm; (ii) the observed variations are determined by the sum of essentially two terms, namely the ring-current and magnetic anisotropy terms; and (iii) for these terms no parameter adjustments are required.

Although the correspondence between observed and calculated shifts is quite good, it is of interest to consider in more detail some of the outliers in Fig. 1. We first note that the cloud of data points has a somewhat rounded form. The largest deviations are found to the left in Fig. 1A. These points have a δ_{calc} that is too large negative due to an overestimation of the ring-current effects, as follows from a comparison of the contributions of the ring current, Fig. 1C, and of the magnetic anisotropy, Fig. 1D, to the total shift. This suggests that the parameter values for ring currents presently used may be slightly too large. Alternatively, the rounded form could also partially be a presentation artefact, originating from the inverse third power dependence on distance of the ring current and the magnetic anisotropy. Uncertainties in the distances will then lead to amplification of large values for δ_{calc}, and thus to the observed rounded form. The points deviating most on the left-hand side almost all concern H2 protons, for which few NOEs are available, i.e. these are the protons with the largest uncertainty in their position. Examination of a number of the outlying H2 chemical shift data points showed that relatively small adjustments in the position of the H2s, i.e. on the order of 0.1–0.2 Å, are sufficient to bring the data point back to the diagonal. Finally, Giessner-Prettre and collaborators have recently shown that for protons close to a ring the large ring-current shifts experienced by these protons may be damped by a dispersion effect (Giessner-Prettre et al., 1992). Such an effect becomes important for shifts larger than 1 ppm. Hence, incorporation of such a term in the present calculations would reduce the magnitude of the calculated chemical shift for the H2 protons. At this point it is difficult to distinguish between these possibilities. However, we do note that the fact that only small changes in position are required to bring the H2 proton chemical shift back to the diagonal has an important practical consequence. It implies that the error in the chemical shift translates into

TABLE 2
OVERVIEW OF STRUCTURAL CHEMICAL SHIFT EFFECTS[a,b]

	δ_{ref}				δ_{ib}				$\delta_{3'b}$				$\delta_{5'b}$			
	A	G	C	T	A	G	C	T	A	G	C	T	A	G	C	T
H1'	5.19 (20)	5.25 (20)	5.48 (24)	5.80 (26)	1.25	1.0	0.65	0.45	–0.6	–0.6	–0.3	–0.3	–	–	–	–
H2'	2.28 (27)	2.28 (27)	2.28 (27)	2.28 (27)	0.3	0.3	–0.2	–0.2	–	–	–	–	–	–	–	–
H2"	2.47 (22)	2.47 (22)	2.47 (22)	2.47 (22)	0.3	0.2	0.1	0.0	–	–	–	–	–	–	–	–
H3'	4.72 (12)	4.72 (12)	4.72 (12)	4.72 (12)	0.2	0.2	0.0	0.0	–	–	–	–	–	–	–	–
H4'	4.17 (22)	4.17 (22)	4.17 (22)	4.17 (22)	0.1	0.1	0.0	0.0	–	–	–	–	–	–	–	–
H5'	4.02 (16)	4.02 (16)	4.02 (16)	4.02 (16)	0.15	0.15	0.0	0.0	–	–	–	–	–	–	–	–
H5"	3.98 (16)	3.98 (16)	3.98 (16)	3.98 (16)	0.15	0.15	0.0	0.0	–	–	–	–	–	–	–	–
H8	8.50 (20) 8.2	8.11 (23) 7.9	–	–	–	–	–	–	[b]	[b]	[b]	[b]	[b]	[b]	[b]	[b]
H6	–	–	7.70 (27) 7.7	7.64 (23) 7.5	–	–	–	–	[b]	[b]	[b]	[b]	[b]	[b]	[b]	[b]
H5 (C)	–	–	6.21 (33) 5.9	–	–	–	–	–	–	–	–	–	–0.8	–0.8	–0.5	–0.5
H5 (T)	–	–	–	2.00 (28) 1.8	–	–	–	–	–	–	–	–	–0.5	–0.5	–0.3	–0.3
H2	8.60 (30) 8.1	–	–	–	–	–	–	–	–1.5	–1.0	–0.8	–0.8	–	–	–	–

[a] The structural chemical shift effects (see text, Eq. 13) are: δ_{ib}, shift from own base, glycosidic torsion angle $\chi = 240°$; $\delta_{3'b}$, shift from 3'-neighboring base; $\delta_{5'b}$, shift from 5'-neighboring base; the values given apply to a double helical environment. Reference shift δ_{ref} with standard deviation between parentheses expressed in terms of the last two digits; for the base protons the second value indicates the random coil shift.

[b] The chemical shifts of the base protons H6 and H8 are influenced by both the 5'- and 3'-neighboring bases; to estimate their effect, the sum of $\delta_{3'b}$ and $\delta_{5'b}$ has been taken, giving for the different base sequence combinations: δ_{exp} of Py-N-Py equals 8.3 ppm (N=A), 7.9 ppm (N=G), 7.55 ppm (N=C), and 7.44 ppm (N=T), so that $\delta_{3'b} + \delta_{5'b}$ equals –0.15–0.2 ppm for N=A, G, C, and T; δ_{exp} of Pu-N-Pu equals 8.05 ppm (N=A), 7.65 ppm (N=G), 7.25 ppm (N=C), and 7.15 ppm (N=T), so that $\delta_{3'b} + \delta_{5'b}$ equals –0.45–0.5 ppm for N=A, G, C, and T; δ_{exp} of Pu-N-Py equals 8.15 ppm (N=A), 7.75 ppm (N=G), 7.35 ppm (N=C), and 7.25 ppm (N=T), so that $\delta_{3'b} + \delta_{5'b}$ equals –0.35–0.4 ppm for N=A, G, C, and T; δ_{exp} of Pu-N-Py equals 8.15 ppm (N=A), 7.8 ppm (N=G), 7.42 ppm (N=C), and 7.3 ppm (N=T), so that $\delta_{3'b} + \delta_{5'b}$ equals –0.30–0.35 ppm for N=A, G, C, and T.

a rather small error in the spatial position, reminiscent of the NOE to distance relationship. On the other hand, on the right in Fig. 1A the deviations are smaller. As will be discussed later in more detail in connection with H2' shifts, these shifts are determined mainly by the magnetic anisotropy term. This suggests that the parameters for the magnetic anisotropy term are essentially correct.

As can be seen from Fig. 1E, the contribution from δ_E is quite small. To calculate the electric field effects we have used approaches that are popular in molecular mechanics and molecular dynamics calculations to mimic solvent effects (Orozco et al., 1990; Veal and Wilson, 1991; Brünger, 1992; Kollman (1995) personal communication). Thus, we used the charges provided as a parameter set in XPLOR, which includes the use of reduced charges on the phosphate oxygens and application of a distance-dependent dielectric constant with $\varepsilon(r) = 4r$ to mimic the solvent (see Materials and Methods). In a rather crude way the effect is accounted for that over longer distances the dielectric constant is on average larger, and it affords incorporation of a screening effect of the charges that follows from the Boltzmann distribution of the counterions. Although some uncertainty exists as to what are the best values for the partial charges, and what distance dependence should be used for the dielectric constant, the small values of δ_E as compared to those of δ_{rc} and δ_{ma}

ensure that small variations in the calculated electric field will not significantly alter the ultimate results for δ_{calc}. The charge effects also play a minor role in determining the H^{α} chemical shifts in proteins (Williamson and Asakura, 1993). In the calculations usually a dielectric constant of 20 is introduced (Williamson and Asakura, 1993). If such a value had been used in the present calculations, an even smaller electric field effect would have been found.

In contrast to nucleic acids, the chemical shift of the H^{α} protons in proteins is mainly determined by the anisotropy of the carbonyl group in the backbone, while the shift contributions from aromatic rings play a rather minor role (Williamson and Asakura, 1993). Of course, this difference in the ring-current contributions in proteins and nucleic acids arises from the fact that every nucleotide contains an 'aromatic' residue, while the abundance of aromatic rings in proteins is relatively low.

We finally note that magnetic anisotropies of sugar ring residues and backbone residues have not been included in the calculations, in view of the finding by Giessner-Prettre and Pullman (1987) that these anisotropies are small and do not significantly influence the chemical shifts.

The reference chemical shift was obtained by fitting the line with slope 1 to the experimental data in a least-squares approximation according to Eq. 5 (see Materials and Methods). It turned out to be necessary for the H1' protons to distinguish between the different nucleotides they belong to. This is demonstrated in Fig. 2. In Fig. 2A, the H1' signals of the four different nucleotides are indicated by differently colored data points. As can be seen, they cluster around different lines with slope 1. If for the four nucleotides these lines are fitted to the data points, different reference shifts are found. Subsequent plotting of the calculated values against the experimental minus the reference shifts then yields a narrower distribution (see Fig. 2B). The reference shifts, for H1' as well as for the other non-exchangeable protons, are summarized in Table 2. For the non-sugar protons the random coil shifts are also listed. The latter turn out to be equal to or somewhat lower than the values derived for the reference shifts (with the exception of the results obtained for the adenine-H2 shifts). This difference between the reference chemical shift and the random coil values is expected, since in the random coil state an effect is expected on average to exist from neighboring residues. Similar findings have been reported by Ösapay and Case for proteins (1994).

Structural origins of the chemical shifts

H2′ resonances

At this stage it is interesting to examine some points that are shifted away from the central vertical line in Fig. 1A. To this end, we first consider the distribution of the H2' signals presented in Fig. 3. It is clear that on the one hand the signals of the pyrimidines and on the other hand the signals of the purines strongly cluster. This has also been observed in an earlier study of a much larger database of chemical shifts (Van de Ven and Hilbers, 1988; Wijmenga et al., 1993), where it was found that the center of the shift distribution of the pyrimidines is located at 2.0 ppm and that of the purines at 2.6 ppm, with very little overlap between them. In Fig. 3 the distributions are also so neatly separated that in a first approximation they can even be used for assignment purposes. In the same study it was found that the positions of H2' signals from double helices are predominantly determined by the base of the nucleotide itself. The physical reason for this may be gleaned from Fig. 4B, where the shift of the H2' signals has been plotted as a function of the glycosidic χ-angle. For the four different nucleotides the behavior of the H2' resonance positions as a function of χ is qualitatively similar, but there are interesting differences. For instance, when χ is around 230°, i.e. in the *anti* region normally found in B-type double helices, small shifts, on the order of 0 to 0.3 ppm, are expected but in opposite directions for purines and pyrimidines. In agreement with this observation, the H2' resonances cluster at small negative ($\delta_{exp} - \delta_{ref}$) values for pyrimidines and at small positive ($\delta_{exp} - \delta_{ref}$) values for purines (viz. Figs. 3 and 4B). Thus, the theory provides a neat qualitative description of the experimental results. From this statement the resonances below –0.5 and above +0.5 ppm are excluded; these are discussed below.

A remarkable aspect of the distribution of the H2' resonances as found here and also earlier in the study by Van de Ven and Hilbers (1988), is the independence of the H2' shift of the 3' neighboring base in a double helix environment. The H2' resonance is positioned rather close to this base and one would expect an effect on its chemical shift by the ring current of this base. Careful consideration of the exact position of the H2' residue with respect to the 3' neighboring base in a regular double helix environment shows, however, that although the H2' is close to the base, it is positioned in the node separating upfield and downfield shifts. This nicely explains the independence of the H2' shift on the 3' neighboring base. It does suggest that the H2' should be rather sensitive to deviations from regular double helix conformation.

The data points in Fig. 3 which are strongly shifted to the right-hand side arise from nucleotides with the adenine and guanine bases in a *syn* orientation. This is in agreement with the calculations presented in Fig. 4B. The curves computed for the H2' resonances clearly show that there is a large increase in chemical shift when the bases are turned into the region between 60° and 150° (which largely coincides with the *syn* and high-*syn* region; Saenger, 1984). In particular when χ is between 60° and 90° a modest change in the glycosidic angle may introduce appreciable shifts. Thus, H2' shifts calculated for nucleotides with, for instance, a guanine base in this orientation may be particularly sensitive to uncertainties in the struc-

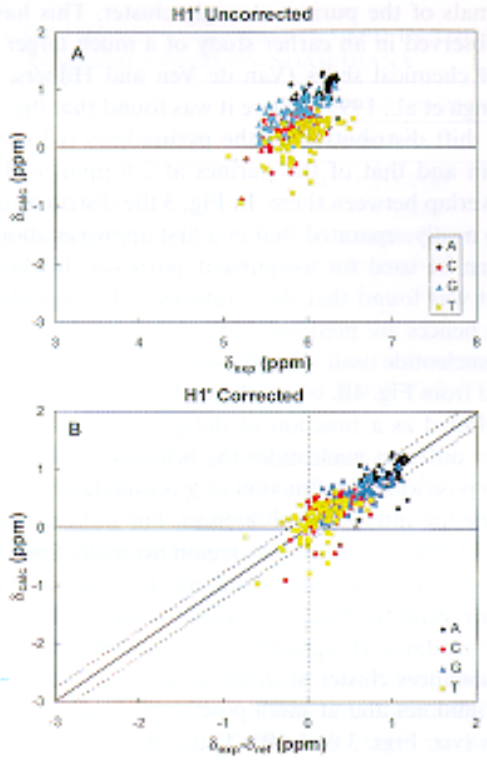

Fig. 2. (A) 'H1' uncorrected': correlations between calculated shifts ('all effects', see Fig. 1) and observed shifts for all H1' protons. (B) 'H1' corrected': correlation between calculated shifts ('all effects') and observed structural shifts, i.e. observed shift corrected for differences in reference shift (see text). The H1' data points are colour coded according to residue type.

ture. An example is provided by the H2' resonance of G10 of the thrombin aptamer studied by Feigon and collaborators (Schulze et al., 1994); the theoretical/empirical correlation corresponding with this resonance is outside the region enclosed by the two standard deviation lines (see Fig. 3). The data point corresponds with the average chemical shift of a set of structures which fulfil the NMR restraints. Figure 4B shows that turning the χ-angle corresponding with the average chemical shift (χ = 70°) by 20°, to the angle found for the structure with the lowest energy state (Schulze et al., 1994), would shift the data point to the expected value on the diagonal. The chemical shift distribution obtained from the set of available structures for other guanine bases with a *syn* orientation, e.g. G5 and G14, is narrow and falls within the region encircled around the corresponding correlation points (Fig. 3). Consideration of other points at the right-hand side of Fig. 3, deviating significantly from the expected values, shows that similar uncertainties, as for G10 in the thrombin aptamer, are present in the corresponding structures.

We now turn to the left-hand side of Fig. 3, where the pyrimidine H2' resonances are clustered. The H2' resonance of residue C7 of the circular dumbbell molecule studied by Altona and collaborators (Ippel et al., 1992, 1995) is located entirely outside the region confined by the standard deviation lines. Its experimental shift is strongly shifted to the left, indicating that it feels a strong

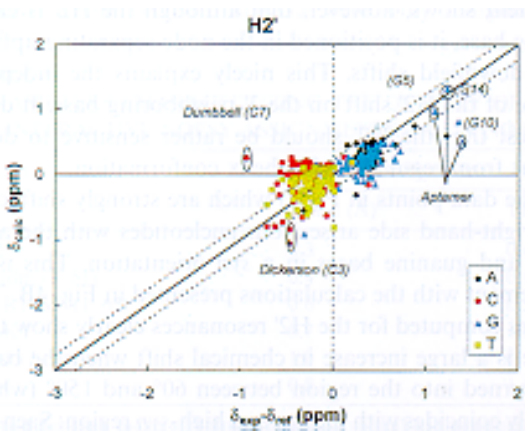

Fig. 3. Correlations between calculated shifts ('all effects', see Fig. 1) and observed structural shifts for all H2' protons. The H2' data points are colour coded according to residue type. The dashed lines run parallel to and at one standard deviation from the diagonal. For a number of outlying data points their origin in the structure data set is indicated (see text).

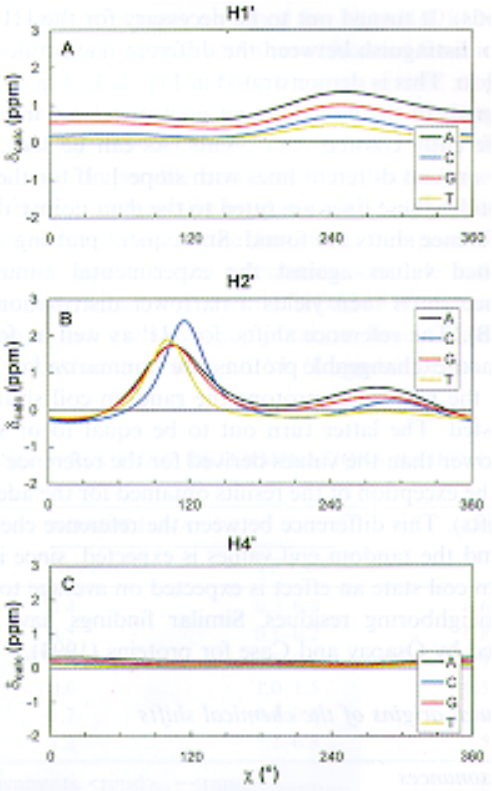

Fig. 4. Calculated shifts ('all effects, see Fig. 1) as a function of the glycosidic torsion angle χ for a mononucleotide with the sugar ring in an S-puckered state; the mononucleotide was taken from the Dickerson dodecamer. (A) Calculated shifts of the H1' protons. (B) Calculated shifts of the H2' protons. (C) Calculated shifts of the H4' protons. The curves are colour coded according to residue type.

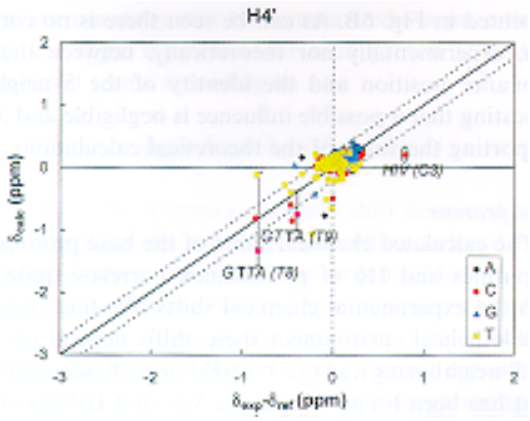

Fig. 5. Correlations between calculated shifts ('all effects', see Fig. 1) and observed structural shifts for all H4' protons. The H4' data points are colour coded according to residue type. The dashed lines running parallel to the diagonal lie at one standard deviation. For a number of outlying data points their origin in the structure data set is indicated (see text).

downfield ring-current shift. On the other hand, the calculated value is close to zero, which is expected in a regular helical environment, where the ring-current effects from neighboring residues are small. In the model used for the calculation, residue C7 is indeed stacked in a regular fashion in-between two guanine residues. In other words, the 3'-neighboring guanine is positioned as in a regular double helix. The present chemical shift calculations, however, show that this cannot be correct. In this circular dumbbell molecule the helical stem is extremely short and in fact consists only of one GC base pair, namely G2•C7. It is capped on one side by a 5'-CTTG- loop and on the other side by a 5'-GTTC- loop. In the former loop a regular C•G base pair is formed, whereas in the latter loop the G•C pair is non-canonical part of the time, with the C residue in a *syn* orientation. Thus, this irregular base pair may, as discussed before for the H2' resonances, easily lead to significant upfield shifts. In fact, the chemical shift calculations show that a modification of the structural model is required; in this case the chemical shifts can be used as constraints to refine the model.

The two resonances in Fig. 3, designated 'Dickerson C3', arise from the so-called double helical Dickerson dodecamer (Wing et al., 1980; Dickerson and Drew, 1981). This double helix is formed by selfcomplementary strands and is therefore expected to have a twofold axis of symmetry. Indeed, as can be derived from the NMR spectrum, the solution structure conforms to this symmetry. Apparently the X-ray structure (due to crystal forces) does not, because the calculated resonance positions of the two symmetry-related C3-H2' protons do not coincide. These differences between the solution and the crystal structure – the latter has been used for the derivation of the theoretical resonance positions – may well shift the calculated resonance positions outside the region confined by the standard deviations. Examination of the other data points outside this region indicates that uncertainties in the structure used to compute their location in Fig. 3 may explain the deviation from the expected values.

H4' resonances

In contrast to the H2' resonances, the positions of the H4' resonances are very modestly influenced by their own base (Fig. 5). In double helices, the same applies also to the neighboring nucleotides. This observation is directly understood when considering Fig. 4C, where the calculated shifts are plotted as a function of the glycosidic angle, χ, for the four different nucleotides. The shifts show only a slight dependence on χ. The H4' positions of adenine and guanine are affected most and this is reflected in the plot of δ_{calc} versus $(\delta_{exp} - \delta_{ref})$ (Fig. 4C). The correlations from the pyrimidines cluster around the (0,0)-coordinate and those from the purines are shifted somewhat to the right, as expected. We can only reconcile the location of the data point in the right-most position, arising from the conserved HIV-1 sequence (see Table 1), if the corresponding resonance has been misassigned or a typo-

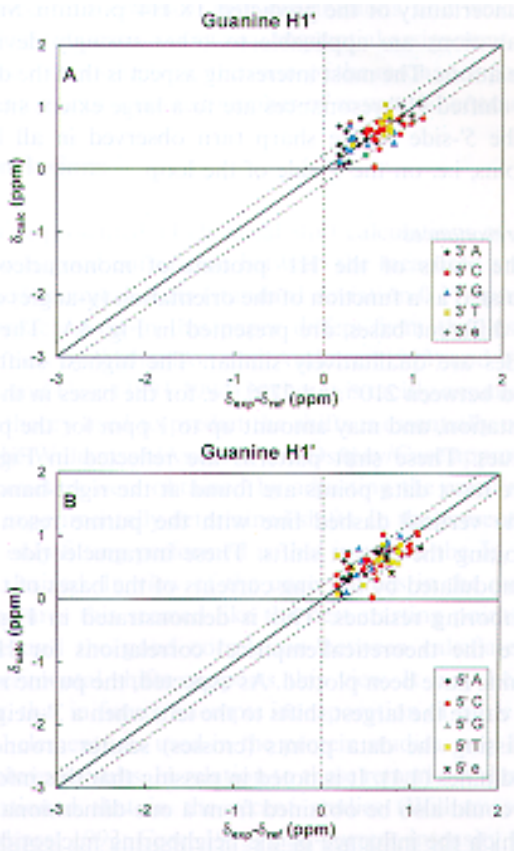

Fig. 6. Correlations between calculated shifts ('all effects', see Fig. 1) and observed structural shifts for all guanine H1' protons. Chemical shift data points are colour coded according to (A) the 3'-neighboring residue type and (B) the 5'-neighboring residue type. The dashed lines running parallel to the diagonal lie at one standard deviation. The labels 3'e and 5'e refer to 3'- and 5'-terminal residues, respectively.

graphical error has crept into the presentation of its resonance position (Schmitz et al., 1992). The data points shifted out further to the left all arise from residues located in loop regions and/or sharp turns, where neighboring bases induce large ring-current shifts (see Table 1). Here the positions of the points falling outside the region enclosed by the standard deviation lines can also be understood if the uncertainties of the structures used for the determination of δ_{calc} are taken into account. Thus, for the hairpin with the -GTTA- loop the H4' signals of its thymine residues are, based on the structures fulfilling the NMR restraints, predicted to scatter between the boundaries of the indicated bars. In this hairpin the T8 residue is situated above the sheared G•A pair formed in the loop, with the thymine base stacking on this base pair. The large downfield shift of the H4' resonance, which corresponds with the large negative ($\delta_{exp} - \delta_{ref}$) value of this proton in Fig. 5, arises from the large ring-current shift generated by the G•A base pair. In the determination of the hairpin loop structure, the hydrogen bonds formed between the G and A bases were not used as constraints, resulting in some uncertainty in the positioning of the bases with respect to one another and thus in the uncertainty of the predicted T8 H4' position. Similar explanations are applicable to other strongly deviating correlations. The most interesting aspect is that the downfield shifted H4' resonances are to a large extent situated on the 5'-side of the sharp turn observed in all DNA hairpins, i.e. on the 3'-side of the loop.

H1′ resonances

The shifts of the H1' protons of mononucleotides, calculated as a function of the orientation (χ-angle) of the four different bases, are presented in Fig. 4A. The shift profiles are qualitatively similar. The highest shifts are found between 210° and 270°, i.e. for the bases in the *anti* orientation, and may amount up to 1 ppm for the purine residues. These shift patterns are reflected in Fig. 2B, where most data points are found at the right-hand side of the vertical dashed line with the purine resonances exhibiting the largest shifts. These intranucleotide shifts are modulated by the ring currents of the bases of the 3'-neighboring residues. This is demonstrated in Fig. 6A, where the theoretical/empirical correlations for H1' of guanine have been plotted. As expected, the purine neighbors cause the largest shifts to the left; when a 3'-neighbor is missing the data points (crosses) scatter around the coordinates (1,1). It is noted in passing that this information could also be obtained from a one-dimensional plot in which the influence of the neighboring nucleotides on the positions of the guanine H1' resonances is considered. The point to make here, however, is that this distribution is correctly predicted by the theoretical calculations. A similar plot, but now reflecting the possible influence of 5'-neighbors on the H1' resonance positions of guanine, is presented in Fig. 6B. As can be seen there is no correlation, experimentally nor theoretically, between the H1' resonance position and the identity of the 5'-neighbor, indicating that a possible influence is negligible and again supporting the value of the theoretical calculations.

Base protons

The calculated chemical shifts of the base protons, H8 of purines and H6 of pyrimidines, correlate quite well with the experimental chemical shifts. We find that in a double helical environment these shifts depend on both the 5'-neighboring and the 3'-neighboring bases, similar to what has been found by Van de Ven and Hilbers (1988) with a larger database. We observe, as they did, that one can distinguish between Pu-N-Pu and Py-N-Py profiles, which have averaged downfield shifts with respect to δ_{ref} of 0.4–0.5 ppm and 0.15–0.2 ppm, respectively, while for Pu-N-Py and Py-N-Pu intermediate values are found. On the other hand, both the H5 protons of cytosine and the methyl protons of thymine depend quite distinctly on the 5'-neighboring base. For cytosine H5 protons one finds on average a downfield shift of approximately 0.5 ppm for a 3'-Py neighbor and a downfield shift of 0.8 ppm for a 5'-Pu neighbor. For the methyl protons the distributions overlap somewhat. The average downfield shifts are approximately 0.25 and 0.5 ppm for the 5'-Py and 5'-Pu neighbors, respectively. These shifts can quite nicely be understood if one considers the stacking patterns in a regular double helix. The H5 and the methyl protons of the Py are positioned above the plane of the 5'-neighbor, in contrast to the H6 and H8 protons, which are positioned in such a way that they feel the effect of both the 3'-neighboring as well as the 5'-neighboring base. Finally, in a regular helix the H2 protons are positioned above the plane of the 3'-neighbor, and indeed a 3'-neighbor effect is experimentally observed. We note that, as for the H2 protons, also for the H5 and methyl protons for the larger observed shift deviations the calculated shift is overestimated, an effect that mainly stems from the ring-current contribution. This again suggests that the unadjusted ring-current strengths used here are somewhat too large.

Summary of structure dependencies

In summary, for all sugar protons the chemical shift is mostly affected by their own base. The effect decreases in magnitude going from H1', H2', and H2" via H3' to H4' and H5' and H5" (see Fig. 4). Furthermore, in a double helix sequential effects are essentially absent for all sugar protons, except for H1'. For the latter it is the 3'-neighboring base that exerts the most influence. Furthermore, the χ-angle dependence is the strongest for H2' resonances (when the sugar is in an S-type conformation). An upfield shift always corresponds to a χ-*syn* torsion angle.

The most important dependencies of the shifts of the different non-exchangeable protons in a nucleotide on

their surroundings can be summarized via an estimated chemical shift, δ_{est}, which for each proton is split in a number of terms according to

$$\delta_{est} = \delta_{ref} + \delta_{ib} + \delta_{3'b} + \delta_{5'b} \quad (13)$$

Here, δ_{ref} is the reference chemical shift discussed earlier. The term δ_{ib} is the chemical shift effect induced by the own base when the χ-angle is around 240°. We have taken this value from the χ-angle dependencies of δ_{calc}, where only the effect of the own base is calculated (see for example Fig. 4). The terms $\delta_{3'b}$ and $\delta_{5'b}$ represent the chemical shifts induced by the 3'- or 5'-neighboring base, respectively. Their values are estimated from the offset of $\delta_{exp} - \delta_{ref} - \delta_{ib}$ from zero for each type of neighbor, using δ_{calc} versus $\delta_{exp} - \delta_{ref}$ plots such as the one shown in Fig. 6. The values for all these terms are summarized in Table 2. Thus, given these values, δ_{est} can easily be calculated. For example, the estimated chemical shift for an H2' proton of a G residue in a double helix is $\delta_{est} = 2.28 + 0.3 = 2.58$ ppm, while the H1' of this residue when having a 3'-cytosine neighbor will have $\delta_{est} = 5.25 + 1.0 - 0.3 = 5.95$ ppm. Thus, Table 2, together with the equation for δ_{est}, serves as a fast guide to estimate the proton chemical shifts in a double helical environment. It is emphasized here that the terms δ_{ib}, $\delta_{3'b}$, and $\delta_{5'b}$, are only valid for double helical structures. As we have seen, neighboring nucleotides may have very strong effects on the resonance positions of the protons of the considered nucleotide when involved in sharp turns or otherwise aberrant conformations. Thus, Table 2 may also serve as a fast guide to trace the origin of unusual shifts.

An interesting spin-off of Table 2 is that it is possible to perform sequential assignments, using chemical shifts alone, via the following protocol.

(1) From COSY or TOCSY experiments, establish that the sugar protons H1', H2', H2" and H3' belong to one and the same residue.

(2) Since the chemical shift values of H2', H2", and H3' depend on their own base alone (see Table 2), it can be established that a sugar moiety is either attached to a purine (R) or to a pyrimidine (Y).

(3) Using the H1' resonance positions, it can subsequently be established that the sugar residue has either a purine or a pyrimidine on the 3'-side or no 3'-neighbor, i.e. it can be established that a sugar moiety is a part of one of the following doublets, 5'**Y**Y, 5'**Y**R, 5'**R**Y, 5'**R**R, 5'**R**- or 5'**Y**-.

(4) Given the possibility to connect the sugar moiety resonances to the resonances of their own base, one can subsequently use the 5'-dependence of H5 and/or T(CH_3) resonances to further specify the surroundings for the pyrimidine sequences, **Y**Y, **Y**R, and **Y**-, i.e. one can discriminate between Y**Y**Y, R**Y**Y, Y**Y**R, R**Y**R, -**Y**Y, and -**Y**R. Thus, one can distinguish between the following triplets, Y**Y**Y, R**Y**Y, Y**Y**R, R**Y**R, (Y/R/-)**R**Y, and (Y/R/-)**R**R. With this knowledge available, the sequential assignment becomes rather simple. As an example we consider the GTTA-hairpin, which has the sequence 5'-ATCCTA-GTTA-TAGGAT (Van Dongen et al., 1997). The following triplets are present: -**R**Y, R**Y**Y, Y**Y**Y, Y**Y**Y, Y**Y**R, Y**R**R, ..., R**Y**R, Y**R**R, R**R**R, R**R**R, R**R**Y, and R**Y**-. Thus, in terms of distinguishable sequences as outlined above we have: one R**Y**Y sequence (can be placed unambiguously); two Y**Y**Y (thus, two positions possible, but each of them lies next to a triplet that can be placed unambiguously; hence each can in fact be placed unambiguously); one Y**Y**R (can be placed unambiguously); five (Y/R/-)**R**R (the two triplets that contain a Y are next to a triplet that can be placed directly, and the two others are next to an already indirectly placed R**R**Y or Y**R**R triplet); one R**Y**R (can be directly placed); one R**R**Y (can be directly placed); one R**Y**- (can be directly placed). Thus, five triplets can be assigned directly; the triplets (Y**Y**Y) can be placed in two positions and the (Y/R/-)**R**R triplets can be placed in five possible positions based on chemical shift data alone, but here one sequential contact allows each of them to be placed uniquely in the sequence, since each borders an already positioned triplet. This demonstrates that on the basis of these simple rules, which are derived from chemical shift distributions, and which have a good foundation in theory, sequential assignment can be made rather simply on the basis of the resonance positions together with a limited number of sequential contacts.

Conclusions

The presented ^{1}H chemical shift calculations of DNAs show a good correlation between the measured and predicted chemical shifts (standard deviation 0.17 ppm). The approach used in this paper differs from that used for proteins by Williamson and Asakura (1993) and by Ösapay and Case (1991,1994) in that it is tacitly assumed that the theory used to predict the shifts is essentially correct. In the Williamson/Asakura and Ösapay/Case approach all parameters were obtained by adjusting the calculated to the experimentally determined shifts. In the present paper the only fitting performed is that in which the δ_{ref} values are derived. In view of the uncertainties in the available structures this seemed like the best starting point at the time and the good correlation between calculated and experimental shifts supports this view. It may be, however, that in future attempts incorporation of more statistical elements, as used in the protein studies, may lead to improved results. In relation to these remarks it should be mentioned that in the recent studies (Williamson and Asakura, 1993; Case, 1995) ring-current intensity factors have been obtained that differ from, i.e. are higher than, those introduced by Giessner-Prettre and Pullman (1987). A reason could be that Giessner-Prettre and Pullman separately account for local anisotropy effects in the rings. This has not been done in the recent studies and

therefore there the introduction of larger ring-current intensity factors is required.

In the previous sections we examined a number of representative examples of theoretical/experimental shift correlations which deviate significantly from the expected values. It turned out that in the majority of cases these effects arise from uncertainties in the structures used to derive the theoretical shifts or from structures for which we had reasons to assume that they deviate (to some extent) from that really present in solution. We expect that in these cases the shift calculations are sufficiently reliable that they may be used to improve or refine currently available structures. Incorporation of chemical shift constraints into XPLOR is underway to investigate these aspects.

The computer program that performs the chemical shift calculations (NUCHEMICS) is available on request from the authors.

References

Asakura, T., Niizawa, Y. and Williamson, M.P. (1992) *J. Magn. Reson.*, **98**, 646–653.

Asakura, T., Taoka, K., Demura, M. and Williamson, M.P. (1995) *J. Biomol. NMR*, **6**, 227–236.

Augspurger, J.D. and Dijkstra, C.E. (1991) *J. Phys. Chem.*, **95**, 9230–9238.

Baleja, J.D., Germann, M.W., Van de Sande, J.H. and Sykes, B.D. (1990a) *J. Mol. Biol.*, **215**, 411–428.

Baleja, J.D., Pon, R.T. and Sykes, B.D. (1990b) *Biochemistry*, **29**, 4828–4839.

Blommers, M.J.J., Van de Ven, F.J.M., Van der Marel, G.A., Van Boom, J.H. and Hilbers, C.W. (1991) *Eur. J. Biochem.*, **201**, 33–51.

Boulard, Y., Gabarro-Arpa, J., Cognet, J.A.M., Le Bret, M., Guy, A., Teoule, R., Guschlbauer, W. and Fazakerley, G.V. (1991) *Nucleic Acids Res.*, **19**, 5159–5167.

Brünger, A. (1992) XPLOR, A system for X-ray crystallography and NMR, Yale University Press, New Haven, CT, U.S.A.

Buckingham, A.D. (1960) *Can. J. Chem.*, **38**, 300–307.

Case, D.A., Dyson, H.J. and Wright, P.E. (1994) *Methods Enzymol.*, **239**, 392–416.

Case, D.A. (1995) *J. Biomol. NMR*, **6**, 341–346.

Celda, B., Biamonti, C., Arnau, M.J., Tejero, R. and Montelione, G.T. (1995) *J. Biomol. NMR*, **5**, 161–172.

Dickerson, R.E. and Drew, H.R. (1981) *J. Mol. Biol.*, **149**, 761–786.

de Dios, A.C., Pearson, J.G. and Oldfield, E. (1993) *Science*, **260**, 1491–1496.

Ghose, R., Marino, J.P., Wiberg, K.B. and Prestegard, J.H. (1994) *J. Am. Chem. Soc.*, **116**, 8827–8828.

Giessner-Prettre, C. and Pullman, B. (1987) *Q. Rev. Biophys.*, **20**, 113–172.

Giessner-Prettre, C., Gresh, N. and Maddaluno, J. (1992) *J. Magn. Reson.*, **99**, 605–610.

Greene, K.L., Wang, Y. and Live, D. (1995) *J. Biomol. NMR*, **5**, 333–338.

Haigh, C.W. and Mallion, R.B. (1980) *Prog. NMR Spectrosc.*, **13**, 303–344.

Hare, D.R., Wemmer, D.E., Chou, S.H., Drobny, G. and Reid, B.R. (1983) *J. Mol. Biol.*, **171**, 319–336.

Hilbers, C.W., Heus, H.A., Van Dongen, M.J.P. and Wijmenga, S.S. (1994) In *Nucleic Acids and Molecular Biology*, Vol. 8 (Eds, Eckstein, F. and Lilley, D.M.J.), Springer, Berlin, Germany, pp. 56–104.

Ippel, J.H., Lanzotti, V., Galeone, A., Mayol, L., Van den Boogaart, J.E., Pikkemaat, J.A. and Altona, C. (1992) *J. Biomol. Struct. Dyn.*, **9**, 821–836.

Ippel, J.H. (1993) Ph.D Thesis, University Leiden, Leiden, The Netherlands.

Ippel, J.H., Lanzotti, V., Galeone, A., Mayol, L., Van den Boogaart, J.E., Pikkemaat, J.A. and Altona, C. (1995) *J. Biomol. NMR*, **6**, 403–422.

Johnson, C.E. and Bovey, F.A. (1958) *J. Chem. Phys.*, **29**, 1012–1014.

Kim, S.C. and Reid, B.R. (1992) *Biochemistry*, **31**, 12103–12116.

Kuszewski, J., Qin, J., Gronenborn, A.M. and Clore, M. (1995a) *J. Magn. Reson.*, **B106**, 92–96.

Kuszewski, J., Gronenborn, A.M. and Clore, M. (1995b) *J. Magn. Reson.*, **B107**, 293–297.

Lamm, G. and Pack, G.E. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 9033–9036.

Lamm, G. and Pack, G.E. (1997) *J. Phys. Chem.*, in press.

Laws, D.D., de Dios, A.C. and Oldfield, E. (1993) *J. Biomol. NMR*, **3**, 1607–1612.

Le, H. and Oldfield, E. (1994) *J. Biomol. NMR*, **4**, 341–348.

Mooren, M.M.W., Pulleyblank, D.E., Wijmenga, S.S., Van de Ven, F.J.M. and Hilbers, C.W. (1994) *Biochemistry*, **33**, 7315–7325.

Mujeeb, A.M., Kerwin, S.M., Egan, W., Kenyon, G.L. and James, T.L. (1992) *Biochemistry*, **31**, 9325–9338.

Oldfield, E. (1995) *J. Biomol. NMR*, **5**, 217–225.

Orozco, M., Laughton, C.A., Herzyk, P. and Neidle, S. (1990) *J. Biomol. Struct. Dyn.*, **8**, 359–373.

Ösapay, K. and Case, D. (1991) *J. Am. Chem. Soc.*, **113**, 9436–9444.

Ösapay, K. and Case, D. (1994) *J. Biomol. NMR*, **4**, 215–230.

Ösapay, K., Theriault, Y., Wright, P. and Case, D.A. (1994) *J. Mol. Biol.*, **244**, 183–197.

Pack, G.R., Garrett, G.A., Wong, L. and Lamm, G. (1993) *Biophys. J.*, **65**, 1363–1370.

Pieters, J.M.L., De Vroom, E., Van der Marel, G.A., Van Boom, J.H., Koning, T.M.G., Kaptein, R. and Altona, A. (1990) *Biochemistry*, **29**, 788–799.

Ribas Prado, R. and Giessner-Prettre, C. (1981) *J. Mol. Struct.*, **76**, 81–92.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer Verlag, Heidelberg, Germany.

Schmitz, U., Sethson, I., Egan, W. and James, T.L. (1992) *J. Mol. Biol.*, **227**, 510–531.

Schulze, P., Macaya, R.F. and Feigon, J. (1994) *J. Mol. Biol.*, **235**, 1532–1547.

Szilagyi, L. (1996) *Prog. Magn. Reson.*, **27**, 325–443.

Van de Ven, F.J.M. and Hilbers, C.W. (1988) *Nucleic Acids Res.*, **16**, 5713–5726.

Van Dongen, M.J.P., Wijmenga, S.S., Van der Marel, G.A., Van Boom, J.H. and Hilbers, C.W. (1996) *J. Mol. Biol.*, **263**, 715–729.

Van Dongen, M.J.P., Mooren, M.M.W., Wijmenga, S.S., Van der Marel, G.A., Van Boom, J.H. and Hilbers, C.W. (1997) *Nucleic Acids Res.*, **25**, 1537–1547.

Veal, J.M. and Wilson, W.D. (1991) *J. Biomol. Struct. Dyn.*, **8**, 1119–1145.

Weiner, S.J., Kollman, P.J., Nguyen, D.T. and Case, D.J. (1986) *Comput. Chem.*, **7**, 230–252.

Williamson, M.P. and Asakura, T. (1993) *J. Magn. Reson.*, **B101**, 63–71.

Williamson, M.P., Kikuchi, J. and Asakura, T. (1995) *J. Mol. Biol.*, **247**, 541–546.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E. (1980) *Nature*, **287**, 573–578.

Wishart, D.S. and Sykes, B.D. (1994) *Methods Enzymol.*, **239**, 363–392.

Wijmenga, S.S., Mooren, M.M.W. and Hilbers, C.W. (1993) In *NMR of Macromolecules. A Practical Approach.* (Ed., Roberts, G.C.K.), IRL Press at Oxford University Press, Oxford, U.K.