

# The use of NMR methods for conformational studies of nucleic acids

Sybre S. Wijmenga\*, Bernd N.M. van Buuren

*Umeå University, Department of Medical Biochemistry and Biophysics, S 901 87 Umeå, Sweden*

Received 10 July 1997

## Contents

1. Introduction	288
2. RNA and DNA synthesis and purification	290
3. Nomenclature	291
4. Distances	291
4.1. Overview of short distances and their general characteristics	292
4.2. Overview of structurally important intra-nucleotide distances	294
4.3. Overview of structurally important sequential and cross-strand distances	295
4.4. Derivation of distances from NOESY spectra and structure characterization using distances	295
4.5. Conclusion	304
5. J-couplings	305
5.1. $^1J_{\text{HC}}$ - and $^1J_{\text{CC}}$ -couplings	307
5.2. Overview of J-couplings in the bases	307
5.3. Ribose sugar	307
5.4. Determination of the $\beta$ torsion angle	310
5.5. Determination of the $\epsilon$ torsion angle	312
5.6. Torsion angle $\gamma$ and H5' and H5'' stereo specific assignment	314
5.7. $\chi$ torsion angle and $^3J_{\text{HC}}$ sugar to base	316
5.8. Measurement of homo- and heteronuclear J-coupling constants	316
5.8.1. Determination of J-couplings from the shape of the signal	316
5.8.2. Determination of J-couplings from E.COSY patterns	318
5.8.2.1. Homonuclear E.COSY	318
5.8.2.2. Heteronuclear E.COSY	319
5.8.2.2.1. Determination of $J_{\text{HP}}$ - and $J_{\text{CP}}$ -couplings	319
5.8.2.2.2. Determination of $J_{\text{HC}}$ -couplings	320
5.8.2.2.3. Determination of $J_{\text{HH}}$ -couplings via HCC-E.COSY spectra	320
5.8.3. Determination of J-couplings from signal intensities	321
5.8.3.1. Determination of $J_{\text{HH}}$ -couplings from homonuclear (H,H) TOCSY transfer	321
5.8.3.2. Determination of J-couplings from heteronuclear experiments	321

\* Corresponding author. Tel: +46 9078 6500; fax: +46 9013 6310;  
e-mail: sybren@indigo.chem.umu.se

6. Chemical shifts . . . . .	322
6.1. Chemical shifts; qualitative aspects . . . . .	324
6.2. Theory . . . . .	324
6.3. $^1\text{H}$ shifts . . . . .	327
6.4. Structurally important $^1\text{H}$ shifts . . . . .	327
6.5. $^{15}\text{N}$ and $^{13}\text{C}$ shifts in DNA and RNA . . . . .	329
6.6. $^{31}\text{P}$ shifts . . . . .	330
7. Assignment methods . . . . .	330
7.1. Assignment without isotope labeling . . . . .	330
7.2. Assignment with isotope labeling . . . . .	335
7.2.1. NOE-based correlation . . . . .	336
7.2.2. Through-bond correlation . . . . .	337
7.2.2.1. Coherence transfer functions . . . . .	337
7.2.2.2. Through-bond amino/imino to non-exchangeable proton correlation . . . . .	337
7.2.2.3. Through-bond H2-H8 correlation . . . . .	344
7.2.2.4. Through-bond base – sugar correlation . . . . .	347
7.2.2.5. Through-bond sugar correlation . . . . .	355
7.2.2.6. Through-bond sequential backbone assignment . . . . .	357
7.2.3. X-filter techniques . . . . .	361
8. Relaxation and dynamics . . . . .	363
9. Calculation of structures . . . . .	375
10. Prospects for larger systems . . . . .	378
11. Conclusions . . . . .	382
Acknowledgments . . . . .	383
References . . . . .	383

**Keywords:** NMR; Conformational studies; Nucleic acids; RNA; DNA; Labeling; Assignment; Structure

## 1. Introduction

Nucleic acid molecules play a central role in cell biological processes. DNA's main role is to act as the carrier of genetic information. Furthermore, DNA is transcribed into RNA by a carefully regulated process, and it is duplicated on cell division. RNA's main role is to communicate the genetic information for protein synthesis to the ribosomes. RNA is, however, very versatile. It can also take on the role of DNA as the carrier of genetic information, and it can function as an enzyme. It has even been hypothesized that early in evolution, life was based entirely on RNA (see, for example, Ref. [1]). All these different processes require different structures. The basic structural elements of RNA and DNA are well established, i.e. DNA forms a B-helix, while RNA may be either single-stranded or may form an A-type helix.

However, the alternate RNA and DNA structures, associated with many of the different processes mentioned above, are less well known. Only since the early 1990s have technological advances in sample preparation, such as isotope labeling and developments in crystallization, made such structural data available, and allowed the structural basis of the biological functions of DNA and RNA to be addressed.

In the past ten years, we have witnessed an explosion in the number of crystal and solution structures of proteins determined by X-ray crystallography and NMR, respectively. In comparison, the increase in the number of nucleic acid structures determined by either X-ray or NMR has been relatively small. This can be attributed to the difficulties encountered when trying to crystallize nucleic acids for detailed X-ray analysis and to the problem of extensive resonance

overlap in NMR spectra of these compounds. Advances in crystallization techniques have in recent years resulted in the structure determination of the RNA hammerhead enzyme [2,3], one of the two folding domains of the group I intron self-splicing RNA [4,5], and a few RNA–protein complexes [6,7]. In addition, the structures of several DNA duplexes, as well as of a DNA quadruplex, have been determined by means of X-ray crystallography [8]. However, despite the two X-ray structures of the hammerhead, the catalytic mechanism of this ribozyme has not yet been clarified. Crystal packing forces sometimes affect RNA or DNA structures. For example, there is still no crystal structure available of a DNA or RNA hairpin, since these tend to crystallize in biologically less relevant extended duplex structures [9,10]. Solution structures, which can be determined via NMR, are therefore particularly important in DNA and RNA structural biology as a complement to crystallography. In addition, nucleic acids often contain regions of higher conformational flexibility. NMR is particularly suited for identifying such regions.

In the field of NMR of nucleic acids, advances were made in the 1980s with the introduction of synthetic methods for preparing well defined DNA sequences. This development also made it possible to produce well defined RNA sequences from DNA templates by enzymatic synthesis via T7-polymerase. These developments led to the determination of several solution DNA and RNA hairpin structures, from which the main folding principles of hairpin loops could be determined [11,12]. In addition, these developments led to the determination of the solution structure of a DNA quadruplex [13,14] and solution structures of triple helix molecules [15], as well as to the determination of a new DNA multi-stranded fold, the C-motif [16]. Still, the overlap encountered in NMR spectra limited the size of the molecules that could be studied and the detail by which the structures could be determined. In the early 1990s, methods were developed to produce  $^{13}\text{C}$  or  $^{15}\text{N}$  enriched RNAs, via enzymatic synthesis, in quantities large enough for NMR studies. This possibility enabled more detailed studies of biologically relevant RNA sequences and folds. Initial NMR studies have been performed and methods have been developed for assignment of resonances of  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled RNAs. The direct result of

this is more reliable resonance assignments. In addition, more extensive constraints lists could be obtained for subsequent structure determination. In the past two years a number of RNA structures with a size up to 30 to 40 nucleotides have been published [17–30], together with RNA–peptide complexes [31–36] and an RNA–protein complex of total molecular weight 22 kDa [37,38]. These studies have also made it clear that the upper size limit for RNAs which can be studied by NMR lies around 30 nucleotides when uniform labeling is employed, a size limit considerably below that for proteins.

Only quite recently has it become possible to enrich DNA with  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopes. Zimmer and Crothers [39] demonstrated that DNA can be enriched via an enzymatic approach, while even more recently  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled DNA phosphoramidites have become available [40], so that  $^{13}\text{C}$  and  $^{15}\text{N}$  enriched DNAs can now also be obtained via chemical synthesis. It is to be expected that these possibilities will also have an effect on NMR structural studies of DNA of larger size. Larger DNA systems, such as those forming three- and four-way junctions, have already been studied [41], but these have not yet produced detailed solution structures, again due to the extensive signal overlap (see, for example, Refs. [42–44]). It is noteworthy that Altona and co-workers used an extremely interesting approach to achieve the assignments in their studies of four-way junctions [43,44]. They used well-determined hairpins as building blocks for the larger four-way and three-way junctions they studied. This made it possible to obtain resonance assignment in very crowded spectra. The future will reveal whether combining this approach with labeling will allow an extension to larger systems, both for RNAs and DNAs.

Naturally, as isotope enriched nucleic acid molecules are now used in NMR studies, we will pay particular attention in this review to the related NMR methods. Various other reviews [36,45–48] have recently appeared, but they have focused generally on specific aspects of the NMR of isotope enriched RNA. We try here to provide a broad overview, covering as much as possible of the various aspects that come into play when performing NMR structural studies of both DNA and RNA molecules. Furthermore, the field is developing rapidly and new aspects have been published since the appearance of

these reviews. For example, a complete overview of  $J$ -couplings in the nucleic acid bases has been published [49] and proton structural chemical shifts have been calculated and compared with experimental data [50]. We will incorporate these aspects into this review, together with a detailed description and critical evaluation of the present state of the art NMR methodology for determining the structure of labeled DNA and RNA molecules. This review is divided into eleven sections. In Section 2,  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling, as well as other labeling methods, are described, albeit briefly, in view of the quite detailed descriptions that have recently appeared. The IUPAC nomenclature is introduced in Section 3. In Section 4, we present an overview of the distances found in DNA and RNA molecules and discuss their relevance for NMR structural studies. Section 5 gives an overview of all homonuclear and heteronuclear  $J$ -couplings and describes their structural dependencies. We also give an overview of the NMR methods that are or can be used to determine these  $J$ -couplings. In Section 6, we describe the chemical shifts and discuss their use both for assignment purposes and as structural parameters. Section 7 forms the heart of this review, and describes and discusses in detail the currently available methods for assignment both in unlabeled and  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled compounds. Section 8 concentrates on a description of relaxation. Isotope enrichment has opened up the way for detailed relaxation studies in the field of proteins. Such relaxation studies are still scarce in the field of nucleic acids. We place relaxation studies on nucleic acids in the context of parallel studies on proteins, and give an overview of the theoretical background. In Section 9 we briefly describe the actual structure determination from NMR data. In Section 10, we discuss the prospects for extension of NMR studies to larger systems and we attempt to draw some conclusions in Section 11.

## 2. RNA and DNA synthesis and purification

Two strategies are available for preparing large quantities of DNA and RNA of defined sequence and high purity for NMR studies: (1) chemical synthesis by the phosphoramidite method, and (2) enzymatic synthesis of RNAs via T7-polymerase and of DNAs via DNA-polymerase. For RNA,

enzymatic synthesis is the usual method of preparation; although chemical synthesis is possible it is still prohibitively expensive when large quantities are required. Chemical synthesis is the usual approach for the preparation of DNAs of defined sequence. Zimmer and Crother [39] have shown how large quantities of DNA can be made via enzymatic synthesis, thus demonstrating the feasibility of  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling of DNA via this method. However,  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled DNA phosphoramidites have also recently become commercially available, so that labeled DNAs can conveniently be prepared via chemical synthesis [40]. We refer the reader to the original papers or reviews for the detailed protocols and for discussions of the relative merits of the various approaches [36,45,47,48,51–59]. Here we will concentrate on some general and qualitative aspects.

A certain amount of confusing terminology has crept into the literature with regard to labeling. We will use the following terms: *uniform labeling*, when every atom of a certain type in the molecule is enriched; *residue-type-specific labeling*, if all residues of a certain type (e.g. all Adenines) in the molecule are enriched; *site-specific labeling*, if a particular residue or a number of particular residues are enriched, e.g. A10; *partial labeling*, if the labeling of a certain residue is on, say, C1' only. In order to indicate that labeling is not 100%, we add the percentage after the word labeling.

For the enzymatic synthesis of RNA, a DNA template is required from which the RNA is transcribed by T7-polymerase using NTPs as building blocks. The  $^{13}\text{C}$  and/or  $^{15}\text{N}$  and/or  $^2\text{H}$  labeled NTPs are usually obtained from *E. coli* cells, which are grown on either  $^{13}\text{C}$  enriched glucose, and/or  $^{15}\text{N}$  enriched ammonium chloride. The RNA isolated from the cells is broken down to  $^{13}\text{C}$  and/or  $^{15}\text{N}$  labeled NMPs, which are subsequently converted into NTPs. This method thus allows uniformly labeled RNAs to be made, or residue-type-specific labeled RNA when the in vitro transcription occurs on a mixture of labeled and unlabeled NTPs. The method can in principle easily be extended to achieve deuteration or partial labeling. For example, Michnicka et al. [60] have suggested partial  $^{13}\text{C}$  labeling using acetate as a carbon source; most recently Nikonowicz et al. [57] have demonstrated uniform  $^2\text{H}/^{15}\text{N}$  labeling via the enzymatic approach. It is more complicated to

achieve site-specific labeling via the enzymatic method (see, for example, Ref. [36]). Site-specific labeling, on the other hand, can quite easily be achieved via chemical synthesis. This would be the method of choice for the preparation of labeled DNA oligonucleotides.

### 3. Nomenclature

For atom numbering and torsion angle definitions in nucleic acids we will follow the IUPAC/IUB guidelines [61]. Accordingly, the chemical structure and atom numbering of the five common bases, the pyrimidines C, T and U, and the purines G and A, are given in Fig. 1(A), and of the  $\beta$ -D-(deoxy) riboses in Fig. 1(B), which also indicates the torsion angles in the sugar-phosphate backbone ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$  and  $\zeta$ ) and the glycosidic torsion angle  $\chi$ . Their definitions are: O3'-P-O5'-C5' ( $\alpha$ ), P-O5'-C5'-C4' ( $\beta$ ), O5'-C5'-C4'-C3' ( $\gamma$ ), C5'-C4'-C3'-O3' ( $\delta$ ), C4'-C3'-O3'-P ( $\epsilon$ ), C3'-O3'-P-O5' ( $\zeta$ ), O4'-C1'-N1-C2 ( $\chi$  (Py)), and O4'-C1'-N9-C4 ( $\chi$  (Pu)). Furthermore, it gives a designation of the chain direction and the unit numbering in a polynucleotide chain. Fig. 1(C) shows the two most common conformations of the  $\beta$ -D-(deoxy)ribose sugar ring, the C2'-endo ( $^2E$ ) and the C3'-endo ( $^3E$ ) conformers, also referred to as S-type and N-type conformers, respectively.

To describe the distances we will use the shorthand notation introduced by Wijmenga et al. [62]. In this notation the distance between the protons *l* and *r* is given by:

$d_i(l;r)$  for intranucleotide distances, e.g.  $d_i(8;2')$

$d_s(l;r)$  for internucleotide distances, e.g.  $d_s(1';6)$

Here, *l* corresponds to the proton in the 5'-nucleotide unit and *r* with the proton in the 3'-nucleotide unit. For methyl protons the *l* or *r* is indicated by the letter M. To indicate that the distance is between H3' in the 5'-nucleotide and H5 or the methyl protons in the 3'-nucleotide we use  $d_s(3';5/M)$ . Cross-strand distances are defined as:

$d_{ci}(l;r)$  for distances within a base pair,

e.g.  $d_{ci}(T-NH3; A-NH_26)$

$d_{cs}(l;r)_p$  for distances between adjacent base paired nucleotides, e.g.  $d_{cs}(1';2)_3$

The symbols NH and NH<sub>2</sub> represent imino and amino protons, respectively. The directionality in the sequential cross-strand distances has to be indicated. Consider two adjacent base pairs, and define the 5'- and 3'-nucleotides. It can be easily seen that  $d_{cs}$  is either between two 3'-nucleotides or between two 5'-nucleotides. This is indicated by the subscript *p*. Alternatively, when two protons *l* and *r* do not fall in any of the above categories the distance is indicated by:

$d(l;r)$  for long-range internucleotide distances,

e.g.  $d(T2-NH3; A9-NH_26)$

Here, T2-NH3 indicates the imino proton of Thymine number 2 and A9-NH<sub>2</sub>6 indicates the amino group of Adenine number 9.

### 4. Distances

Proton to proton distances are essential parameters for the three-dimensional structure determination of biomolecules by NMR. Since only short distances ( $< 5-6$  Å) can be obtained by NMR, it is difficult to determine global features, such as bending of the helix. On the other hand, local features can be determined quite well and most NMR structural studies have focused on these aspects. Consequently, it is of paramount importance to have a good overview of the short distances in the main structural elements, such as the sugar ring, the bases, the base pairs, etc. and of how these distances determine the structural features of those elements. Another aspect is that several of the short distances do not depend on conformation nor do they take on well defined values for the two major helical conformations, A- and B-helices. For this reason, it is particularly useful to have at hand an overview of these distances and their characteristics, so that one can focus on the relevant data for the more interesting structural aspects.

In the next sections we therefore discuss the short distances and how they reflect structural characteristics, by first giving a more general overview and

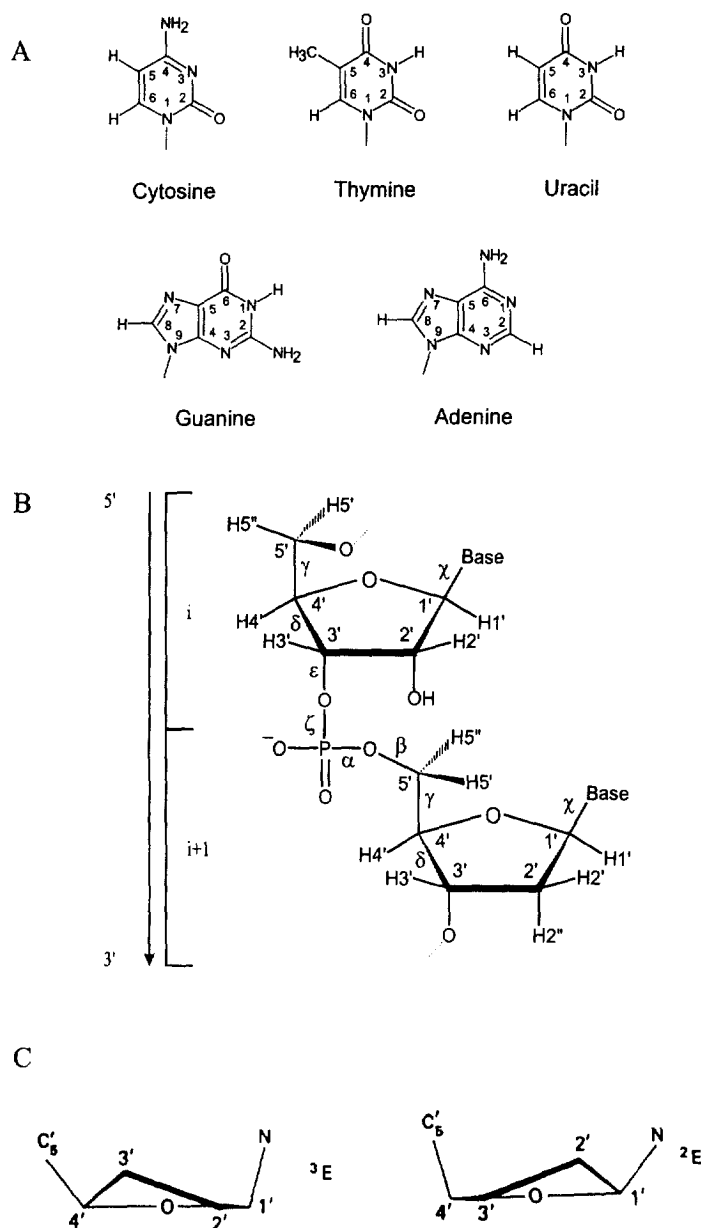


Fig. 1. Structure and atom numbering in nucleic acids, according to the IUPAC/IUB guidelines [61], of the five common bases (pyrimidines C, T and U; purines G and A) (A), and of the  $\beta$ -D-(deoxy)ribose (B and C). (B) also shows the torsion angles in the sugar-phosphate backbone ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$  and  $\zeta$ ) and the glycosidic torsion angle  $\chi$  (the exact definition is given in the text), a designation of the chain 5' to 3' direction and the unit numbering in a polynucleotide chain. (C) shows the puckering of the two most common  $\beta$ -D-(deoxy)ribose sugar ring conformations, the C2'-endo (or S-type) and the C3'-endo (or N-type) conformations.

subsequently going into more detail. Finally, we discuss their derivation from NOESY spectra and their use as constraints in simulated annealing protocols.

#### 4.1. Overview of short distances and their general characteristics

In Table 1 we have summarized the short distances

Table 1  
Overview of short distances per residue

Type	%	M	A	B	C	C'	D
<b>Intra-nucleotide</b>							
1. constant	5	3	3	0	0	0	0
2. sugar–sugar	16	10	0	8	2	2	1
3. sugar–5'/5"	13	8	0	4	4	0	0
4. base–sugar	12	7	0	1	6	4	4
sum	46	28	3	13	12	6	5
<b>Inter-nucleotide</b>							
<i>I. non-exchangeable</i>							
1. sequential sugar–base	20	12	0	0	12	12	12
2. sequential base–base	6	4	0	0	4	4	0
3. sequential sugar–sugar	20	12	0	0	12	2	2
4. cross-strand	(3%)	(2)	0	0	(2)	(2)	(2)
sum	46	28	0	0	28	18	14
<i>II. exchangeable (imino/amino)</i>							
1. within base pair	2	1	0	0	1	1	0
2. sequential	3	2	0	0	2	2	0
3. cross-strand	3	2	0	0	2	2	2
sum	8	5	0	0	5	5	2
Total	100	61	3	13	45	29	20
%		100	5	21	74	48	33

M: measurable distances  $< 5$  to  $6 \text{ \AA}$ [62]; A: completely conformation independent distances; B: distances that are conformation independent within approximately  $\pm 0.2 \text{ \AA}$ ; C: 'structural distances', i.e. conformation dependent distances with variation  $> 0.2 \text{ \AA}$  (see text); C': NMR accessible 'structural' distances; D: NMR accessible 'structural' distances that are different in A- and B-helices. The **intra-nucleotide** distances: 1. The constant distances  $d_i(2';2'')$ ,  $d_i(5';5'')$ ,  $d_i(6;5)$  and  $d_i(6;M)$ ; 2. the sugar-to-sugar distances,  $d_i(1'-4';1'-4'')$ ; they all fall into group B, except for  $d_i(2'';4')$  and  $d_i(1';4')$ , which fall into groups C and C', while group D only contains  $d_i(2'';4'')$ ; 3. the distances  $d_i(2'-4';5'/5'')$ ; group B contains the distances  $d_i(2'/2'');5'/5'')$ ; group C contains  $d_i(3'/4';5'/5'')$ ; none of them fall into groups C' or D; 4. sugar-to-base distances  $d_i(6/8;1'-5'')$ ; they are subdivided according to: group B,  $d_i(6/8;4')$ , group C,  $d_i(6/8;1'-3', 5'/5'')$ , group C', excluded  $d_i(6/8;5'/5'')$ , group D,  $d_i(6/8;1'-3')$ . The distances  $d_i(5/M;1'-5'')$  are not taken into account since they are larger than  $5 \text{ \AA}$ [62]. The **inter-nucleotide** distances (considered are the distances to and from Cytosine in a GCG trinucleotide sequence (see Fig. 2). I. Non-exchangeable protons: 1. sequential sugar-to-base distances,  $d_s(1'-3';8)$  and  $d_s(1'-3';6/5)$ ; all of them fall into categories C, C' and D; 2. base-to-base distances,  $d_s(6/5;8)$  and  $d_s(8;6/5)$ ; all of them fall into categories C, C' and D; 3. sequential sugar-to-sugar distances,  $d_s(1'-2'',4';5'')$ ,  $d_s(5'';1'-2'',4')$ ,  $d_s(2';3')$ ,  $d_s(3';2')$ ,  $d_s(2';2'')$  and  $d_s(2';2'')$ ; all of them are conformation dependent (category C), but only  $d_s(2';3')$  and  $d_s(3';2')$  are easily accessible and differ between A-type and B-type helices. II. Inter-nucleotide distances involving exchangeable protons: 1. The distances within a base pair are  $d_c(\text{NH}_2; \text{NH})$ ; this distance depends on conformation (category C), is NMR accessible (C'), but does not differ between A-type and B-type helices; 2. the sequential distances are  $d_s(\text{NH}_2; \text{NH})$  and  $d_s(\text{NH}; \text{NH}_2)$ ; they both fall into category C and C', but not into category D; 2. the cross-strand distances are  $d_c(\text{NH}_2; \text{NH}_2)_5$  and  $d_c(\text{NH}_2; \text{NH}_2)_3$ ; they fall into category C, C' and D (however, note that the NH<sub>2</sub> resonances of G may be broadened making them inaccessible for NMR).

( $< 5\text{--}6 \text{ \AA}$ ) and categorized them into two main groups, intra-nucleotide and inter-nucleotide distances, with further subdivision to reflect more detailed conformational characteristics. The inter-nucleotide distances fall into the two broad groups of sequential and cross-strand distances involving non-exchanging protons and exchanging protons, respectively. The sequential distances involving non-exchanging protons are again subdivided into sugar-to-sugar distances, base-to-base distances and sugar-to-base distances. Within each category their

dependence on conformation is indicated (A to D), with category A referring to conformation independent distances, category B to distances that can vary by less than  $\pm 0.2 \text{ \AA}$ , and category C to 'structural' distances, i.e. distances that convey structural information since they can vary by more than  $\pm 0.2 \text{ \AA}$ . The 'structural' distances, category C, are subdivided into two further categories to indicate their usefulness; category C' contains those 'structural' distances that are reasonably well accessible by NMR, and category D refers to NMR accessible 'structural' distances that

also show differences depending on whether they are present in an A- or B-type helix.

As can be seen from Table 1 approximately 60 distances per residue can in principle be measured. The number of distances that are constant within  $\pm 0.2$  Å is rather high. They represent about 26% of the total number of measurable distances. Their percentage is even higher for the intra-nucleotide distances, of which they represent about 57% (16 out of 28). The distances that convey relevant structural information (in helices) and are reasonably well accessible by NMR represent less than half (48%) of the total number of distances, while only 20 are different between A- and B-type helices (33%). Note also the small number of structurally very important cross-strand and sequential distances involving exchanging protons which establish base pairing (8%), and the small number of cross-strand distances involving non-exchanging protons (3%). On the other hand, sequential sugar-to-base and sugar-to-sugar distances, which are so important for establishing base stacking and defining the phosphate backbone, are both relatively large in number (20%). The former are mostly reasonably well accessible by NMR, whereas the latter are extremely difficult to establish. Thus, a rather uneven spread in the short distances is found through the chemical structure. As a consequence, important structural features such as base pairing often hinge on the presence of a particular NOE contact reflecting one short distance.

#### 4.2. Overview of structurally important intra-nucleotide distances

The **intra-nucleotide** distances in DNA and RNA can conveniently be subdivided according to the categories indicated in Table 1, i.e. (1) conformation independent distances, (2) distances between sugar protons, (3) distances between H2'/2''/3'/4' and H5'/5'', (4) distances between H1' through H5'/5'' and base protons.

1. The conformation independent distances are: the geminal proton distances,  $d_i(2';2'')$  and  $d_i(5';5'')$ , of 1.8 Å,  $d_i(5;6)$  ( $\approx 2.45$  Å) in Cytosine and Uracyl, and  $d_i(6,M)$  in Thymidine.
2. The distances within the sugar ring are all independent of its conformation, except for  $d_i(2'';4')$  and

$d_i(1';4')$ . Only  $d_i(2'';4')$  differs significantly between S-type and N-type conformers, with  $d_i(2'';4') = 4.2$  Å for the S-type conformer (pseudorotation angle  $P = 160^\circ$ ) and  $d_i(2'';4') = 2.8$  Å for the N-type conformer ( $P = 10^\circ$ ). Although it is in principle possible to determine the sugar conformation from the  $d_i(2'';4')$  distance, the accuracy of the determination is limited. The  $d_i(2'';4')$  distance is difficult to determine from NOE intensities because of spin diffusion effects, due to the close proximity of the H2' and H2'' protons. Also note that in RNA the H2'' proton is absent, so that these sugar distances cannot be used at all to determine the puckering. The distance  $d_i(1';4')$  is almost identical for N-type and S-type sugars (3.4 Å), but has a lower value for sugar rings with an intermediate pseudorotation angle,  $d_i(1';4') = 2.6$  Å for  $P = 90^\circ$ . Here again spin diffusion can adversely affect the accuracy distance of the determination.

3. The distances  $d_i(3';5'/5'')$  depend only weakly on the sugar ring conformation, but significantly on the  $\gamma$  torsion angle, while the distances  $d_i(4';5'/5'')$  only depend on the  $\gamma$  torsion angle. The distances  $d_i(3'/4';5'/5'')$  therefore allow the determination of the torsion angle [62,63]. This can be done in conjunction with relevant  $J$ -couplings (see Section 5). Given an uncertainty in these distances of  $\pm 0.2$  Å, they do not discriminate well between the different ranges of the  $\gamma$  torsion angle, in particular when an equilibrium between  $g^+$  and  $g^-$  rotamers exists. The distances  $d_i(2'/2'';5'/5'')$  depend on both the sugar puckering and the torsion angle  $\gamma$ , but their dependence is weak, and they are of the order of 5 to 6 Å [62].
4. The distance between H1' and H8/6,  $d_i(1';6/8)$ , depends **only** on the glycosidic torsion angle  $\chi$ . It thus provides a means for determining this torsion angle. However, the maximum difference in the values of  $d_i(1';6/8)$  for  $\chi$  in the *syn* domain ( $\chi = 60^\circ$ ) and in the *anti* domain ( $\chi = 240^\circ$ ) is only about 1.2 Å. Given that in practice the uncertainty in the distance determination from NOE data is of the order of  $\pm 0.2$  Å to  $\pm 0.5$  Å, it is to be expected that the use of  $d_i(1';6/8)$  is a rather imprecise means to determine the  $\chi$  torsion angle. The other sugar proton to base proton distances,  $d_i(2'/2''/3'/4';6/8)$ , depend on **both** the sugar puckering



and the  $\chi$  torsion angle. The distance  $d_i(4';6/8)$  does not convey useful structural information since its dependence on these parameters is weak [62]. The distances  $d_i(2'/2''/3';6/8)$  are on the other hand quite useful. Each of these distances defines the  $\chi$  torsion angle quite well, because of their quite strong dependence on the torsion angle  $\chi$  [62]. Their dependence on the sugar pucker is rather weak, in particular for the distances  $d_i(2'/2'';6/8)$  [62]. Despite this weak dependence on the sugar pucker, a concerted use of  $d_i(2'/2''/3';6/8)$  makes it possible to determine the percentage N-type or S-type pucker, but to achieve a reasonable level of precision requires that the uncertainty in their values should be less than  $\pm 0.5 \text{ \AA}$  [62]. We finally note that Lane and co-workers [64] have shown the improved reliability of sugar pucker determination using these distances together with  $J$ -couplings.

The H5'/H5'' to base proton distances,  $d_i(5'/5'';6/8)$ , depend on three torsion angles,  $\gamma$ ,  $\delta$  and  $\chi$ . Their dependence on the sugar pucker ( $\delta$ ), and on the glycosidic torsion angle ( $\chi$ ) in the usual *anti* domain (180–240°) is weak, but they depend quite strongly on the  $\gamma$  torsion angle. In particular, for  $\gamma^+$  both  $d_i(5';6/8)$  and  $d_i(5'';6/8)$  are long (3.7 to 4.5 Å), while for  $\gamma^-$  the distance  $d_i(5'';6/8)$  becomes short (2.5 to 2.9 Å). As has been shown, with uncertainties in the distance estimates in the order of  $\pm 0.2 \text{ \AA}$ , they determine quite well the torsion angle  $\gamma$  [62]. The distances  $d_i(5'/5'';6/8)$  can be quite useful in NOESY spectra of DNA, since the related NOE cross peaks do not reside in a crowded spectral region. This does not hold true for RNA where these cross peaks overlap with the other H6/8 to H2'/3' NOE cross peaks. On the other hand, the distances  $d_i(3'/4';5'/5'')$  all relate to cross peaks in crowded spectral regions for both DNA and RNA and are thus difficult to establish.

#### 4.3. Overview of structurally important sequential and cross-strand distances

Helical conformations form an important part of nucleic acid structures. We therefore present an overview of the distances in the two most commonly found helix types, A-helices and B-helices. Fig. 2,

reproduced from Wijmenga et al. [62], gives the sequential distances,  $d_s(l;r)$ , and cross-strand distances,  $d_{ci}(l;r)$  and  $d_{cs}(l;r)$ , found in A-DNA, B-DNA and RNA helices.

The cross-strand distances,  $d_{ci}(l;r)$  and  $d_{cs}(l;r)$ , involving exchangeable protons are indicative of base pair formation. The sequential distances involving either exchanging or non-exchanging protons are indicative of base stacking. However, only a limited number depend on the type of helix conformation. In both A- and B-type helices, short base-to-base distances,  $d_s(6/8/5/M;6/8/5/M)$ , are present, depending on the sequence. Similarly, all distances involving exchanging protons are very similar in A- and B-type helices. The differences occur for the cross-strand and sequential distances involving H2 protons,  $d_{cs}(2;1'/2)3'$  and  $d_s(2;1')$ , the sequential sugar-to-base distances,  $d_s(2'/2''/3';6/8/5)$ , and for a number of sequential sugar-to-sugar distances,  $d_s(2'/2'';5'/5'')$ ,  $d_s(2';3')$ ,  $d_s(2'';2'')$  and  $d_s(1';5'')$ . Short cross-strand, as well as sequential H2 to H1' distances, are present in A-type helices, but absent in B-type helices. Short sequential H2' to H6/8 distances and long H2'' to H6/8 distances are seen in A-helices, while in B-helices the reverse is found. The sugar-to-sugar distances show the following pattern: Short sequential H2'/H2'' to H5'/5'' distances in A-helices, while in B-helices these distances are long; rather long, but measurable, sequential H2' to H3' distances in A-helices, which are over 7 Å and thus not measurable in B-helices; finally, long ( $> 7 \text{ \AA}$ ) sequential H2'' to H2' and H1' to H5'' distances in A-helices, which are relatively short in B-helices. While the distances involving H2 and the H2'/2'' to base distances are quite accessible from NMR spectra, the sugar-to-sugar distances are difficult to determine since the sugar proton resonances reside in quite crowded spectral regions.

#### 4.4. Derivation of distances from NOESY spectra and structure characterization using distances

We will discuss here the three aspects of NMR accessible distances that are of particular relevance for structure determination. First, how precisely can distances be derived from NOE data? Secondly, how does this precision affect the precision of the determined structure? Thirdly, how does the spread and

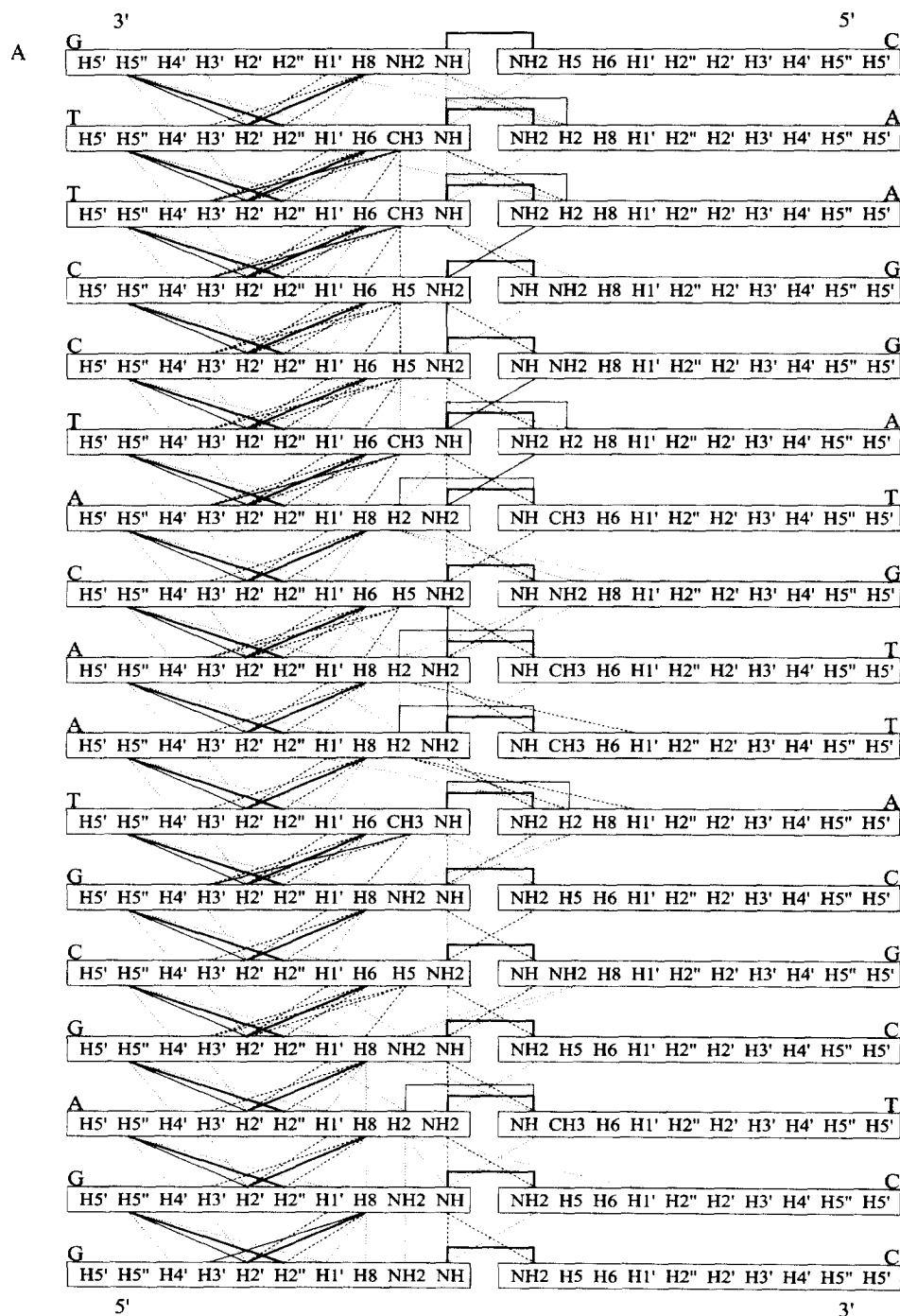


Fig. 2. Overview of short sequential and inter-strand proton–proton distances for all possible combinations of base stacking in A-DNA (A), B-DNA (B) and RNA (C). The meaning of the symbols is: 0–2.5 Å (thick solid line), 2.5–3.0 Å (solid line), 3.0–4.0 Å (dashed line), 4.0–5.0 Å (dotted line).

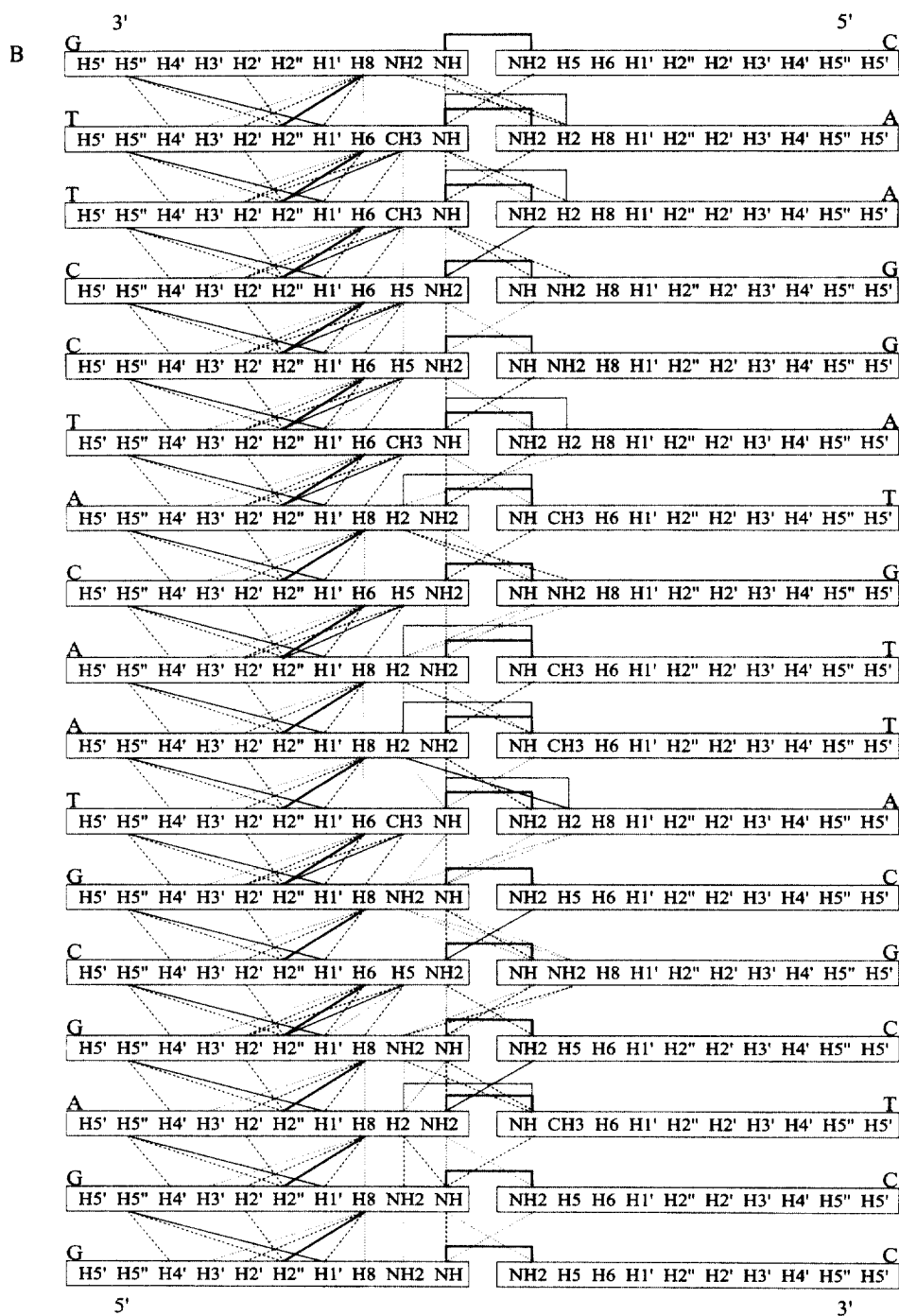


Fig. 2. (Continued).

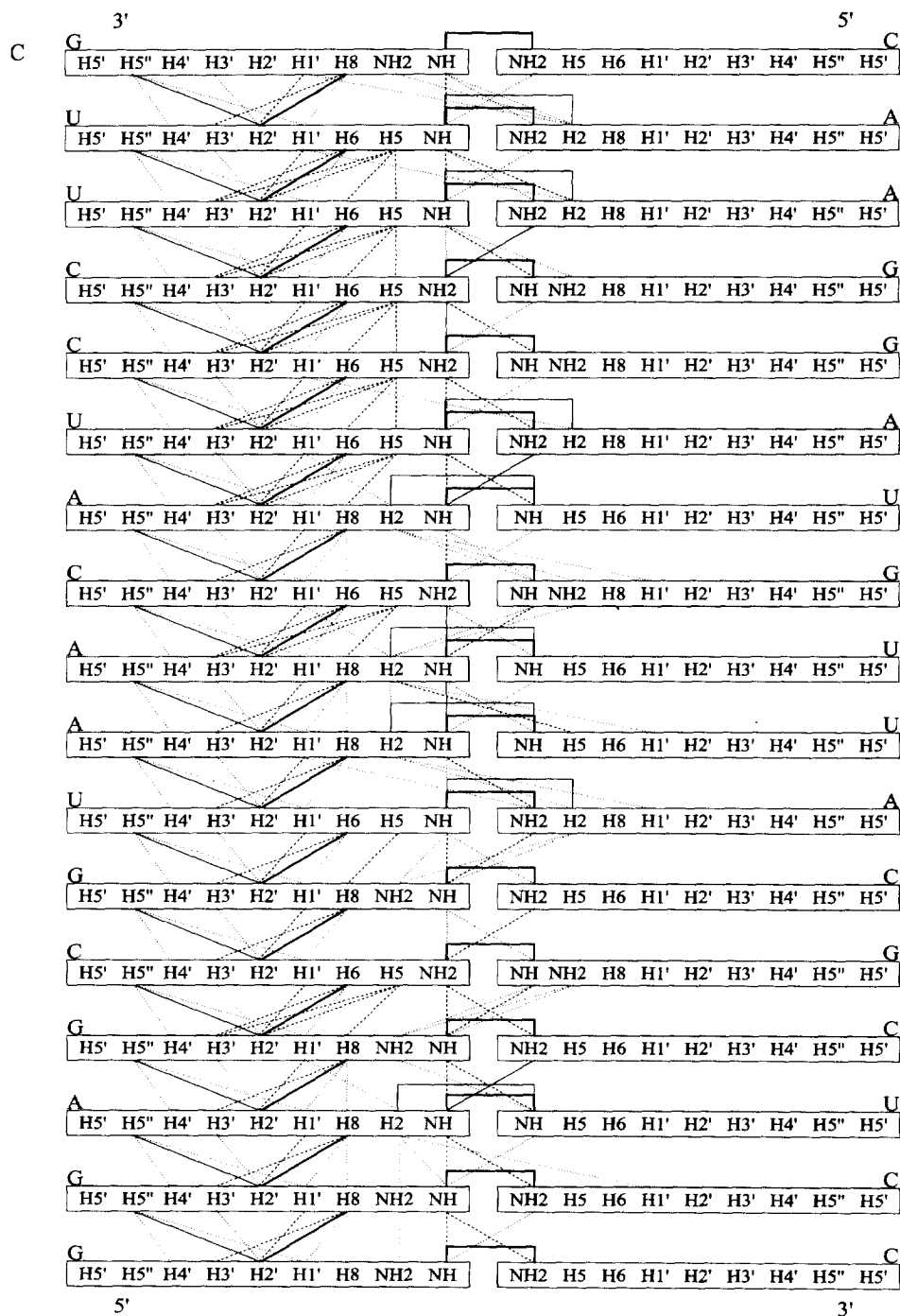


Fig. 2. (Continued).

number of distance constraints affect the precision of the determined structure?

The intensities of the cross peaks in a NOESY spectrum are related to distances between spins via the relaxation matrix,  $R$ :

$$NOE = S \exp(-R\tau_m) = S(1 - R\tau_m + \frac{1}{2}(R\tau_m)^2 - \dots) \quad (1)$$

Here  $\tau_m$  is the mixing time and  $S$  is a scaling factor taken to be equal to 1. The elements  $R_{ij}$  of the relaxation matrix are given by (see also Section 8):

$$R_{ij} = \frac{q}{r_{ij}^6}(6J(2\omega) - J(0)) \quad (2)$$

where  $q = \gamma^4(h/2\pi)^2/4$ , and  $r_{ij}$  is the distance between protons  $i$  and  $j$ . For a rigid isotropically tumbling molecule, the spectral density function,  $J(\omega)$ , can be written as:

$$J(\omega) = \frac{2}{5} \frac{\tau_c}{1 + \omega^2\tau_c^2} \quad (3)$$

so that in the slow tumbling limit ( $\omega\tau \gg 1$ ) one obtains:

$$R_{ij} = -\frac{q}{r_{ij}^6}J(0) = -\frac{q}{r_{ij}^6}\tau_c \quad (4)$$

This is the equation used in most relaxation matrix calculations (see below). Internal dynamics or conformational flexibility implies that interconversion between different conformers takes place, with each conformer having a different set of distances. The NOE intensity is then derived from the average relaxation rate,  $\langle R_{ij} \rangle$ , which is either proportional to  $\langle 1/r^6 \rangle$ , when the motion is slower than the overall tumbling, or to  $\langle 1/r^3 \rangle^2$ , when the averaging is faster than the overall tumbling time [62]. When the internal motion is of limited scope (libration motions) the average distance may be approximated by the middle distance. As a result, fast internal libration motions on a ps to ns time-scale can be accounted for by using the Lipari and Szabo approach [65,66] by introducing a scaling factor ( $S^2$ ) in Eq. (4) (see Section 8):

$$R_{ij} = -\frac{q}{r_{ij}^6}J(0) = -\frac{q}{r_{ij}^6}S^2\tau_c \quad (5)$$

The fast rotation of the methyl protons also leads to a scaling down of the NOE intensity. Corrections for

fast rotation of the methyl groups and fast internal motion (Eq. (5)) are generally incorporated into relaxation matrix programs (see below). More difficult is the situation for interconversion between conformers, which have distinctly different proton–proton distances. An example is the rapid interconversion between N-type and S-type sugars observed in nucleic acids, which leads to very different intra-nucleotide sugar-to-base distances. In this case the average distance derived from NOEs tends to be heavily biased towards the shorter distances. To account for the latter effects requires ensemble averaging, a method which has not been implemented in relaxation matrix programs mainly because of the major computational effort involved. Finally, nucleic acids are not spherically but asymmetrically shaped, leading to anisotropic instead of isotropic tumbling. Consequently, the spectral density function has to be replaced by a more complicated form (see Section 8). Anisotropic tumbling is not considered in most relaxation matrix programs (see below).

For short mixing times only the first term in the expansion given in Eq. (1) is required. The isolated spin pair approximation, ISPA, then applies, and the NOE between protons  $i$  and  $j$  is proportional to their distance to the inverse sixth power:

$$NOE_{ij} = C \frac{1}{r_{ij}^6} \quad (6)$$

Plotting the  $NOE^{-1/6}$  versus known distances, according to

$$NOE_{ij}^{-1/6} = C^{-1/6}r_{ij} \quad (7)$$

allows the determination of the constant  $C$ . Distances can then be calculated via

$$r_{ij} = C^{-1/6}NOE_{ij}^{-1/6} \quad (8)$$

For longer mixing times spin diffusion becomes more effective. As a result NOE intensity is lost in cross peaks involving spins that are close in space, while NOE intensity may be gained in cross peaks involving spins that are relatively far apart. The net effect is that the  $NOE^{-1/6}$  versus distance curve flattens and can be described by

$$NOE_{ij}^{-1/6} = A + B^{-1/6}r_{ij} \quad (9)$$

The constants  $A$  and  $B$  can be determined as before from known distances. This 'modified' ISPA approach allows the estimation of unknown distances via

$$r_{ij} = (NOE_{ij}^{-1/6} - A)B^{1/6} \quad (10)$$

Ultimately, all NOE intensity will be uniformly spread out through the network of proton spins, and, consequently, becomes independent of the distances, and distance estimates can no longer be made. This method of accounting for spin diffusion is rather crude. In effect the method more or less assumes that each spin pair is surrounded by a uniform network of other spins, which provide the spin diffusion pathways. Although this approach does take into account that short and long distances are differently affected by spin diffusion, it does not allow for the fact that different short distances may be differently affected by spin diffusion. A rigorously correct approach is to take the full relaxation matrix into account, since ultimately only the full expression correctly gives the NOE matrix. The relaxation matrix,  $R$ , and thereby the distances, can be calculated from the NOE matrix, if the complete NOE matrix is known:

$$R = \frac{-\ln(NOE)}{\tau_m} \quad (11)$$

In this way the complete spin network is considered and spin diffusion can be fully accounted for. One can now in fact derive from the NOE matrix the conformation of the spin network. In practice, not all NOE intensities are known. Consequently, one needs to somehow build up a complete NOE matrix. Several approaches have been developed for this purpose [67–74]. These approaches generally proceed as follows. In the first step a model is chosen, from which a model NOE matrix is calculated. Next, those theoretical NOEs for which measured values are available are replaced by experimental NOEs; the rest remain at the model values. Thus, a hybrid model/experimental NOE matrix is constructed which is complete and from which relaxation rates can be calculated and distances extracted. In the IRMA [73] and the MORASS [70] approaches these distances are immediately used in a further cycle of simulated annealing refinement to obtain new model distances and then a new NOE matrix. In the MARDIGRAS approach [67] the updated set of distances is directly

used to calculate a new model NOE matrix. The whole procedure of substitution, back-calculation, etc. is repeated until a good fit between experimental and calculated NOEs is obtained. In the NO2DI method [74] ISPA is used to estimate zeroth-order distances. Subsequently, a relaxation matrix is built of all spins for which measured NOEs are available. To complete the relaxation matrix the distances for which no zeroth-order distances are obtained are set to a large value. The NOE matrix is then back-calculated and from the  $NOE_{ij}(\text{calc})$  a new distance estimate obtained,  $r_{ij} = r_{ij}^6 \sqrt{NOE_{ij}(\text{calc})/NOE_{ij}(\text{exp})}$ . The new distance is used to calculate the NOE matrix again, giving a new calculated distance, until  $NOE_{ij}(\text{calc})$  and  $NOE_{ij}(\text{exp})$  are within a certain range (1%). This process is repeated for each distance, starting with short distances and working up to the longer ones, thus obtaining first-order estimates for each distance. The process can then be repeated until convergence is reached. In the NO2DI method, obviously the spin network consists of the other spins for which an experimental NOE was observed. If the number of experimental NOEs is too sparse the network may become too sparse, and model distances may be included. While in the 'modified' ISPA approach, uniform surrounding is assumed for each spin pair, here the surrounding of each spin pair does not need to be uniform and is built up from both estimated and/or model distances. It turns out that the ultimate results do not depend strongly on the choice of model [67], although the best results are obtained with model distances which are as close as possible to the true distances [68].

What are the possible sources of error in the ISPA and relaxation matrix approaches? Most easily assessed are errors in the distances resulting from the noise and integration errors in the NOE intensity. The absolute noise in a spectrum can, with modern spectrometers, be quite small and varies from 0.01% to 0.3% depending on the quality of the spectrum (assuming that the NOE matrix is scaled to 1). In addition, there is a relative error from peak integration which amounts to 10–15% of the peak volume [68]. In our studies, peak amplitudes yield good estimates of the peak volumes, when determined from highly digitized NOESY spectra. The advantage of using amplitudes over integration is that errors due to base line distortion and peak overlap are minimized. We

found that the peak amplitudes and volumes from a 3D TOCSY–NOESY spectrum agree within 10%. A second source of possible error is that in practice, NOESY spectra are recorded with relatively short relaxation delays (RD) inbetween FID recordings, i.e.  $RD < 5 T_1$ . This will affect the signal amplitude from different spins differently, depending on their  $T_1$  relaxation rates. Such effects are generally not accounted for in the above described programs. A third, more important issue, is internal dynamics. Fast internal libration motions on a ps to ns time-scale can be accounted for via Lipari and Szabo's formalism [65,66] by introducing a scaling factor ( $S^2$ ) (see Section 8). The fast rotation of the methyl protons also leads to a scaling down of the NOE intensity. These corrections are generally incorporated into the programs discussed above. To account for inter-conversion between conformers with distinctly different proton–proton distances requires ensemble averaging, a method which has not been implemented in the programs discussed above, mainly because of the enormous computational effort involved. The fourth, and major, source of error in the derivation of distances from NOEs seems to originate from the fact that the more spin diffusion contributes to the intensity of a cross peak, the more difficult it becomes to retrieve the direct contribution, even when relaxation matrix approaches are used. This can be understood qualitatively from the decreased dependence on distance, which is found when spin diffusion plays a role. This has a profound effect on how errors in NOE intensities translate into errors in the derived distances. At short mixing times, when ISPA applies, errors in the NOE intensity lead to strongly damped errors in the derived distances, because of the inverse sixth power relationship. On the other hand, when spin diffusion becomes highly effective, the distance dependence of the NOEs becomes less strong, and the error in the NOE translates into much larger errors in the derived distances.

How accurately can distances be obtained and how much better does a relaxation matrix approach perform as compared to a simple ISPA approach? The ultimate errors in the distance estimates are nicely illustrated by the model calculations performed by van de Ven et al. [74] and Borgias and James [67]. In the calculation of van de Ven et al., the NOEs of a B-DNA duplex were calculated assuming a mixing

time of 200 ms and a tumbling time of 2 ns. In addition, 0.1% random noise was added to the NOE intensities and 20% of the NOEs were deleted. Figure 7(a) of van de Ven et al. [74] gives the initial ISPA estimates of the distances, while Fig. 7(b) and (c) represent distance estimates after five cycles of NO2DI refinement using the ISPA as starting values or a wrong model as source for starting values of the distances, respectively. Van de Ven et al. found that the initial distance estimates, obtained via the ISPA approach, have in fact reasonably small error bounds; the distances are correct within  $\pm 0.3 \text{ \AA}$  at  $2.0 \text{ \AA}$  up to  $\pm 0.6 \text{ \AA}$  at  $5 \text{ \AA}$  (as judged from Fig. 7(a) in Ref. [74]). As expected the shorter distances are being calculated too large and the longer distances calculated too short on average. A 'modified' ISPA estimate of the distances removes this bias and leads to narrower error bounds of  $< \pm 0.2 \text{ \AA}$  at  $2.0 \text{ \AA}$  to  $\pm 0.6 \text{ \AA}$  at  $5 \text{ \AA}$  (again as judged from Fig. 7(a) in Ref. [74]). Most importantly, we note that all distances are correctly estimated within these bounds. Van de Ven et al. find that after five rounds of relaxation matrix refinement the shorter distances are now correctly estimated (within  $\pm 0.2 \text{ \AA}$  for distances up to  $3.0 \text{ \AA}$ ), while for larger distances the majority of the estimated distances have errors up to  $\pm 0.7 \text{ \AA}$  at  $5 \text{ \AA}$ , which is the same order of magnitude as the error when using the ISPA estimates. Similar error bounds for ISPA and relaxation matrix refined distances have been obtained by Borgias and James using MARDIGRAS [67] for NOE data with 0.3% error. Thus, it seems that rather similar error bounds are obtained for 'modified' ISPA derived and relaxation matrix refined distances, at least under these conditions. In view of these results we have investigated the derivation of distances from NOE data for the case of a 3D TOCSY–NOESY spectrum of a 12-mer RNA duplex [75–77]. The NOE mixing time was 200 ms, so that spin diffusion does affect the NOEs, albeit to a limited extent. The NOE intensities derived from the 3D TOCSY–NOESY spectrum were corrected for  $T_2$  effects and missing TOCSY peaks. Distances were then estimated using a 'modified' ISPA approach; the  $NOE^{-1/6}$  intensities for known distances were plotted along the y axis against the known distances along the x axis; the known distances covered the complete 1.8 to  $5.0 \text{ \AA}$  range. Calibration was done by fitting a straight line through the data according to Eq. (9). The known

distances could be determined with an accuracy of about  $\pm 0.3$  Å for short distances (1.8–3.0 Å), of  $\pm 0.4$  Å for distances between 3 and 4 Å, and of  $\pm 0.5$  Å for distances between 4.0 and 5.0 Å [77].

When considering the relaxation matrix refined distances obtained by van de Ven et al. [74], it is disconcerting to find that, in addition to the majority of correctly determined distances, a number are calculated to be too large, with errors ranging up to 5 Å, and in the extreme case they are said to have ‘exploded’. These ‘exploded’ distances not only occur for large true distances, but also for true distances as small as 2.6 Å. These wrong distances probably result from the difficulties in estimating the spin diffusion contribution to the total NOE intensity. This may be due to errors in the NOE intensities or errors in the model distances or a combination of these effects. To account in the calculations for the first source of error the following method has been implemented in MARDIGRAS. The calculations are repeated at least 30 times while randomly varying NOE intensities with a certain noise level for each NOE 2D dataset (e.g. 0.002–0.003 for absolute intensity and 5–10% integration error) [68]. The final distances are taken as the average; for the error one can conservatively take the maximum and minimum distance values or, as Schmitz and James suggest [68], some intermediate range, leading to error ranges of  $\pm 0.25$ – $0.4$  Å. This procedure does not however identify the erroneous distances with certainty and errors resulting from a wrong starting model may still evolve. Different starting models could be employed for estimating unknown fixed distances. Schmitz and James [68] note that a starting model closer to the true model improves the estimates. The advantage in this respect of MORASS and IRMA is that restrained MD structure calculations are done in each iteration step, thus improving the model estimates in each iteration step, thus reducing this possible source of error. On the other hand, we have found that errors in the distances are very difficult to detect in the restrained MD calculations (XPLOR, [78]). The structure is often adjusted to compensate for the erroneous distance constraint in such a way that distance violations show up not at the site of the erroneous constraint but elsewhere in the structure. Thus, simulated annealing does not provide a certain means to identify erroneous distance constraints. In

fact, it is our experience that it is of prime importance for simulated annealing to derive distance constraints which are assured to be correct. How to identify the types of errors? One simple approach, which we have not seen suggested in the literature, however, is to compare the relaxation matrix estimates with ISPA or ‘modified’ ISPA derived distances. Since the latter are assuredly correct, within albeit somewhat larger error bounds, large deviations directly pinpoint ‘suspect’ distance estimates.

In summary, the relaxation matrix approaches result for most distances (especially the short distances  $< 3.0$  Å) in better estimates than those obtained with a ‘modified’ ISPA approach. On the other hand, the relaxation matrix approaches may also give distances that are wrong by a large amount. These erroneous distances are difficult to pinpoint. In contrast, the simulations by van de Ven et al. [74] or Borgias and James [67] indicate that the simple ISPA (or a modified ISPA approach when applied) gives distance estimates that are assuredly correct, albeit with slightly larger error bounds. The advantage of the ISPA method is its simplicity, which allows one to easily identify those distances that are likely to be affected most by spin diffusion. In addition, one can experimentally minimize spin diffusion effects by choosing relatively short mixing times when recording the NOESY spectra. Furthermore, several factors that may affect the distance calculations are not well accounted for in the relaxation matrix equations. Libration motions may lead to  $S^2$  values varying from values of 0.8 to 0.6, thereby affecting NOEs proportionally. When considering that spin diffusion leads to a lower effective distance dependence of the NOEs, e.g. from the sixth to the fourth inverse power, this results in relative errors of 5–8% in the distances. Nucleic acids do not behave as isotropic tumbling molecules, but are rather asymmetric. This also affects the NOE calculations (see Section 8). Thus, minimal error bounds of at least 10% corresponding to  $\pm 0.2$  Å at 2 Å are called for. These issues provide another incentive for sticking to somewhat conservative distance constraints, which may as well be derived from the less precise ‘modified’ ISPA approach. We suggest therefore the use of conservative error bounds for the non-exchanging protons of:  $\pm 0.2$ – $0.3$  Å up to 2.6 Å,  $\pm 0.3$  Å from 2.6 Å to 3.3 Å,  $\pm 0.4$  Å from 3.2 Å to 4.0 Å,  $\pm 0.5$  Å from 4.0 Å



to 5.0 Å, and  $> \pm 0.6$  Å for distances  $> 5.0$  Å. For exchanging protons, one could use 0 to 5 Å, but see Schmidt and James [68]. Deriving distance constraints via a 'modified' ISPA with conservative error bounds as described above prevents on the one hand putting too tight constraints on the long distances and still providing reasonable estimates for short distances, while on the other hand it also prevents choosing unduly loose error bounds.

Are conservative or 'loose' distance constraints detrimental to the structure determination compared to 'tight' distance constraints? The answer seems to be that they are not. This has recently been elegantly demonstrated by Allain and Varani [79] using NOE data from the RNA hammerhead as a model system. They classified distances loosely into three categories, i.e. 0–2.9 Å (corresponding to  $d_{\text{est}} \pm 0.55$  Å if one considers that the shortest distance is 1.8 Å anyhow), 2.9–3.5 Å ( $d_{\text{est}} \pm 0.3$  Å), and 3.5–5.0 Å ( $d_{\text{est}} \pm 0.75$  Å). These distance ranges correspond in fact quite closely to the ones mentioned above. These ranges are used for all protons except the exchangeable or strongly overlapping protons, for which the bounds were set to 0–5.0 Å. When tightening the constraints to ranges of 1.8–2.4 Å ( $d_{\text{est}} \pm 0.3$  Å), 1.8–2.9 Å ( $d_{\text{est}} \pm 0.55$  Å), 2.5–3.5 Å ( $d_{\text{est}} \pm 0.5$  Å), and 3.5–5.0 Å ( $d_{\text{est}} \pm 0.75$  Å), the precision of the structures does not significantly improve, i.e. the rmsd of the final structures improves by 0.2 Å at best (Table 2 of Allain and Varani [79]). Van de Ven and Hilbers [80] (see also Hilbers et al. [81]) have also investigated how the precision of the distance data affects the precision of the structures. Employing all  $d_s(6/8/1'/2'/2''/3';6/8/5)$  distances and assuming that they are determined with a precision of  $\pm 0.25$  Å, they find that in a G to G dinucleotide step the twist is determined with a range of about 42° and the rise with a range of roughly 1.5 Å. For the unrealistically high precision of  $\pm 0.1$  Å, they find similar values, namely, 40° for the twist and 1.2 Å for the rise. The increased tightness of the distance ranges does not considerably improve the precision of the helix parameters. Van de Ven and Hilbers [80] and Wijmenga et al. [62] have investigated how precisely torsion angles within a nucleotide unit are determined by distances given a  $\pm 0.2$  Å uncertainty in the distances. It is found that a combination of  $d_i(6;2'/2''/3')$  distances quite accurately determine the glycosidic

torsion angles and also allows the determination of the fraction of N- or S-pucker of the sugar pucker, while the distance  $d_i(6;3')$  by itself does not determine well the fraction of N- or S-pucker. As pointed out in Ref. [62] uncertainties  $> \pm 0.5$  Å make it virtually impossible to determine the fraction of N- or S-pucker. The glycosidic torsion angle remains well determined by the distance  $d_i(6/8;2')$  even with an uncertainty as large as  $\pm 0.7$  Å. With an uncertainty of  $\pm 0.2$  Å in the distance  $d_i(3';5'/5'')$  the fraction of  $\gamma^+$  rotamer is defined, but rather loosely. On the other hand, a combination of  $d_i(6;5'/5'')$  is sufficient to define this fraction rather well [62]. Uncertainties  $> \pm 0.5$  Å make it virtually impossible to determine the fraction of  $\gamma^+$  rotamer (see Fig. 3.1 in Ref. [62]). The same error bounds for ISPA estimates derived above start from about  $\pm 0.2$  Å for distances around 2.0 Å and broaden to about  $\pm 0.5$  Å at 5.0 Å; the error in the 'loose' as well as the 'tight' constraint sets used by Allain and Varani also increases from roughly  $\pm 0.3$ –0.5 Å for short distances between 2.0 and 3.0 Å to  $\pm 0.7$  Å for distances between 3.5 and 5.0 Å. These error bounds correspond to uncertainties which fall somewhere in between the  $\pm 0.2$  Å and  $\pm 0.5$  Å range. Thus, for both sets of constraints the error bounds on the intra-residue distances are sufficiently narrow that the glycosidic torsion angle and sugar pucker, as well as the  $\gamma^+$  rotamer, can be determined reasonably well. The conclusion that one is forced to draw from these investigations is that rather conservative distance constraints (see above) can be employed without detrimental effect on the ultimate precision of the structure.

How does the number of constraints and their spread through the molecular structure affect the precision of the derived structure? As can be seen from Table 1, the spread in space of the NMR accessible distances in a helix is rather uneven. A large proportion of the total number of short distances are intra-residue (48%), a considerable percentage of which do not confer structural information. In addition, the most easily measured sequential distances, the group of H8/H6 to sugar proton distances, constitute a large part of the sequential distances (20%). Unfortunately, these sequential sugar-to-base distances only define one side of the base plane. The same applies for the base-to-base distances (6%). A large number of the sequential distances involve sequential sugar-to-sugar

distances (20%). They define the backbone quite well, but are difficult to measure since the protons involved reside in very crowded regions of the NMR spectra. Very few detectable cross-strand distances exist, except for those involving imino and amino protons and H2 to H1' protons in an A-helix. Furthermore, in contrast to  $\alpha$ -helices in proteins where one finds NH(i) to NH(i + 3) distances, in nucleic acids no such long-range distances are found. This rather uneven spread through the chemical structure of the NMR detectable distances, together with the lack of long-range distances, is expected to detrimentally affect the precision by which the helical structure can be determined and thereby the precision of torsion angles and/or stacking and helical parameters such as twist and rise. How extensive should the constraint set be in order to be able to define structural elements reasonably well?

We find in the case of a 12-mer RNA duplex with a tandem GA base pair that very high precision is achieved when about 30 'loose' distance constraints per residue are employed; the constraint set included, besides the easily accessible constraints, sequential sugar-to-sugar distances and torsion angle constraints for  $\delta$  and  $\gamma$ , but not for  $\alpha$ ,  $\beta$ ,  $\epsilon$  and  $\zeta$  [76,77]. The pairwise rmsd of the center part of the duplex consisting of four base pairs which included the tandem GA base pairs was found to be 0.6 Å. This highly precise structure allowed the determination of the base pairing of the tandem GA base pair (no base pair constraints were applied for the tandem GA part of the structure). This suggests that the inclusion of the sequential sugar-to-sugar distances is quite important. Gorenstein and co-workers [71] also found that inclusion of sequential sugar-to-sugar distances gives considerable improvement in the definition of the derived structure. Van de Ven and Hilbers [80] (see also Ref. [81]) have investigated how the spread of the distance constraints through the molecule affects the precision of the derived structure. When using distance constraints with a precision of  $\pm 0.25$  Å they find that in a G to G dinucleotide step the twist is determined with a range of  $42^\circ$  and the rise with a range of roughly 1.5 Å (see above). Better results are obtained when the dinucleotide is C to G; the range for the twist is then  $22^\circ$  and for the rise 0.8 Å. This improved definition results from a better spatial spread of the protons in the CG step as

compared to that in the GG step. In the CG dinucleotide one finds distances involving H6 and H5 protons, whereas in the GG dinucleotide only distances involving H8 protons are found. Similarly, the systematic study of Allain and Varani [79], using the hammerhead as a model system, shows that inclusion of loose constraints (see above) involving exchangeable protons greatly increases the definition, while adding sugar-to-sugar constraints and torsion angles gives further improvement but to a lesser extent. For the so-called realistic 'loose' constraint set they find approximate ranges of  $10^\circ$  for the twist and 0.4 Å for the rise (see above for the distance ranges; this set contains no sugar-to-sugar constraints but includes constraints involving exchangeable protons apart from sugar-to-base and base-to-base constraints). These ranges are narrower than those found by van de Ven and Hilbers, which again demonstrates the advantageous effect of a larger number of constraints on the precision. Note that van de Ven and Hilbers considered isolated G to G or C to G dinucleotide steps, while in the study of Allain and Varani the bases are part of a helix, which limits the allowed conformational space. This also demonstrates that the number of constraints required to obtain high structural definition is quite context dependent. In loop regions, and bulge regions where no base pair constraints are present, either the precision will be lower with the same number of constraints or a larger number of constraints is required to achieve the same high level of precision as in a helix. Finally, we note that the in-depth studies of James and coworkers [68,69] and of Luxon and Gorenstein [72] on DNA duplexes show that using a large number ( $> 10$ /residue) of rather precise constraints (see above), which are well spread through the molecular structure, leads to highly defined NMR derived structures. It appears to be possible to distinguish sequence dependent structural effects.

#### 4.5. Conclusion

For the precision of NMR derived structures, the number of (structurally relevant) distance constraints is more important than the precision of constraints. The number of structurally relevant constraints should be around 15 to 30 per residue. They should as far as possible be uniformly spread through the

Table 2

One-bond coupling constants  $^1J_{\text{HC}}$  (Hz) in the base and sugar moieties of  $^{13}\text{C}/^{15}\text{N}$  labeled 5'-AMP, 5'-GMP, 5'-UMP and 5'-CMP

$^1J_{\text{CH}}$	5'-AMP <sup>a</sup>	5'-GMP <sup>b</sup>	5'-UMP <sup>a</sup>	5'-CMP <sup>b</sup>
H8	215.9	216.0		
H2	203.2			
H6			184.9	184.1
H5			178.7	175.8
H1'	166.5	166.1	170.2	169.5
H2' <sup>c</sup>	150.9	151.4	151.7	151.2
H3' <sup>c</sup>	152.6	152.0	152.2	151.3
H4'	150.9	151.1	150.3	148.9
H5' <sup>d</sup>	148.3	146.9 <sup>e</sup>	148.3	147.2 <sup>f</sup>
H5'' <sup>d</sup>	145.2	146.9 <sup>e</sup>	144.9	145.0 <sup>f</sup>

<sup>a</sup> average values derived from 1D  $^1\text{H}$  and  $^{13}\text{C}$  spectra recorded at 500 and 700 MHz.<sup>b</sup> average values derived from 1D  $^{13}\text{C}$  spectrum recorded at 600 MHz.<sup>c</sup> in cyclic nucleotides, which have a defined sugar conformation, one finds for N- and S-type sugars  $^1J_{\text{H2'C2'}}$  values of 158.4 Hz and 149.2 Hz, respectively, and for  $^1J_{\text{H3'C3'}}$  values of 149.6 Hz and 156.8 Hz, respectively. The error in the coupling constant values is given to be about  $\pm 0.3$  Hz [49].<sup>d</sup> for 5'-AMP, 5'-GMP, 5'-UMP and 5'-CMP the  $\gamma$  torsion angle is a mixture of rotamers [49] 2209]; in the cyclic nucleotides,  $\alpha(\text{pApA})$ ,  $\alpha(\text{pGp(dG)})$ , the  $\gamma$  torsion angle is in a defined conformation, gauche + ; the  $^1J_{\text{H5'CS'}}$  and  $^1J_{\text{H5''CS'}}$  are then 151 Hz and 141 Hz, respectively.<sup>e</sup> near isochronous resonances H5' and H5''; value given is from proton-coupled 1D  $^{13}\text{C}$  spectrum.<sup>f</sup> derived from 1D  $^1\text{H}$  spectrum at 400 MHz. Table adapted from Ref. [49].

chemical structure of the molecule, i.e. rather than including a large number from one category (e.g. intra-nucleotide), constraints should be derived in such a way that each category is represented (intra-nucleotide and its subdivisions or sequential and cross-strand inter-nucleotide and its subdivisions, see Table 1). In view of the uncertainties in the distance determination from NOE data, precise distances are prone to error and rather conservative distance ranges should be used for structure refinement.

## 5. $J$ -couplings

With the development of  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling techniques for nucleic acids a large number of

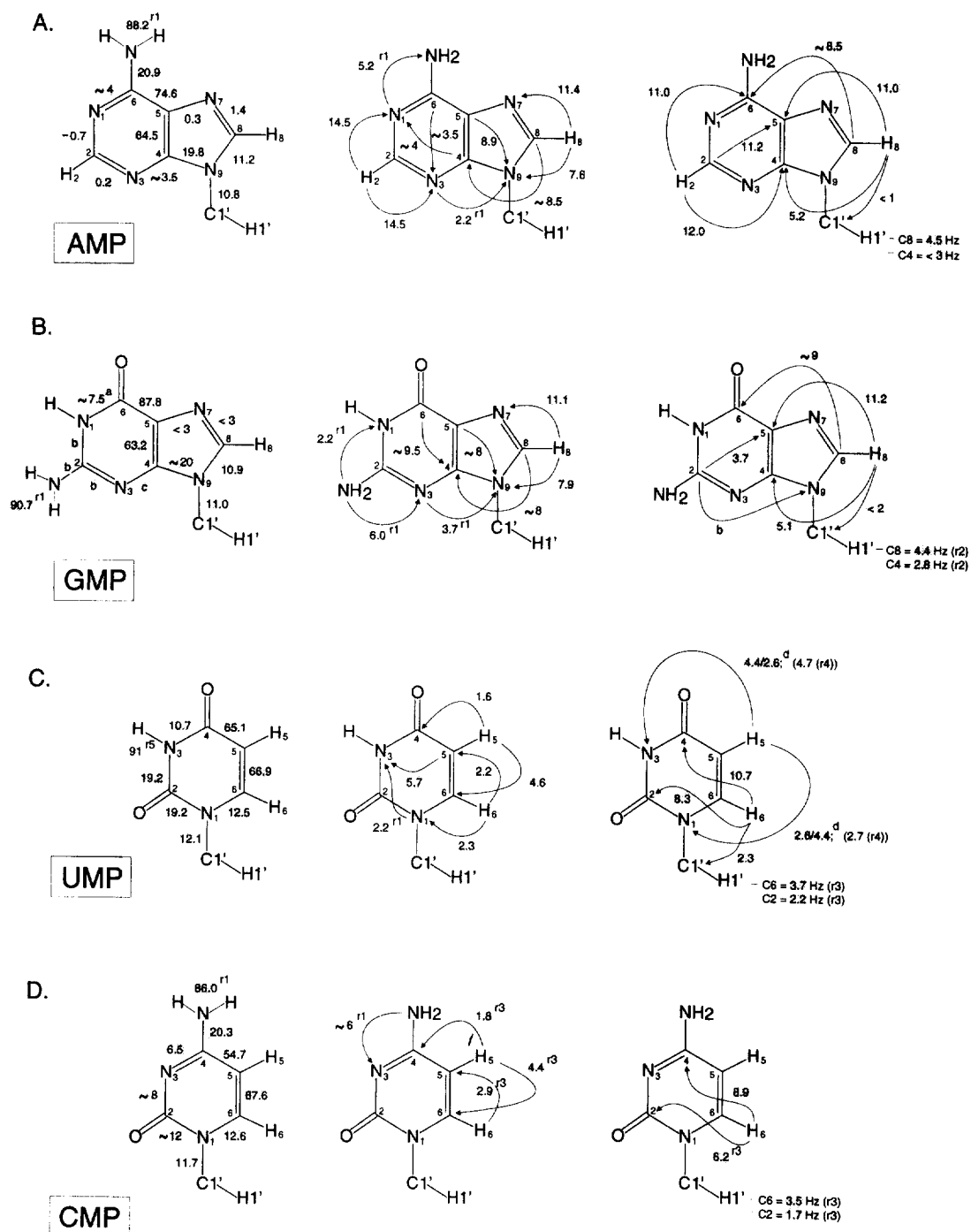
new heteronuclear  $J$ -couplings have become accessible. Knowledge of the values of  $J$ -coupling constants is essential for the rational design and application of resonance assignment techniques based on through-bond coherence transfer. Furthermore, these couplings provide important additional parameters for the determination of torsion angles. In this section we present an overview of these heteronuclear  $J$ -couplings as well as of proton–proton  $J$ -coupling constants, their relation to torsion angles, and describe NMR techniques for determining their values. The  $J$ -couplings are discussed according to their torsion angle dependence, except for the  $J$ -couplings in the bases and the  $^1J_{\text{CH}}$ -couplings which are discussed separately.

Table 3

One-bond coupling constants  $^1J_{\text{CC}}$  (Hz) in the base and sugar moieties of  $^{13}\text{C}/^{15}\text{N}$  labeled 5'-AMP, 5'-GMP, 5'-UMP and 5'-CMP

$^1J_{\text{CC}}$	5'-AMP <sup>a</sup>	5'-GMP <sup>a</sup>	5'-UMP <sup>a</sup>	5'-CMP <sup>a</sup>
C1'-C2'	42.2	42.6	43.0	43.4
C2'-C3'	38.1	37.8	37.8	37.4
C3'-C4'	38.0	38.3	38.5	38.7
C4'-C5'	42.3	42.9	42.9	43.0

<sup>a</sup> average values derived from 1D  $^{13}\text{C}$  spectra recorded at 500 and 700 MHz (5'-AMP and 5'-UMP) from 1D  $^{13}\text{C}$  spectrum recorded at 600 MHz (5'-GMP and 5'-CMP). The error in the coupling constant values is given to be about  $\pm 0.4$  Hz [49]. Table adapted from Ref. [49].



### 5.1. $^1J_{\text{HC}}$ - and $^1J_{\text{CC}}$ -couplings

The  $^1J_{\text{CH}}$  and  $^1J_{\text{CC}}$ -couplings are given in Table 2 and Table 3, respectively [49]. To a large extent these  $J$ -couplings are independent of conformation, but have often indicated a *small conformation dependent variation*. For example, the  $^1J_{\text{C1'H1'}}$ -couplings have been found to convey  $\chi$  torsion angle information [49,82]. The  $J$ -couplings,  $^1J_{\text{H3'C3'}}$  and  $^1J_{\text{H2'C2'}}$ , depend on the puckering [49]. Finally, the  $^1J_{\text{H5'CS'}}$  and  $^1J_{\text{H5''CS'}}$  have distinctly different values, namely 151 Hz and 141 Hz, respectively, when the torsion angle  $\gamma$  is in the gauche + domain; these differences are retained when the rotamer population is not purely gauche + [49]. Thus, the latter  $J$ -couplings can be used for stereospecific assignment of the H5' and H5'' resonances. These conformational dependencies are extremely useful, especially since the  $^1J_{\text{HC}}$ -couplings are very easy to measure. They will be discussed in more detail below in the appropriate sections.

### 5.2. Overview of $J$ -couplings in the bases

A complete overview of the homonuclear and heteronuclear  $J$ -couplings, found in the bases A, G, U and C, involving both newly measured as well as redetermined  $J$ -couplings, has recently been published by Ippel et al. [49]. They are summarized here in Fig. 3 (and Tables 2 and 3). As can be appreciated from Fig. 3 (and also Tables 2 and 3), the homonuclear and heteronuclear  $J$ -couplings in the bases form a complex network. Consequently, such a detailed, accurate, and complete set of these

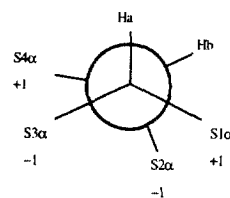
$J$ -couplings such as given here should be of great value for the rational design and improvement of through-bond assignment strategies discussed in Section 7.

### 5.3. Ribose sugar

Traditionally, the sugar pucker has been determined mainly from the  $^3J_{\text{HH}}$ -couplings in the sugar ring. They can be derived from the generalized Karplus equation [62,83,84] which is given by:

$$^3J_{\text{HH}} = P1 \cos^2 \phi + P2 \cos \phi + \sum_{i=1}^4 (P3 + P4 \cos^2(\zeta_i \phi + P5|\Delta\chi_i|)) \Delta\chi_i + \Delta J \quad (12)$$

$$\Delta\chi_i = \Delta\chi_{i,\alpha} + P6 \sum_{j=1}^3 \Delta\chi_{ij,\beta} \quad (13)$$



Scheme 1.

Here  $\Delta\chi_{i,\alpha}$  and  $\Delta\chi_{ij,\beta}$  are the difference in Huggins electronegativity between hydrogen and the  $\alpha$  and  $\beta$  substituents in the  $\text{HaS1}\alpha(\text{S11}\beta, \text{S12}\beta, \text{S13}\beta)$

Fig. 3. From left to right,  $^1J$ -,  $^2J$ - and  $^3J$ -coupling constants (in Hz) in the bases of 5'-AMP, 5'-GMP, 5'-UMP and 5'-CMP. This figure is adapted from a similar one in Ref. [49]. All the given  $J$ -coupling values were (re)determined in that study, except those indicated with r1, r2, r3, r4 and r5, which were taken from Refs. [231,112,259,233,232], respectively.  $J$ -couplings for which the assignment was, in Ref. [49], found to be ambiguous are indicated by the lower case letters a, b, c and d, with: (a) Two equal  $^6J_{\text{CN}}$ -couplings were found on C6, with a sum value of 14.9 Hz; assigning 7.5 Hz to  $^1J_{\text{C6N1}}$  leaves  $\sim 7.5$  Hz for the other  $J_{\text{C6N}}$ -coupling, which can arise either from  $^3J_{\text{C6N2}}$  or  $^3J_{\text{C6N3}}$  or  $^2J_{\text{C6N7}}$  or  $^3J_{\text{C6N9}}$ . (b) The C2 resonance exhibits a multiplet pattern consisting of 15 lines (intensity, 1:2:2:2:2:3:4:3:2:2:2:2:1) in the 1D ( $^1\text{H}$ )-decoupled  $^{13}\text{C}$  spectrum. Since the  $^3J_{\text{C2CS}}$  has a value of 3.7 Hz, the observed multiplet can be simulated by incorporating four additional  $J$ -couplings with values of approximately 3.9, 7.6, 15.2 and 23 Hz; the 23 Hz coupling can be assigned to  $^1J_{\text{C2N2}}$  on the basis of a similarly large value for the corresponding coupling in CMP and AMP; the  $J$ -couplings of 7.6 and 15.2 Hz can tentatively be assigned to  $^1J_{\text{C2N1}}$  and  $^1J_{\text{C2N3}}$ , respectively, or the reverse; the smallest of these four  $J$ -couplings, i.e. 3.9 Hz, can then be assigned to  $^3J_{\text{C2N9}}$ , assuming that the four-bond coupling,  $^4J_{\text{C2N7}}$ , is undetectably small. (c) Computer simulation of the G C4 resonance multiplet in the 1D ( $^1\text{H}$ )-decoupled  $^{13}\text{C}$  spectrum yields two  $J_{\text{CN}}$ -coupling constants of approximately 20 and 8.5 Hz, respectively. The largest coupling constant can be assigned to  $^1J_{\text{C4N9}}$  in analogy to that in 5'-AMP; the  $J$ -coupling of 8.5 Hz can be assigned to either one of the couplings  $^1J_{\text{C4N3}}$ ,  $^2J_{\text{C4N7}}$ ,  $^3J_{\text{C4N1}}$  or  $^3J_{\text{C4N2}}$ . (d) Computer simulation of the H5 resonance multiplet in the fully coupled 1D  $^1\text{H}$  spectrum yielded two more not yet assigned values, namely 2.6 and 4.4 Hz, which can be attributed to  $^3J_{\text{H5N1}}$  and  $^3J_{\text{H5N3}}$ , respectively, or the reverse.

S3 $\alpha$ (S31 $\beta$ S32 $\beta$ S33 $\beta$ ) -C1-C2- S2 $\alpha$ (S21 $\beta$ S22 $\beta$ S23 $\beta$ ) S4 $\alpha$ (S41 $\beta$ S42 $\beta$ S43 $\beta$ )Hb fragment;  $\Delta\chi_{i,\alpha}$ ,  $\Delta\chi_{ij,\beta} = 1.3$  (O), 0.4 (C), 0.85 (N) and  $-0.05$  (P) (Scheme 1). The parameter  $\zeta_i$  is  $+1$  or  $-1$  depending on the orientation of the substituent as indicated. The parameters  $P1$  to  $P6$  depend on the number of non-hydrogen  $\alpha$  substituents:

# $\alpha$ substituents	$P1$	$P2$	$P3$	$P4$	$P5$	$P6$
2	13.89	-0.96	1.02	-3.40	14.9	0.24
3	13.22	-0.99	0.87	-2.46	19.9	0.00
4	13.24	-0.91	0.53	-2.41	15.5	0.19
overall	13.70	-0.73	0.56	-2.47	16.9	0.14

From this generalized Karplus equation it follows that the  $^3J_{\text{HH}}$ -couplings in the ribose ring are given by:

$$^3J_{\text{HH}} = 13.22 \cos^2 \phi - 0.99 \cos \phi + \sum_{i=1}^4 (0.87 - 2.46 \cos^2(\zeta_i \phi + 19.1 |\Delta\chi_i|)) \Delta\chi_i + \Delta J \quad (14)$$

$$\Delta\chi_i = \Delta\chi_{i,\alpha} \quad (15)$$

(for  $J_{1'2'}$ ,  $J_{1'2''}$ ,  $J_{2'3'}$  and  $J_{2'3''}$ )

$$^3J_{3'4'} = 13.24 \cos^2 \phi - 0.91 \cos \phi + \sum_{i=1}^4 (0.53 - 2.41 \cos^2(\zeta_i \phi + 15.5 |\Delta\chi_i|)) \Delta\chi_i + \Delta J \quad (16)$$

$$\Delta\chi_i = \Delta\chi_{i,\alpha} - 0.19 \sum_{j=1}^3 \Delta\chi_{ij,\beta} \quad (17)$$

In addition, a correction for the so-called Barfield transmission effect is required for  $J_{1'2''}$  and  $J_{2'3''}$ :

$$\Delta J_{1'2''} = -2.0 \cos^2(P - 234) \quad 144^\circ < P < 324^\circ \quad (18)$$

$$\Delta J_{2'3''} = -0.5 \cos^2(P - 288) \quad 180^\circ < P < 360^\circ; \quad 0^\circ < P < 18^\circ \quad (19)$$

The torsion angle  $\phi$  in the above equations can be

related to the pseudorotation angle  $P$  and pucker amplitude  $\varphi_m$  via [62]:

$$\phi_{1'2'} = 121.4 + 1.03\varphi_m \cos(P - 144) \quad (20)$$

$$\phi_{1'2''} = 0.9 + 1.03\varphi_m \cos(P - 144)$$

$$\phi_{2'3'} = 2.4 + 1.03\varphi_m \cos(P)$$

$$\phi_{2'3''} = 121.9 + 1.03\varphi_m \cos(P)$$

$$\phi_{3'4'} = 124.0 + 1.03\varphi_m \cos(P + 144)$$

The values of  $^3J_{1'2'}$  to  $^3J_{3'4'}$  are given as a function of pseudorotation angle  $P$  and pucker amplitude  $\varphi_m$  in Fig. 4(A)–(E). Ribose sugar rings are not rigid but interconvert rapidly between N- and S-type conformations. These sugar puckering states are described by their pseudorotation angles and amplitudes, namely,  $P^N$  and  $\varphi_m^N$ , for the N-state, and  $P^S$  and  $\varphi_m^S$ , for the S-state. The relative population can be found via the fraction S-conformer,  $pS$ . The parameters  $\varphi_m^N$  and  $\varphi_m^S$  tend to be fairly constant and range from 32 to 40, while  $P^N$  ranges from  $-10^\circ$  to  $20^\circ$  and  $P^S$  lies between  $120^\circ$  and  $180^\circ$ . The  $^3J_{\text{HH}}$ -couplings in the ring are then the weighted average of the  $^3J_{\text{HH}}$ -couplings in the two conformations:

$$^3J_{\alpha\beta}^{\text{av}} = (1 - pS) \cdot ^3J_{\alpha\beta}^N + pS \cdot ^3J_{\alpha\beta}^S \quad (21)$$

Thus,  $^3J_{\alpha\beta}^N$  and  $^3J_{\alpha\beta}^S$  depend on  $P$  and  $\varphi_m$  in their respective states. Although  $P^N$ ,  $P^S$ ,  $\varphi_m^N$  and  $\varphi_m^S$  can be determined in principle from the complete set of ribose  $^3J_{\text{HH}}$ -couplings, in practice, these values are assumed to be known for the least populated state, and to correspond to the middle values of  $P$  and  $\varphi_m$  in the N- or S-puckered state, i.e.  $P^N = 10$  and  $\varphi_m^N = 35$ ,  $P^S = 160$  and  $\varphi_m^S = 35$ . For the most highly populated state it is thus possible to derive  $P$  and  $\varphi_m$ . We finally note that a straightforward check whether the assumption of an equilibrium between N-type and S-type conformers applies, is to plot  $^3J_{3'4'}$  against  $^3J_{1'2'}$ . These two  $J$ -couplings have a reverse dependence on the fraction of N-type sugar, which is incompatible with a  $P$  value intermediate between N-type and S-type sugars [62,80]. The most commonly used program to derive  $P$ ,  $\varphi_m$  and  $pS$  from  $^3J_{\text{HH}}$ -couplings is PSEUROT, developed by van den Hoogen et al. [85]. Alternatively, one can add NOE information to determine the puckering state [62,64].

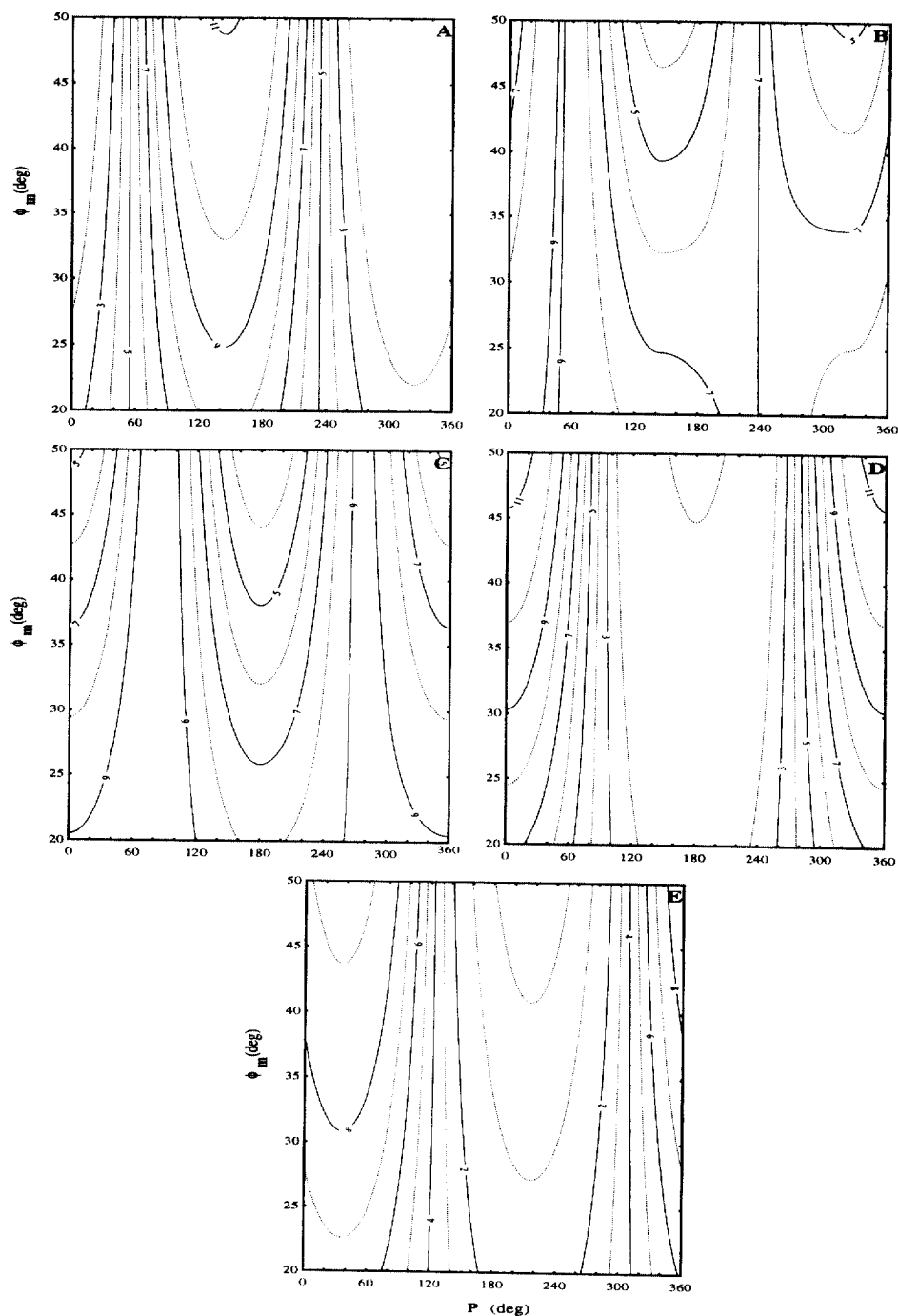


Fig. 4. (A)–(E) Contour lines of  $^3J_{HH}$ -coupling constants (Hz) in the sugar ring as a function of  $P_m$  and  $\phi_m$ . The coupling constants were calculated with the aid of the EOS-Karplus equation and corrected for the Barfield transmission effect (see text). (A)  $^3J_{1'2'}$ , (B)  $^3J_{1'2'}$ , (C)  $^3J_{2'3'}$ , (D)  $^3J_{2'3'}$  and (E)  $^3J_{3'4'}$ .

Conte et al. [64] use a program that incorporates both  $^3J_{\text{HH}}$ -couplings and NOE information. Finally, as will be discussed in more detail in Section 5.3, the  $^3J_{\text{HH}}$ -couplings can also be derived from TOCSY data. In the program developed by van Duynhoven et al. [86] the puckering state, i.e.  $P$ ,  $\varphi_m$  and  $pS$ , are directly derived from the (H,H) TOCSY cross peak intensities.

Thanks to the availability of labeled compounds,  $^3J_{\text{HC}}$ -couplings can be used as additional parameters to determine the sugar pucker. Ippel et al. [49] have recently determined parameters for the Karplus relations between  $^3J_{\text{H}3'\text{C}1'}$ ,  $^3J_{\text{H}1'\text{C}3'}$ ,  $^3J_{\text{H}2'\text{C}4'}$  and  $^3J_{\text{H}4'\text{C}2'}$  and the sugar puckering. Since the ribose ring is in rapid equilibrium between N- and S-conformers, the  $^3J_{\text{HC}}$ -couplings are the weighted average of the  $^3J_{\text{HC}}$ -couplings in the two conformations:

$$^3J_{\alpha\beta} = pS \cdot ^3J_{\alpha\beta}^{\text{S}} + (1 - pS) \cdot ^3J_{\alpha\beta}^{\text{N}} \quad (22)$$

Here  $^3J_{\alpha\beta}^{\text{S}}$  and  $^3J_{\alpha\beta}^{\text{N}}$  are the  $^3J_{\text{HC}}$ -coupling values in the S- and N-puckered state, respectively, i.e.  $^3J_{\text{H}3'\text{C}1'}^{\text{S}} = 6.3$  Hz,  $^3J_{\text{H}3'\text{C}1'}^{\text{N}} = 1.4$  Hz,  $^3J_{\text{H}2'\text{C}4'}^{\text{S}} = 0.3$  Hz,  $^3J_{\text{H}2'\text{C}4'}^{\text{N}} = 4.2$  Hz,  $^3J_{\text{H}1'\text{C}3'}^{\text{S}} = 0.3$  Hz,  $^3J_{\text{H}1'\text{C}3'}^{\text{N}} = 2.7$  Hz,  $^3J_{\text{H}4'\text{C}2'}^{\text{S}} = 1.1$  Hz and  $^3J_{\text{H}4'\text{C}2'}^{\text{N}} = 1.0$  Hz. Thus, this relation provides another opportunity to determine the sugar puckering. Note that  $^3J_{\text{H}3'\text{C}1'}$  and  $^3J_{\text{H}2'\text{C}4'}$  differ strongly between the N- and S-state and can thus be used quite well to determine  $pS$ . The three-bond couplings,  $^3J_{\text{H}1'\text{C}3'}$  and  $^3J_{\text{H}4'\text{C}2'}$ , remain small for both the S- and N-puckered states and are thus less useful indicators of the sugar puckering. We finally note that the  $^3J_{\text{H}3'\text{C}5'}$ -coupling, in principle, also monitors the puckering, namely via the torsion angle  $\delta$ . Unfortunately, its value does not differ significantly between the N- and S-state [49].

A third source of information on the puckering state of the sugar ring is the  $^2J_{\text{HC}}$ -couplings. The two-bond couplings  $^2J_{\text{H}2'\text{C}1'}$  and  $^2J_{\text{H}3'\text{C}2'}$  are found to be negative for S-puckered rings and positive for N-puckered rings, while the reverse holds for  $^2J_{\text{H}2'\text{C}3'}$  and  $^2J_{\text{H}3'\text{C}4'}$  [49]. These two-bond  $J$ -couplings are thus good indicators of the sugar ring conformation. However, the other sugar ring two-bond couplings,  $^2J_{\text{H}1'\text{C}2'}$  and  $^3J_{\text{H}4'\text{C}3'}$ , do not change sign and remain negative for both the N-puckered and the S-puckered state [49]. These signs of the  $^2J_{\text{HC}}$ -couplings are in complete accordance with the projection rule proposed by Bock and Pederson [87]. According to this rule the

sign of the  $^2J_{\text{HC}}$ -couplings in H-C- $^{13}\text{C}$ -X1(X2) systems can simply be predicted from the Newman projection and the orientation of the electronegative substituents X1 and X2 relative to the position of the proton coupled to  $^{13}\text{C}$ ; X1 and X2 can either be oxygen or nitrogen. A trans orientation of this H relative to X leads to a positive value for  $^2J_{\text{CH}}$ , whereas a gauche orientation gives rise to a negative value for  $^2J_{\text{CH}}$ . If two X substituents are present on  $^{13}\text{C}$  the effects are additive. Thus, if both H to X1 and H to X2 are oriented gauche to each other a negative value is obtained for  $^2J_{\text{HC}}$ . On the other hand, a trans orientation of H relative to X1 and a gauche orientation of H relative to X2 will give rise to compensating effects and a  $^2J_{\text{HC}}$  value which is approximately zero. For example,  $^2J_{\text{H}2'\text{C}3'}$  is predicted by this rule to be negative for an N-conformer, since O3' is oriented gauche with respect to H2', and positive for an S-conformer, since O3' is oriented trans with respect to H2'. This is exactly what is observed experimentally (see above). The other  $^2J_{\text{HC}}$ -couplings in the sugar ring also follow this projection rule.

Finally, the one-bond couplings  $^1J_{\text{CH}}$  also convey sugar pucker information;  $^1J_{\text{H}3'\text{C}3'}$  has values of 149.6 Hz and 156.8 Hz for N- and S-type sugars, respectively, while  $^1J_{\text{H}2'\text{C}2'}$  shows the reverse trend, namely values of 158.4 Hz and 149.2 Hz for N- and S-type sugars, respectively. When the sugar ring occurs as a mixture of N- and S-conformers with intermediate values for  $pS$ , intermediate  $^1J_{\text{H}2'/3'\text{C}2'/3'}$  values are also found. We note that the torsion angle dependencies of the  $^1J_{\text{HC}}$ -couplings are not very well understood, in contrast to the case of the  $^3J_{\text{HC}}$ - and the  $^2J_{\text{HC}}$ -couplings. Nevertheless these experimental observations remain useful as indicators of the sugar pucker.

#### 5.4. Determination of the $\beta$ torsion angle

The  $\beta$  torsion angle has traditionally been determined by the  $^3J_{\text{HP}}$ -coupling constants. The Karplus equation describing the  $^3J_{\text{H}5'/5''\text{P}5'}$ -couplings is

$$^3J_{\text{H}5'/5''\text{P}5'} = 15.3 \cos^2 \phi - 6.2 \cos \phi + 1.5 \quad (23)$$

with  $\phi = \beta - 120^\circ$  for H5' and  $\phi = \beta + 120^\circ$  for H5''. The most recent parametrization has been used here [62,88]. The  $^3J_{\text{H}5'\text{P}5'}$ - and  $^3J_{\text{H}5''\text{P}5'}$ -couplings describe the  $\beta$  torsion angle as shown in Fig. 5. As can be



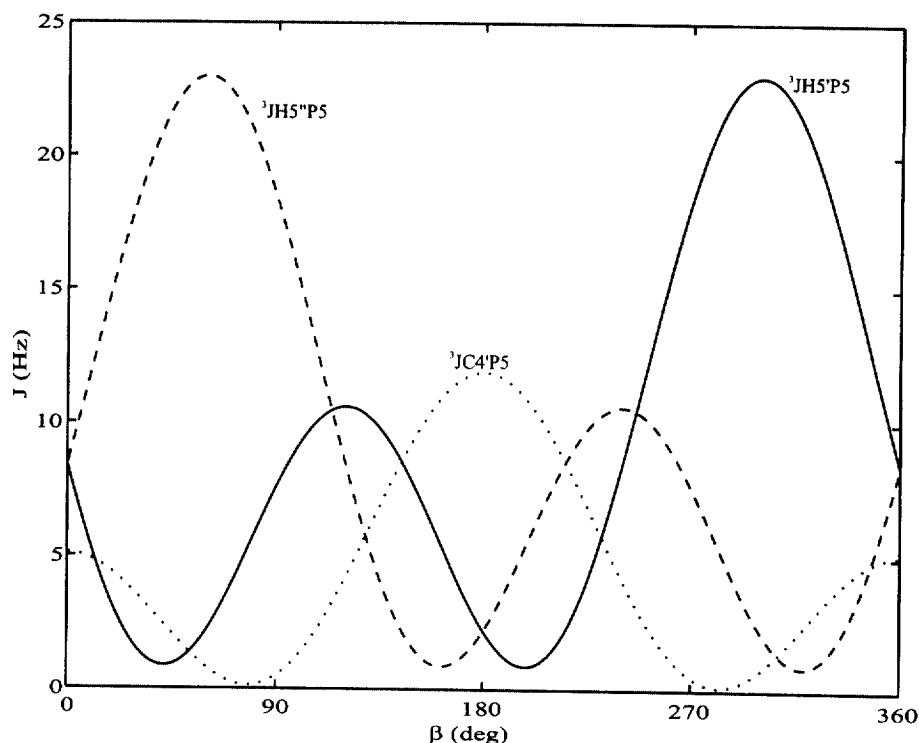


Fig. 5. The  ${}^3J_{\text{H5}'\text{P5}'}$ ,  ${}^3J_{\text{H5}''\text{P5}}$  and  ${}^3J_{\text{C4}'\text{P5}}$ -coupling constants calculated as a function of the torsion angle  $\beta$  on the basis of their Karplus relations (see text).

seen, these two  $J$ -coupling constants define the  $\beta$  torsion angle quite well for the whole range of  $\beta$  values.

The  ${}^3J_{\text{C4}'\text{P5}'}$ -couplings are described by

$${}^3J_{\text{C4}'\text{P5}'} = 8.0 \cos^2 \beta - 3.4 \cos \beta + 0.5 \quad (24)$$

Here again the latest parametrization, as given by Mooren et al. [88], is used. Fig. 5 also shows  ${}^3J_{\text{C4}'\text{P5}'}$  as a function of  $\beta$ . As can be seen, a fairly narrow range of  $\beta$  torsion angles can be derived from the additional knowledge of this  $J$ -coupling constant.

Whether the flexibility plays a role or not can be established by the concerted use of the different  $J$ -coupling constants. If the three  $J$ -coupling values do not indicate one value for the torsion angle, it can be concluded that angular averaging is taking place. The angular averaging can be either the result of libration motions within one rotamer domain or rapid inter-conversion between different rotamers. It may even be possible under certain circumstances and making certain assumptions, for example that only one

rotamer fraction is populated, to establish the range of allowed torsion angles for that rotamer, i.e. the width of the libration motion. But usually one tries to derive from the  $J$ -couplings the relative rotamer populations. An estimation of the fraction of the most common conformer,  $\beta^1$ , can be obtained from the equation

$$f_1 = \frac{[25.5 - (J_{\text{H5}'\text{P5}'} + J_{\text{H5}''\text{P5}})]}{20.5} \quad (25)$$

This equation has been derived under the assumption that  $\beta$  torsion angle values of the  $\beta^+$  and  $\beta^-$  conformers are  $60^\circ$  and  $300^\circ$ , respectively. The coupling values  $J_{\text{H5}'\text{P5}'}$  and  $J_{\text{H5}''\text{P5}}$  of these rotamers are then known. The value of  $J_{\text{C4}'\text{P5}'}$  can also be used to calculate this fraction in a similar way:

$$f_1 = \frac{J_{\text{C4}'\text{P5}'} - 1.3}{9.8} \quad (26)$$

Finally, the  ${}^4J_{\text{H4}'\text{P5}'}$ -coupling can reach values as large

as  $\sim 1.5$ – $2.5$  Hz if  $\beta$  falls in the trans region and the torsion angle  $\gamma$  in the gauche + region, i.e. when the molecular fragment  $\text{H4}'\text{--C4}'\text{--C5}'\text{--O5}'\text{--P5}'$  has a close to planar W-shaped conformation; for other conformations a much smaller value of the  $^4J_{\text{H4}'\text{P5}'}$ -coupling will be found.

### 5.5. Determination of the $\epsilon$ torsion angle

The  $^3J_{\text{H3}'\text{P3}'}$ -coupling describes the  $\epsilon$  torsion angle via the Karplus equation

$$^3J_{\text{H3}'\text{P3}'} = 15.3 \cos^2(\epsilon + 120^\circ) - 6.2 \cos(\epsilon + 120^\circ) + 1.5 \quad (27)$$

Here the parametrization by Mooren et al. is used [88]. The torsion angle  $\epsilon$  can for steric reasons only take values between approximately  $170^\circ$  and  $300^\circ$ , depending somewhat on the pucker of the sugar ring [88,89]. The usual value for  $\epsilon$  in A-DNA, B-DNA and A-RNA is trans, but as has been shown by Schroeder

et al. [90], B-DNA can be either in a BI or a BII conformation, characterized by  $\epsilon$  ranging from  $160$ – $220^\circ$  and from  $260$ – $300^\circ$ , respectively. These ranges fall within the sterically allowed region for  $\epsilon$  in DNA for both C2'-endo and C3'-endo sugars [88]. For C2'-endo these ranges correspond to energy minima at  $\epsilon \approx 180^\circ$  and  $\epsilon \approx 288^\circ$  [91], while for C3'-endo sugar puckers only a small range of  $\epsilon$  values centered around  $195^\circ$  is energetically favorable. For RNA, Mooren et al. [88] have shown that  $\epsilon$  has a narrower range, namely  $170^\circ < \epsilon < 280^\circ$ , for C2'-endo sugar puckers than for the usual C3'-endo sugar puckers,  $185^\circ < \epsilon < 280^\circ$ . Also for RNA a maximum tends to exist around  $\epsilon \approx 210^\circ$ . With these data in mind it is interesting to consider the Karplus curve of the  $^3J_{\text{H3}'\text{P3}'}$ -coupling versus the torsion angle  $\epsilon$  as shown in Fig. 6. As can be seen the two expected regions,  $160$ – $220^\circ$  and  $260$ – $300^\circ$ , have quite similar  $^3J_{\text{H3}'\text{P3}'}$ -coupling constants. Thus, determination of the  $^3J_{\text{H3}'\text{P3}'}$ -coupling constant is of limited value for determining the torsion angle  $\epsilon$ .

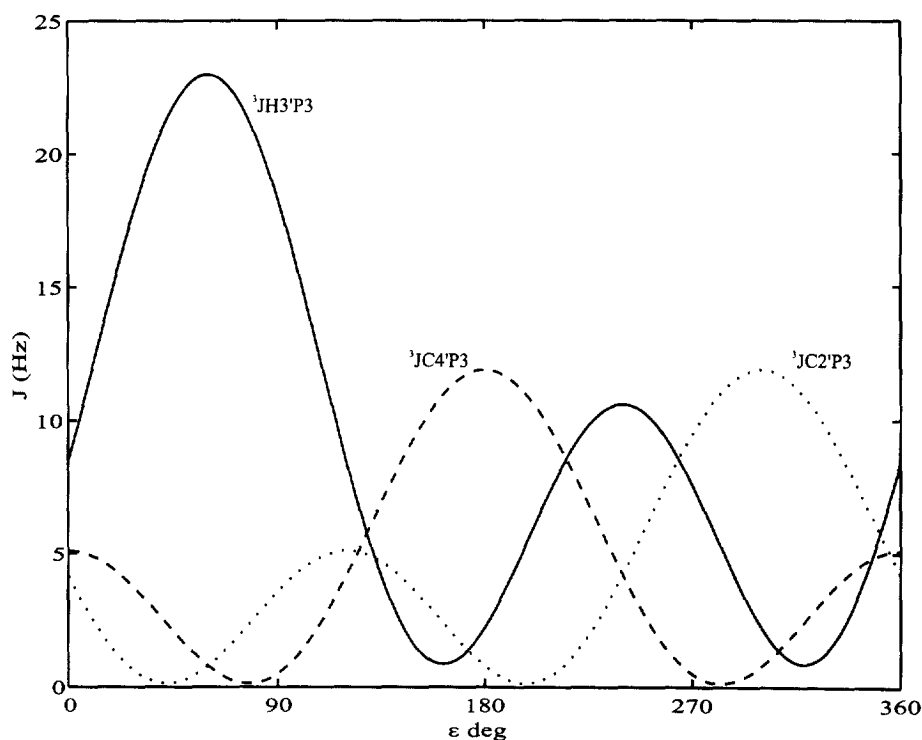


Fig. 6. The  $^3J_{\text{C4}'\text{P3}'}$ ,  $^3J_{\text{C2}'\text{P3}'}$  and  $^3J_{\text{H3}'\text{P3}'}$ -coupling constants calculated as a function of the torsion angle  $\epsilon$  on the basis of their Karplus relations (see text).

Fig. 6 also gives  ${}^3J_{C4'P3'}$  and  ${}^3J_{C2'P3'}$  as a function of  $\varepsilon$  as derived from the Karplus equation

$${}^3J_{C4'/C2'P3'} = 8.0 \cos^2 \phi - 3.4 \cos \phi + 0.5 \quad (28)$$

with  $\phi = \varepsilon$  for C4' and  $\phi = \varepsilon - 120^\circ$  for C2'. Here again the latest parametrization as given by Mooren et al. [88] is used. As can be seen,  ${}^3J_{C4'P3'}$  and  ${}^3J_{C2'P3'}$  together define the  $\varepsilon$  torsion angle quite well for the trans range as well as for the gauche range. An even

narrower determination can be made when these  $J$ -couplings are used together with the  ${}^3J_{H3'P3'}$ -coupling. In addition, there is the  ${}^4J_{H2'P3'}$ -coupling, which has a low value when  $\varepsilon$  is in the trans region, while its value is  $\approx 2.0$ – $3.0$  Hz when  $\varepsilon$  is in the gauche region. The  ${}^4J_{H2'P3'}$ -coupling can therefore further confirm the presence of a gauche rotamer for  $\varepsilon$ .

As for the torsion angle  $\beta$ , the additional  $J$ -couplings make it also possible to determine the

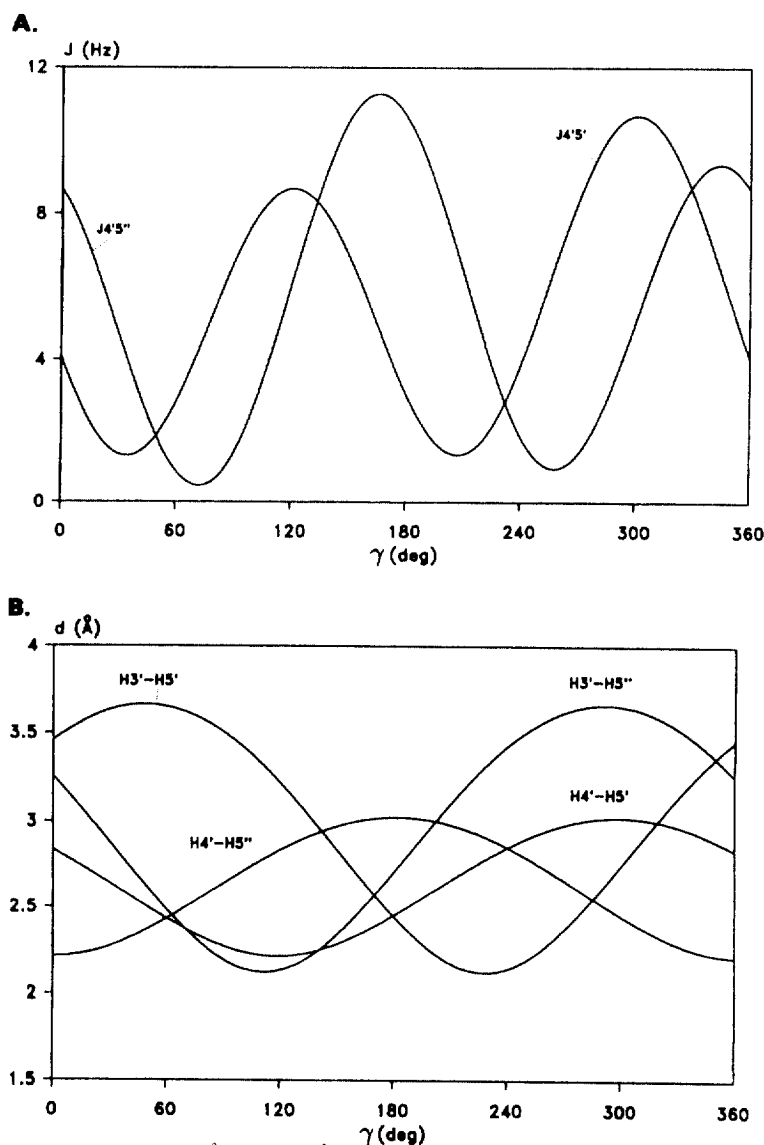


Fig. 7. The  $\gamma$  torsion angle dependence of: (A) The  ${}^3J_{H4'H5'}$  and  ${}^3J_{H4'H5''}$ -coupling constants (calculated according to their Karplus relations (see text)); (B) The  ${}^1H$ - ${}^1H$  distances,  $d_i(3';5')$ ,  $d_i(3';5'')$ ,  $d_i(4';5')$  and  $d_i(4';5'')$ .

rotamer fractions; in the case when only one rotamer is populated it is, at least in principle, possible to determine the width of the libration motion.

### 5.6. Torsion angle $\gamma$ and H5' and H5'' stereospecific assignment

Traditionally, the torsion angle  $\gamma$  as well as the H5'/H5'' stereospecific assignment have been determined from a combination of the  $^3J_{4'5'/5''}$ -couplings (Fig. 7(A), [63]), and the distances  $d_i(4'/3';5'/5'')$  (Fig. 7(B), [62]). With the availability of  $^{13}\text{C}$  labeled compounds it is now possible to also utilize a number of heteronuclear  $^1J_{\text{CH}}$ -couplings for this purpose [49,92]. Table 4 summarizes all the  $J$ -coupling constants and distances involved, and gives their values for the three main rotamers. As can be seen, a large number of additional  $J$ -couplings has become available. We will consider below, for each parameter in turn, its conformational dependence, and then how they can be used to determine the torsion angle  $\gamma$  and the H5'/H5'' stereospecific assignment.

The  $^3J_{4'5'}$ -coupling and  $^3J_{4'5''}$ -coupling depend

Table 4

Overview of  $J$ -couplings and distances for the determination of the torsion angle  $\gamma$  and for the stereospecific assignment of the H5' and H5'' resonances<sup>a</sup>

	$\gamma^+$	$\gamma^1$	$\gamma^-$
$^3J_{\text{H}5'\text{C}3'}$	+5.2	+1.3	+1.3
$^3J_{\text{H}5''\text{C}3'}$	+1.3	+1.3	+5.2
$^2J_{\text{H}5'\text{C}4'}$	-4.8	+1.2	-4.8 <sup>b</sup>
$^2J_{\text{H}5''\text{C}4'}$	+1.0	-4.2	-4.8 <sup>b</sup>
$^2J_{\text{H}4'\text{C}5'}$	+3.5	-1.5	-1.5 <sup>b</sup>
$^1J_{\text{H}5'\text{C}5'}$	151 <sup>c</sup>	151 <sup>c</sup>	151 <sup>c</sup>
$^1J_{\text{H}5''\text{C}5'}$	141 <sup>c</sup>	141 <sup>c</sup>	141 <sup>c</sup>
$^3J_{4'5'}$	+2.5	+2.7	+10.1
$^3J_{4'5''}$	+1.1	+10.2	+3.6
$d_i(3'5')^d$	3.6	2.4	2.7
$d_i(3'5'')^d$	2.5	2.7	3.7
$d_i(4'5')^d$	2.4	2.4	3.0
$d_i(4'5'')^d$	2.4	3.0	2.5

<sup>a</sup>  $J$ -coupling values are given in Hz and distances are in Å.

<sup>b</sup> The values are estimated on the basis of the known values of these  $J$ -couplings for the  $\gamma^+$  and  $\gamma^1$  rotamers and the Bock projection rule [86] (see text).

<sup>c</sup> There is some uncertainty considering the data in [49] that this relation holds for other torsion angle values than  $\gamma^-$ .

<sup>d</sup> Estimated values at torsion angles  $\gamma$  of 60°, 300° and 290° corresponding to the  $\gamma^+$ ,  $\gamma^1$  and  $\gamma^-$  rotamers, respectively.

on the torsion angle  $\gamma$  as determined by the generalized  $^3J_{\text{HH}}$  Karplus, given by Eqs. (12) and (13), and as shown in Fig. 7(A) [63]. In Table 4, these  $J$ -couplings are given for the  $\gamma^+$ ,  $\gamma^1$  and  $\gamma^-$  rotamers at the usual angles of 60°, 180° and 290°, respectively. As can be read off from this figure and seen in Table 4, when the torsion angle  $\gamma$  is in its usual  $\gamma^+$  domain both the  $^3J_{4'5'}$ -coupling and the  $^3J_{4'5''}$ -coupling are rather small, about 1.1 and 2.5 Hz, respectively, while for  $\gamma^1$  the  $^3J_{4'5'}$ -coupling is large (about 10.2 Hz) and the  $^3J_{4'5''}$ -coupling is small (about 2.7 Hz), while the reverse holds for the  $\gamma^-$  conformation. Assuming that the least populated rotamers, gauche(t) and gauche(-), have middle values for  $\gamma$  of 180° and 290°, it can be derived using the Karplus equation that the fraction gauche(+) can be calculated from [62]

$$f_g = \frac{13.3 - (^3J_{4'5'} + ^3J_{4'5''})}{9.7} \quad (29)$$

The distances  $d_i(3';5'/5'')$  also depend on the torsion angle  $\gamma$  Fig. 7(B). As can be seen, when the torsion angle  $\gamma$  is in its usual  $\gamma^+$  domain the distance  $d_i(3';5')$  is large and the distance  $d_i(3';5'')$  is small, for  $\gamma^1$  both the distances are small while for the  $\gamma^-$  conformation the reverse holds as for  $\gamma^+$  (see Table 4). How precisely one has to know these distances to determine the torsion angle  $\gamma$  or the fraction  $\gamma^+$  has been discussed in detail in Section 4 (see also Ref. [62]). Together with the  $J$ -couplings these distances also provide the H5'/H5'' stereospecific assignment. For example, when both the  $^3J_{4'5'}$ -coupling and  $^3J_{4'5''}$ -coupling are small the  $\gamma$  torsion angle is gauche(+), but in this case it is not possible to obtain the stereospecific assignment from these  $J$ -couplings. The stereospecific assignments of H5' and H5'' can then be obtained from the two distances  $d_i(3';5')$  and  $d_i(3';5'')$ , since for  $\gamma^+$  the latter is small while the former is large (see Table 4).

In addition to the classical parameters discussed above, a number of heteronuclear  $J$ -couplings have become accessible. Firstly, we consider the  $^3J_{\text{H}5'/\text{H}5''\text{C}3'}$ -couplings. Although no well established Karplus equation exists for the  $\gamma$  torsion dependence of this  $J$ -coupling, a qualitative assessment of the torsion angle  $\gamma$  is still possible from the available data (see Ref. [49]). For  $\gamma^+$ , the trans coupling  $^3J_{\text{H}5'\text{C}3'}$  is about +5.2 Hz and larger than the gauche coupling,  $^3J_{\text{H}5''\text{C}3'}$ , which is about +1.2–1.4 Hz. For  $\gamma^1$ , both

H5' and H5'' are gauche with respect to C3', so that both  ${}^3J_{\text{H5}'\text{C3}'}$  and  ${}^3J_{\text{H5}''\text{C3}'}$  are expected to be small, i.e.  $\approx 1.2$ – $1.4$  Hz (see Table 4, [49]). Secondly, the  ${}^2J_{\text{H5}'\text{C4}'}$  and  ${}^2J_{\text{H5}''\text{C4}'}$ -couplings also monitor the torsion angle  $\gamma$  and can provide stereospecific assignment of the H5' and H5'' protons [49,92]. For  $\gamma^+$ , the coupling  ${}^2J_{\text{H5}'\text{C4}'} = -4.6$  to  $-5.0$  Hz, while  ${}^2J_{\text{H5}''\text{C4}'}$  is small positive,  $+0.3$ – $1.4$  Hz; for  $\gamma^-$ ,  ${}^2J_{\text{H5}'\text{C4}'}$  and  ${}^2J_{\text{H5}''\text{C4}'}$  are  $+1.2$  Hz and  $-4.2$  Hz, respectively (see Table 4, [49,92]). Finally,  ${}^2J_{\text{H4}'\text{C5}'}$  can be estimated at  $3.5$  Hz and  $-1.5$  Hz for the  $\gamma^+$  and  $\gamma^-/\gamma^+$  states, respectively (see Table 4, [49,92]). These variations in  ${}^2J_{\text{HC}}$  are in accordance with the projection rule ([87], see also discussion in Section 5.3). For example, for  $\gamma^+$  H5' is oriented gauche with respect to O4' giving rise to a negative value for  ${}^2J_{\text{H5}'\text{C3}'}$ , while H5'' is trans with respect to O4' so that  ${}^2J_{\text{H5}''\text{C3}'}$  is positive.

The easily measurable  ${}^1J_{\text{H5}'\text{C5}'}$ -couplings have been discussed in Section 5.1. They seem to depend mainly on the stereospecificity, with the larger value for  ${}^1J_{\text{H5}'\text{C5}'}$  (see Table 4, [49]).

These additional heteronuclear  $J$ -couplings provide an alternative means for the determination of the torsion angle  $\gamma$  and the H5'/H5'' stereospecific assignment. Their usage for determining the torsion angle  $\gamma$  may or may not require stereospecific assignment of the H5' and H5'' resonances. For example, the  ${}^2J_{\text{H4}'\text{C5}'}$ -coupling depends only on the torsion angle  $\gamma$ , and does not involve the H5'/H5'' protons. It is thus ideally suited for the estimation of the torsion angle  $\gamma$ , since it does not require a knowledge of the stereospecific assignment of the H5' and H5'' resonances. A positive value immediately ascertains  $\gamma^+$  (see Table 4). On the other hand, although  ${}^3J_{\text{H5}'\text{H5}''\text{C3}'}$  and  ${}^2J_{\text{H5}'\text{H5}''\text{C4}'}$  also depend on the torsion angle  $\gamma$ , their unambiguous use requires the stereospecific assignment of the H5' and H5'' resonances; the same applies to  ${}^3J_{\text{HH}}$ -couplings,  ${}^3J_{\text{H5}'\text{H4}'}$  and  ${}^3J_{\text{H5}''\text{H4}'}$ , discussed previously. Finally, the  ${}^1J_{\text{H5}'\text{C5}'}$ - and  ${}^1J_{\text{H5}''\text{C5}'}$ -couplings seem not to depend on the torsion angle  $\gamma$ , but they can provide stereospecific assignment of the H5' and H5'' resonances, since  ${}^1J_{\text{H5}'\text{C5}'} > {}^1J_{\text{H5}''\text{C5}'}$  [49]. Another aspect

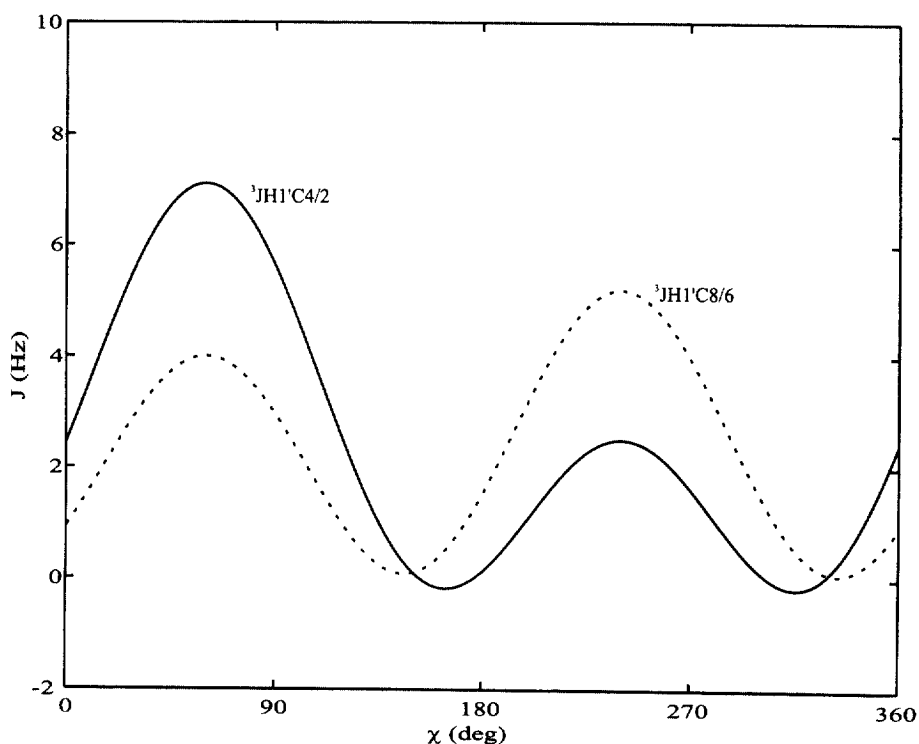


Fig. 8. The  ${}^3J_{\text{H1}'\text{C6/8}}$  and  ${}^3J_{\text{H1}'\text{C2/4}}$ -coupling constants calculated as a function of the torsion angle  $\chi$  on the basis of their Karplus relations (see text).

is that these extra  $J$ -couplings make it possible both to determine the torsion angle  $\gamma$  and obtain the stereospecific assignment of the H5' and H5'' protons, from various combinations of these  $J$ -couplings. In addition, NOE or distance data may also be available, which can be used as an extra source of information. Table 4 can be consulted to assess which particular combination of  $J$ -coupling and/or NOE data is required. For example, it can be gleaned from Table 4 that for the most common situation, namely that of  $\gamma^+$ , various different combinations of  $J$ -couplings suffice to both establish the torsion angle and obtain stereospecific assignment: (1) ( $^3J_{4'5'/5''}$ ,  $^2J_{H5'/H5''C4'}$ ), (2) ( $^3J_{4'5'/5''}$ ,  $^3J_{H5'/H5''C3'}$ ), (3) ( $^2J_{H4'C5'}$ ,  $^2J_{H5'/H5''C4'}$ ), (4) ( $^2J_{H4'C5'}$ ,  $^3J_{H5'/H5''C3'}$ ), (5) ( $^2J_{H5'/H5''C4'}$ ,  $^3J_{H5'/H5''C3'}$ ), (6) ( $^2J_{H4'C5'}$ ,  $^1J_{H5'/H5''C5'}$ ).

### 5.7. $\chi$ torsion angle and $^3J_{\text{HC}}$ sugar-to-base

The three-bond couplings  $^3J_{H1'C8/6}$  and  $^3J_{H1'C4/2}$  convey information about the glycosidic torsion angle  $\chi$ . Ippel et al. [49] have derived a new parametrization for the  $^3J_{H1'C8/6}$  Karplus equation:

$$^3J_{H1'C8/6} = 4.5 \cos^2(\chi - 60^\circ) - 0.6 \cos(\chi - 60^\circ) + 0.1 \quad (30)$$

and for the  $^3J_{H1'C4/2}$  Karplus equation they derive

$$^3J_{H1'C4/2} = 4.7 \cos^2(\chi - 60^\circ) + 2.3 \cos(\chi - 60^\circ) + 0.1 \quad (31)$$

As can be seen from Fig. 8, the combined use of  $^3J_{H1'C8/6}$  and  $^3J_{H1'C4/2}$  makes it possible to discriminate between the *anti*- and the *syn*-range for the torsion angle  $\chi$ . The  $^3J_{H1'C8/6}$ -coupling has a value of about 4.5 Hz in both the  $\chi_{\text{anti}}$  and the  $\chi_{\text{syn}}$  domain. On the other hand, the  $^3J_{H1'C4/2}$ -coupling has quite different values in the  $\chi_{\text{anti}}$  and the  $\chi_{\text{syn}}$  domains, 2.0 Hz and 6.0 Hz, respectively. Thus, in a qualitative sense, when  $^3J_{H1'C4/2} < ^3J_{H1'C8/6}$  the glycosidic torsion angle is in the *anti* domain, while for  $^3J_{H1'C4/2} > ^3J_{H1'C8/6}$  the glycosidic torsion angle is in the *syn* domain. We note that Davies and co-workers [82] also derived a correlation between the glycosidic torsion angle  $\chi$  and the one-bond coupling  $^1J_{H1'C1'}$ .

### 5.8. Measurement of homo- and heteronuclear $J$ -coupling constants

We discuss here the various methods that are available for measuring  $J$ -coupling constants in nucleic acids. A large number of (novel) approaches exist for the determination of  $J$ -couplings, the majority of which have been tested and applied to the determination of  $J$ -couplings in proteins (for reviews see, for example, Refs. [93,94]). The application to nucleic acids has been more limited. Here we discuss the established and new approaches as they pertain to their application in the determination of  $J$ -coupling constants in nucleic acids: (1) Determination of  $J$ -couplings from the shape of the signal; (2) Determination of  $J$ -coupling constants with the E.COSY principle; (3) Determination of  $J$ -coupling constants from signal intensities.

#### 5.8.1. Determination of $J$ -couplings from the shape of the signal

These methods apply generally well for small molecules where the line width is smaller than the  $J$ -coupling. The two main factors that complicate the determination are (i) the complexity of the multiplet pattern and (ii) the  $J$ -coupling to the line width ratio,  $J/LW$ . The higher the complexity of the multiplet pattern and the smaller the value of  $J/LW$  the more advanced methods are required.

Direct measurement of the  $J$ -couplings from either 1D NMR spectra or from 2D COSY in-phase or anti-phase multiplets works well only when the  $J$ -couplings are larger than or equal to the line width, otherwise corrections have to be applied. Kim and Prestegard [95] have proposed a simple and straightforward method for correcting for the line width in a doublet. Unfortunately, in nucleic acids no such simple doublets occur, except on the H1' resonance in RNAs which only couples to H2'. This  $^3J_{H1'H2'}$ -coupling is for RNAs the main  $^3J_{\text{HH}}$ -coupling used in practice, to determine the sugar pucker state. The more complex  $J$ -coupling patterns generally need to be simulated in order to obtain quantitative values for the  $J$ -couplings involved. Programs such as, for example, Sphinx/Linsha [96], can be used to extract the  $^3J_{\text{HH}}$ -couplings in a ribose ring by simulation of the multiplet patterns in a ( $^1\text{H}$ ,  $^1\text{H}$ ) DQF-COSY spectrum. More qualitative or less detailed

but still quantitative knowledge of  $J$ -couplings can be derived from complex multiplets under certain conditions without the aid of simulation. For example, certain rules have been proposed for the extraction of sums of  $^3J_{\text{HH}}$ -couplings from the multiplets in ( $^1\text{H}$ ,  $^1\text{H}$ ) DQF-COSY spectra (see e.g. Ref. [62]). Furthermore, as pointed out by Blommers et al. [97] the presence of passive  $^3J_{\text{HP}}$ -couplings in the ( $^1\text{H}$ ,  $^1\text{H}$ ) DQF-COSY multiplets allows the determination of  $^3J_{\text{HP}}$  from these spectra as well. Recently, Pikkemaat and Altona [98] published a set of simulated multiplet patterns found in ( $^1\text{H}$ ,  $^{31}\text{P}$ ) COSY spectra (see also ECOSY principle, Section 5.8.2). By visually comparing such patterns with experimentally observed patterns, the allowed ranges of the  $^3J_{\text{HP}}$  and  $^3J_{\text{HH}}$ -couplings involved can be determined.

All the aforementioned approaches have the danger that  $J$ -couplings are underestimated or overestimated depending on whether the splitting is measured in an in-phase or anti-phase doublet, respectively, since the line width is not known a priori. Macaya et al. [99] have published a procedure (CHEOPS) where they refine the simulated COSY multiplet patterns including line width and chemical shift position against the experimental COSY pattern. Similarly, Conte et al. [64] have published such procedures and extensively tested their method and described the accuracy of the  $J$ -couplings obtained. They have in addition included NOE information (Section 4) to improve the accuracy of the determined sugar puckering state. Leijon et al. [100] have also devised a method which allows interactive adjustment of the  $J$ -couplings while the multiplet is displayed. It uses SPHINX/Linsha to simulate two-dimensional multiplet patterns and includes a line width estimate.

An alternative approach for extracting  $J$ -couplings from a variety of spectra, but which has not been used extensively in the field of nucleic acids, is the idea of  $J$ -doubling, originally suggested by McIntyre and Freeman [101], later extended by Titman and Keeler [102] and then called the FIDs procedure. In the original approach a FID (free induction decay) is obtained by IFT from a slice through the multiplet, after which it is multiplied by  $\sin \pi J_{\text{test}} t$ . The subsequently fourier transformed FID will, in the case of a doublet with a splitting of  $J$  Hz, show an anti-phase signal in which the middle signal cancels when  $J_{\text{test}} = J$ . For more complex multiplets other minima

will occur, so that these  $J$ -couplings can also be extracted. However, the complexity of the multiplet increases by this procedure, making it less useful for nucleic acids. Furthermore, it does require the line width to be small compared to the  $J$ -coupling.

An extension of the above approach is the FIDs procedure by Titman and Keeler [102]. Here two spectra are acquired. One is the original spectrum and the other a reference spectrum which lacks the  $J$ -coupling of interest, and which is obtained via decoupling, for example. The reference spectrum can then be compared, via the FIDs procedure, i.e. multiply the time domain signal by  $\cos \pi J_{\text{test}} t$ , with the original spectrum and subtracted from it. A minimum is obtained when  $J = J_{\text{test}}$ . In the section on ECOSY this method will be discussed as well, since it allows the determination of  $J_{\text{HP}}$  and  $J_{\text{CP}}$  from heteronuclear E.COSY experiments. As far as we know it has not been applied to DQF-COSY spectra to obtain  $J_{\text{HP}}$ -couplings or to selectively decoupled DQF-COSY to obtain small  $J_{\text{HH}}$  couplings. The FIDs approach has been applied to the determination of  $^nJ_{\text{HC}}$  couplings in a HMBC spectrum [102]. The multiplet contains both  $J_{\text{HH}}$ -couplings and  $J_{\text{HC}}$ -couplings. It has dispersive character due to the evolution of the  $J_{\text{HH}}$ -couplings. The reference spectrum containing the  $J_{\text{HH}}$ -couplings can be obtained from, for example, a TOCSY experiment, by taking a 1D slice through a cross peak. In order to reproduce the dispersive character, the FID is reproduced from the 1D slice via IFT, and subsequently this FID is time shifted, which simulates the evolution of the  $J_{\text{HH}}$ -coupling. After fourier transformation of the time-shifted FID a 1D reference spectrum with the dispersive character is obtained. The  $J_{\text{HC}}$ -coupling is obtained via multiplication of the FID by  $\cos \pi J_{\text{HCtest}} t$  and fourier transformation. The test spectrum is subtracted from the original and a minimum is obtained when  $J_{\text{HCtest}} = J_{\text{HC}}$ .

Recently, Stonehouse and Keeler [103] have presented a new method akin to the FIDs procedure, which is more generally applicable. In this approach no reference spectrum is required. Instead, the multiplet is extracted from the spectrum and a time domain signal is obtained via inverse fourier transformation. The time domain signal  $S(t)$  is multiplied by  $\cos \pi J_{\text{test}} t$  or  $\sin \pi J_{\text{test}} t$  and the resultant time domain signal  $S(t, J_{\text{test}})$  is integrated over the region where  $S(t) < 0$ .

The integral is definite positive and has a minimum for  $J_{\text{test}} = J$ . Stonehouse and Keeler demonstrated the procedure on a number of different types of spectra, e.g. HSQC, HMBC, NOESY and COSY spectra, and investigated the effect of line width on the accuracy. The modified FIDs procedure works quite well for relatively simple multiplets, i.e. up to quartets of both in-phase and anti-phase nature.

Finally, the Inverse Fourier Transformation (IFT) method of in-phase multiplets can be applied [104]. Extraction of the multiplet and subsequent IFT gives a FID which can be zero-filled and provides a well digitized spectrum after FT. Alternatively, a test FID with a certain  $T_2$  decay rate can be fitted to this FID, giving both  $J$ -coupling and line width. This method is quite useful for simple doublets because it can be used for larger molecules as well as smaller molecules. Furthermore, only one spectrum is required, i.e. a NOESY spectrum which has to be recorded anyhow.

### 5.8.2. Determination of $J$ -coupling constants from E.COSY cross peak patterns

A large number of methods for the determination of coupling constants employ E.COSY multiplet patterns [94]. In E.COSY spectra the multiplet patterns are simplified with respect to those in a regular DQF-COSY spectrum. The  $J$ -couplings in such spectra can be directly read off from the frequency shift of two in-phase patterns. A large number of pulse sequences have been published that utilize the E.COSY principle.

The essence of the method is best illustrated using an AMX spin system (see Fig. 9). Both A and M are coupled to spin X, with  $J$ -couplings equal to  $J_{AX}$  and

$J_{MX}$ , respectively. We consider a cross peak between spins A and M created via some transfer method (e.g. TOCSY, COSY, NOESY) without influencing the spin state of X Fig. 9(a). The acquired signal will then contain two components:

$$S(t) = M \exp(i\omega_A t_1) \exp(j\omega_A t_2) \cos(\pi J_{AX} t_1) \cos(\pi J_{MX} t_2) \\ + M \exp(i\omega_A t_1) \exp(j\omega_A t_2) \sin(\pi J_{AX} t_1) \sin(\pi J_{MX} t_2) \quad (32)$$

The two-dimensional multiplet can thus be considered to consist of the sum of two spectra. In the first spectrum the  $J_{AX}$ -coupling and the  $J_{MX}$ -couplings to spin X are passive, giving rise to in-phase splittings in both dimensions Fig. 9(b). In the second spectrum the  $J_{AX}$ -coupling and the  $J_{MX}$ -coupling to spin X are active, giving rise to anti-phase splittings in both dimensions Fig. 9(c). The sum spectrum shows the well known E.COSY pattern, from which both the  $J_{AX}$ -coupling and the  $J_{MX}$ -coupling can be read off Fig. 9(d). The direction of the shift depends on the sign of the  $J$ -couplings.

**5.8.2.1. Homonuclear E.COSY.** In the homonuclear E.COSY experiment the effect outlined above is not obtained by the retention of the spin state of the passive spin but by suitable combination of several multiple quantum correlations. To derive the multiplet structure in an E.COSY spectrum one proceeds however in more or less the same way. Consider, as before, the cross peak between spins A and M, which are part of an AMX spin system with all three spins mutually  $J$ -coupled. Again one has to distinguish between active and passive  $J$ -couplings, with the former giving rise to anti-phase splittings and the latter to in-phase splittings. The E.COSY or P.E.COSY spectrum is the sum of two spectra. In the first spectrum the cross peak multiplet structure is derived by taking the  $J_{AM}$ -coupling as the active one and  $J_{AX}$  and  $J_{MX}$  as the passive  $J$ -couplings. In the second spectrum three  $J$ -couplings are active. The net result is, as before, an E.COSY pattern. The most straightforward approach to achieve homonuclear E.COSY patterns is to record a P.E.COSY spectrum [105]. This is in fact a regular COSY with a small flip angle for the read pulse. From this spectrum the dispersive diagonal peaks can be removed by subtracting a 2D spectrum with zero

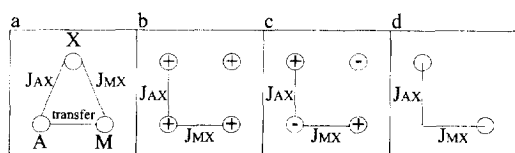


Fig. 9. Schematic illustration of the E.COSY principle: (a) the three-spin system, A, M and X; both A and M are  $J$ -coupled to X; magnetization or coherence is assumed to be transferred from A to M in an (A,M) 2D correlation spectrum, resulting in an (A,M) cross peak with splittings given by  $J_{AX}$  and  $J_{MX}$ , respectively. The (A,M) cross peak has an E.COSY multiplet pattern as shown in (d), which is the sum of a multiplet with all in-phase splittings (b), and a multiplet with all anti-phase splittings (c).



flip angle for the read pulse. This spectrum can simply be obtained arithmetically from a 1D spectrum by shifting the data points in the time domain to the left.

P.E.COSY spectra are often recorded for the determination of  $J_{\text{HH}}$ -couplings in DNA sugars. Indeed, simplification will be observed for a number of cross peaks, such as the most easily accessible  $\text{H1}'\text{H2}'$  and  $\text{H1}'\text{H2}''$  cross peaks. These cross peaks will have very distinct patterns depending on whether the sugar is in an S-puckered or an N-puckered state. The  $\text{H3}'\text{H2}'$ ,  $\text{H3}'\text{H2}''$  and  $\text{H2}'\text{H2}''$  cross peaks will also be simplified. The same applies for the group of protons involved with  $\text{H4}'$ ,  $\text{H5}'$  and  $\text{H5}''$ . The latter all strongly overlap, however, which may prevent the possibility of extracting the  $J$ -couplings. For RNA the  $\text{H2}''$  is lacking, so that on the  $\text{H1}'\text{H2}'$  cross peak no simplification can be attained. Only the cross peaks involving the  $\text{H4}'$ ,  $\text{H5}'$  and  $\text{H5}''$  resonances will be simplified, but they all strongly overlap.

**5.8.2.2. Heteronuclear E.COSY.** E.COSY patterns can be produced in a variety of ways in heteronuclear spectroscopy. Consider again three spins A, M and X, and assume that spins A and M are correlated in a heteronuclear (A,M) correlation spectrum. Assume further that both A and M are  $J$ -coupled to a spin X whose spin state is not changed during the experiment. Since the spin state of X remains unchanged, both in-phase and anti-phase splittings will show up and an E.COSY pattern will evolve as shown in Fig. 9. Alternatively, E.COSY patterns will show up in homonuclear spectra, where an E.COSY pattern will evolve if both homonuclear spins A and M are coupled to a heteronuclear spin X. It is also possible to obtain E.COSY patterns in triple-resonance spectra, for example, in an HNCA correlation spectrum [106] of a polypeptide. Here  $\text{H}_\text{N}$  coherence is transferred via  $^{15}\text{N}$  to  $\text{C}_\alpha$ , creating a  $\text{H}_\text{N}\text{C}_\alpha$  cross peak;  $\text{H}_\alpha$  is here the passive  $J$ -coupling partner of  $\text{H}_\text{N}$  and  $\text{C}_\alpha$ . In the  $t_1$  dimension  $\text{C}_\alpha$  is split up by the  $J_{\text{H}\alpha\text{C}\alpha}$ -coupling of 150 Hz and in the  $t_2$  dimension the  $\text{H}_\text{N}$  is split by the  $J_{\text{H}_\text{N}\text{H}\alpha}$ -coupling in an E.COSY-like fashion. The large  $J_{\text{H}\alpha\text{C}\alpha}$ -coupling makes it possible to determine the small  $J_{\text{H}_\text{N}\text{H}\alpha}$ -coupling very accurately, i.e. without it being disturbed by line width.

#### 5.8.2.2.1. Determination of $J_{\text{HP}}$ - and $J_{\text{CP}}$ -couplings.

For nucleic acids one could use a NOESY, TOCSY or COSY experiment to create the ( $\text{H}_\text{a}$ ,  $\text{H}_\text{b}$ ) cross peak which shows an E.COSY pattern of  $J_{\text{HaP}}$ -couplings and  $J_{\text{HbP}}$ -couplings [97]. The major problem is the extreme overlap in such spectra. This is particularly true for the ( $\text{H3}'$ ,  $\text{H2}'$ ) cross peak regions in a NOESY, COSY or TOCSY spectrum of RNA. For DNA, the ( $\text{H3}'$ ,  $\text{H2}'$ ) cross peak lies well off the diagonal in a less crowded region. In any case, the E.COSY pattern would here consist of a generally small  $^4J_{\text{H2}'\text{P3}'}$ -coupling and a  $^3J_{\text{H3}'\text{P3}'}$ -coupling which tends to lie between 4 and 8 Hz. As mentioned before the  $^3J_{\text{H3}'\text{P3}'}$ -coupling is not very informative with regard to the  $\epsilon$  torsion angle. The  $\beta$  torsion angle is monitored by  $^3J_{\text{H5}'\text{P5}'}$  and  $^3J_{\text{H5}''\text{P5}'}$ , while  $^4J_{\text{H4}'\text{P5}'}$  has a significant value ( $\approx 3$  Hz) only when the segment  $\text{H4}'\text{--C4}'\text{--C5}'\text{--O5}'\text{--P5}'$  is W-shaped, which occurs when  $\gamma$  is gauche + and  $\beta$  is trans. In the usual trans conformer for  $\beta$  both  $^3J_{\text{H5}'\text{P5}'}$  and  $^3J_{\text{H5}''\text{P5}'}$  are small, 3–4 Hz, while in both the  $\beta^+$  and  $\beta^-$  states one of them is quite large, up to 20 Hz. Thus, E.COSY patterns could be quite useful for determining  $\beta$ . However, the ( $\text{H5}'$ ,  $\text{H5}''$ ) cross peaks reside in an extremely crowded region; the same applies for the ( $\text{H4}'$ ,  $\text{H5}'/\text{5}''$ ) cross peaks.

The use of heteronuclei, either by enrichment or via natural abundance, can resolve a large part of the overlap. Schmieder et al. [107] have recorded for a circular DNA a natural abundance ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectrum without  $^{31}\text{P}$  decoupling. In the HSQC spectrum E.COSY patterns are formed on the different cross peaks involving the passive  $J$ -couplings with  $^{31}\text{P}$  nucleus, i.e. ( $J_{\text{H2}'\text{P3}'}$ ,  $J_{\text{C2}'\text{P3}'}$ ), ( $J_{\text{H3}'\text{P3}'}$ ,  $J_{\text{C3}'\text{P3}'}$ ), ( $J_{\text{H4}'\text{P3}'}$ ,  $J_{\text{C4}'\text{P3}'}$ ), ( $J_{\text{H4}'\text{P5}'}$ ,  $J_{\text{H4}'\text{C5}'}$ ), ( $J_{\text{H5}'\text{P5}'}$ ,  $J_{\text{C5}'\text{P5}'}$ ) and ( $J_{\text{H5}''\text{P5}'}$ ,  $J_{\text{C5}'\text{P5}'}$ ). In addition, the multiplets contain in-phase splittings in the  $f_2$  dimension due to  $J_{\text{HH}}$ -couplings and in the case of  $\text{C4}'$ , also in-phase splittings in the  $f_1$  dimension due to  $J_{\text{CP}}$ -couplings. Analysis of this HSQC spectrum allowed the determination of a large number of the  $J_{\text{HP}}$ - and  $J_{\text{CP}}$ -couplings. The sensitivity at natural abundance is low, putting some limitations on the approach, for example the  $J_{\text{HH}}$ -couplings could not be extracted from this spectrum. Van Buuren et al. [108] have applied high resolution HSQC to a  $^{13}\text{C}$  labeled small hairpin molecule. Here the same E.COSY patterns are observed, except for the extra splitting in the  $f_1$  dimension by the large  $^1J_{\text{CC}}$ -

coupling ( $\approx 35$  Hz). A much higher sensitivity is obtained here thanks to the  $^{13}\text{C}$  labeling. With the aid of simulation, virtually all the above mentioned  $J_{\text{HP}}$ - and  $J_{\text{CP}}$ -couplings could be extracted from this spectrum, and virtually all  $J_{\text{HH}}$ -couplings and the  $J_{\text{CC}}$ -couplings. The cross peaks involving H3' and H2' were rather difficult to interpret. This study showed that the approach is quite useful but limited to small systems. As will be discussed later, alternatively intensity-based methods could be used to obtain  $J_{\text{HP}}$ - and  $J_{\text{CP}}$ -couplings.

Schwalbe et al. [109] described an approach which uses both E.COSY patterns as well as the FIDs procedure (see above) to determine  $^3J_{\text{HP}}$ - and  $^3J_{\text{CP}}$ -coupling constants in a  $^{13}\text{C}$  labeled RNA fragment. They recorded four HSQC spectra, namely, one with full  $^{31}\text{P}$  decoupling, one without  $^{31}\text{P}$  decoupling, one with  $^{31}\text{P}$  decoupling in  $\omega_1$ , and one with  $^{31}\text{P}$  decoupling in the  $\omega_2$  dimension [109]. When no  $^{31}\text{P}$  decoupling is employed the E.COSY patterns discussed above are obtained. If the line width to  $J_{\text{HP}}$  or  $J_{\text{CP}}$  ratio is such that the  $J$ -couplings are not fully resolved, these  $J$ -couplings cannot be directly read off from the multiplet pattern. One can, as suggested above, resort to simulation of the multiplet or use the FIDs procedure to extract the  $J$ -coupling of interest. The latter approach is described by Schwalbe et al. [109]. In the FIDs procedure a reference spectrum is required which does not contain the  $J$ -coupling of interest. This spectrum is for example the  $\omega_1$ -decoupled spectrum. By addition of the  $\omega_1$ -decoupled spectra shifted by  $+J_{\text{test}}/2$  and  $-J_{\text{test}}/2$ , a simulated spectrum is obtained. This spectrum is then subtracted from the  $^{31}\text{P}$ -coupled spectrum. When  $J_{\text{test}} = J_{\text{CP}}$  the residual signal after subtraction is minimal. This procedure provides an alternative to the use of simulation of cross peak multiplets.

**5.8.2.2.2. Determination of  $J_{\text{HC}}$ -couplings.** The  $\omega_1$ -filtered (H,H) TOCSY (see e.g. Ref. [49]) can be used for determining  $^nJ_{\text{HC}}$  at natural abundance as well as for  $^{13}\text{C}$  labeled compounds. Instead of using a  $90^\circ$  rf pulse for excitation of the proton magnetization, an  $\omega_1$ -half filter is used which creates in-phase  $^1\text{H}$  magnetization of protons bound to a  $^{13}\text{C}$  nucleus and suppresses magnetization of protons bound to  $^{12}\text{C}$ . This method has also been used by Ippel et al. [49] to derive the  $J_{\text{HC}}$ -couplings in small circular

nucleotides and in  $^{13}\text{C}$  labeled NMPs. Hines et al. [92] have used a 3D HMQC-(H,H)TOCSY, originally proposed by Wijmenga et al. [110], to determine  $J_{\text{HC}}$ -couplings in a small  $^{13}\text{C}$  labeled RNA. The HMQC is used to spread the cross peaks in the  $^{13}\text{C}$  dimension. In this way, for example,  $J_{\text{H4}'\text{C5}'}$  and  $J_{\text{H5}'\text{C5}'}$  E.COSY patterns evolve, which can easily be measured thanks to the large  $J_{\text{H5}'\text{C5}'}$ -coupling. The limitation lies here in the H4' to H5' or H5'' TOCSY transfer, which for  $\gamma^+$  is limited, so that the cross peak of interest may have very low intensity. Similarly, in an N-puckered sugar ring the H1' to H3' transfer is small, so that the E.COSY cross peak containing  $J_{\text{H1}'\text{C3}'}$  becomes difficult to analyze. The same applies for  $^3J_{\text{H3}'\text{C1}'}$ . On the other hand, the transfer between H2' and H4' is not limited in an N-puckered sugar ring, so that the determination of  $^3J_{\text{H2}'\text{C4}'}$  and  $^3J_{\text{H4}'\text{C2}'}$  is expected to work well. This is borne out by the results obtained by Hines et al. [92].

Interestingly, Glaser et al. [111] have recently suggested a selective TOCSY for the sugar ring. This method allows selective transfer to occur between, say, C3' and C1', so that a 2D spectrum with only a (H1'C3') cross peak can be obtained. An E.COSY pattern will evolve with a large  $^1J_{\text{H1}'\text{C1}'}$  (150 Hz) and a small  $^3J_{\text{C3}'\text{H1}'}$  (2–6 Hz).

**5.8.2.2.3. Determination of  $J_{\text{HH}}$ -couplings via HCCH-E.COSY spectra.** The HCCH-E.COSY experiment has been proposed by Schwalbe et al. [112] and applied to a  $^{13}\text{C}$  labeled RNA 19-mer to determine the  $J_{\text{HH}}$ -couplings in the ribose sugar ring. To generate the E.COSY patterns the  $90^\circ$  rf pulses in the INEPTs are replaced by  $35^\circ$  rf pulses. The experiment is essentially employed in 2D format, but it may also be run as a 3D version to improve resolution. In the 2D version, for example, C1' is correlated with H2'. The cross peak will then show an E.COSY type multiplet pattern with  $J_{\text{H1}'\text{C1}'}$  and  $J_{\text{H1}'\text{H2}'}$  as the passive  $J$ -couplings (see Fig. 9). The large  $J_{\text{H1}'\text{C1}'}$  can effectively be used to resolve the smaller  $J_{\text{H1}'\text{H2}'}$ -coupling. Zimmer et al. [113] have applied 2D and 3D versions of this experiment to a  $^{13}\text{C}$  labeled DNA 10-mer duplex. It turns out that even with the labeling resonance, overlap prevents the determination of a full set of  $J_{\text{HH}}$ -coupling constants, for example only three out of ten  $J_{\text{H3}'\text{H4}'}$ -couplings could be determined.

### 5.8.3. Determination of $J$ -coupling constants from signal intensities

**5.8.3.1. Determination of  $J_{\text{HH}}$ -couplings from homonuclear ( $H,H$ ) TOCSY transfer.** This method has been proposed and used by van Duynhoven et al. [86] to determine the sugar pucker and the  $\gamma$  torsion angle from the cross peak intensities in TOCSY spectra. The approach has the advantage over DQF-COSY or P.E.COSY that, due to the relay effect in TOCSY, free lying cross peaks which contain important  $J$ -coupling information are more likely to be available. The method allows the determination of the sugar pucker with good accuracy for both DNA [12,86] and RNA [77]. The method not only works well for 2D TOCSY spectra, but can also be used for 3D TOCSY-NOESY spectra [75,114] and  $^{13}\text{C}$ -edited TOCSY spectra [260].

**5.8.3.2. Determination of  $J$ -couplings from heteronuclear experiments.** -coupling values have

traditionally been determined from the splittings in 1D spectra or  $n\text{D}$  spectra using a variety of methods as discussed above. Recently, Bax et al. introduced a number of procedures for the determination of homo- and heteronuclear  $J$ -couplings using quantitative  $J$ -correlation (for a review see Ref. [93]). In these experiments the  $J$ -coupling is derived from the intensity ratio between two resonances. Special care needs therefore to be taken that both resonances are identically affected by relaxation and digitization, so that the ratio only depends on the size of the active  $J$ -coupling. Bax et al. have demonstrated that the quantitative  $J$ -correlation experiments can provide accurate  $J$ -coupling values.

These methods have not been applied to nucleic acids, except for the HMBC type experiment [49,115] to determine  $^3J_{\text{H}^1\text{C}8/6}$  and  $^3J_{\text{H}^1\text{C}4/2}$ . Other applications of such methods could be of value for nucleic acids as well, although it does require modification of the pulse sequences. For example, for nucleic acids a particularly appealing experiment

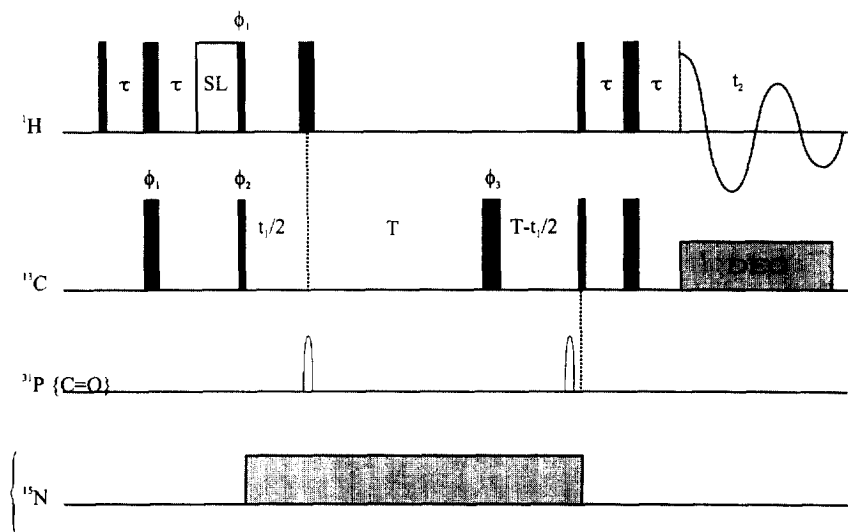


Fig. 10. Pulse sequence of the CT-HSQC experiment, adapted from Vuister and Bax [116], for the measurement of  $J_{\text{CP}}$ -couplings. Narrow and wide pulses denote  $90^\circ$  and  $180^\circ$  flip angles, respectively. All proton, carbon and phosphor pulses can be of high power. SL denotes a spin-lock pulse to suppress the residual HDO resonance. Unless indicated otherwise all pulses are applied along the  $x$  axis. Minimum phase cycle resulting in artifact-free spectra is as follows:  $\phi_1 = y, -y$ ;  $\phi_2 = x$ ;  $\phi_3 = 2(y), 2(-y), 2(-x), 2(x)$ ;  $\phi_4 = 8(x), 8(-x)$ ; receiver =  $2(x, -x), 4(-x, x), 2(x, -x)$ . Quadrature detection in  $t_1$  is obtained by the States-TPPI technique incrementing  $\phi_2$ . The delay  $\tau$  should be set to  $1/4J_{\text{HC}} \approx 1.7$  ms; the constant time  $2T$  should for nucleic acids be adjusted to a multiple of  $1/J_{\text{CC}}$  for optimal refocusing of the ribose carbon-carbon couplings of 35 Hz. The  $^{13}\text{C}$  (as well as optional  $^{31}\text{P}$ ) decoupling can be accomplished with a GARP sequence, relatively low rf field strengths can be used,  $\approx 2.5$  to 4.7 kHz (and  $\approx 1$  to 2 kHz for  $^{31}\text{P}$ ). The correspondence with the parent experiment of Vuister and Bax [116] is indicated in the sequence, by C=O in parenthesis on the  $^{31}\text{P}$  line, and by the  $^{15}\text{N}$  line in curly brackets.

is the CT-HSQC, which can be performed either with or without  $^{31}\text{P}$  decoupling. The ratio of the cross peaks in the two experiments is directly proportional to the  $J_{\text{CP}}$ -coupling. The pulse sequence of this experiment is shown in Fig. 10. This pulse sequence has been proposed and used by Vuister and Bax [116] to determine  $J_{\text{CN}^-}$  couplings with good results. In a modified form this type of experiment has been used to determine  $\epsilon$  and  $\beta$  torsion angles in a large 44 nucleotide RNA pseudoknot [260].

## 6. Chemical shifts

Chemical shifts are the most easily measured NMR

parameters and carry important structural information. Their values depend on the electron densities around the nuclei, which is influenced by the surroundings in variety of ways. Owing to the uncertainties in their dependence on structural parameters the use of chemical shifts as a tool to determine three-dimensional characteristics has been overtaken by  $J$ -couplings and NOEs, for which the relationship with structural parameters is much better established. Chemical shifts have however experienced renewed interest as a structural tool in the field of protein NMR. The large database of chemical shifts and independently determined protein structures now available makes it possible to reliably interpret chemical shifts in terms of structural

Table 5  
Overview of chemical shift ranges (ppm) in DNA and RNA

Name	Thymidine	Uridine	Cytosine	Adenine	Guanine
H1'	5–6	5–6	5–6	5–6	5–6
H2' (DNA)	1.7–2.3		1.7–2.3	2.3–2.9	2.3–2.9
H2' (RNA)		4.4–5.0	4.4–5.0	4.4–5.0	4.4–5.0
H2''	2.1–2.7	2.1–2.7	2.1–2.70	2.4–3.1	2.4–3.1
H3'	4.4–5.0	4.4–5.0	4.4–5.0	4.4–5.2	4.4–5.2
H4'	3.8–4.3	3.8–4.3	3.8–4.3	3.8–4.3	3.8–4.3
H5'	3.8–4.3	3.8–4.3	3.8–4.3	3.8–4.3	3.8–4.3
H5''	3.8–4.3	3.8–4.3	3.8–4.3	3.8–4.3	3.8–4.3
H6/8	6.9–7.9	6.9–7.9	6.9–7.9	7.7–8.5	7.5–8.3
H5	1.0–1.9	5.0–6.0	5.0–6.0		
NH (G,T,U)	13–14	13–14			12–13.6
NH2 (C,G,A)			6.7–7.0 8.1–8.8	5–6 7–8	5–6 8–9
C1' (RNA)	87–94	87–94	87–94	87–94	87–94
C1' (DNA)	83–89	83–89	83–89	83–89	83–89
C2' (DNA)	35–38		35–38	35–38	35–38
C2' (RNA)		70–78	70–78	70–78	70–78
C3'	70–78	70–78	70–78	70–78	70–78
C4'	82–86	82–86	82–86	82–86	82–86
C5'	63–68	63–68	63–68	63–68	63–68
C2	154	154	159	152–156	156
C4	169	169	166–168	149–151	152–154
C5	15–20	102–107	94–99	119–121	117–119
C6	137–142	137–144	136–144	157–158	161
C8				137–142	131–138
N1/N3 (imino)	156	156–162			146–149
N2/6/4 (NH2)			94–98	82–84	72–76
N1(Py)/N9(Pu)	144	142–146	150–156	166–172	166–172
N3			210	220–226	167
N1(Pu)				214–216	
N7(Pu)				224–232	228–238

Data are taken from Dieckman et al. [225], Hilbers et al. [76], Ippel et al. [49], Nikonowicz et al. [148], Pardi [48], Van de Ven and Hilbers [133], van Dongen et al. [144], Varani et al. [36] and Wijmenga et al. [56,62,50]. The ranges are applicable for both DNA and RNA, if not otherwise indicated.

parameters in a number of ways (see, for example, Refs. [117–119] for recent reviews). Firstly, it has become possible to assess the reliability of quantitative chemical shift calculations. For  $^1\text{H}$  shifts semi-empirical relationships have been derived which quite accurately (rmsd = 0.25 ppm) predict experimental shifts ([120–123]). From these studies it appears that the conformation dependent shifts of both the  $\text{H}_\alpha$  and the NH resonances result mainly from the magnetic anisotropy of the peptide bond. The conformation dependent shift resulting from ring current effect is limited because of the small number of aromatic residues present in proteins. Improved and extensive quantum mechanical calculations show for heteronuclei such as  $^{13}\text{C}$  quite good correlations with experimental data [124]. In addition, purely empirical correlations have been derived such as the chemical

shift index [118]. This has led to the introduction of chemical shifts as constraints in simulated annealing [125,126].

All nucleic acid residues are aromatic in nature, in contrast to amino acid residues in proteins. It is therefore expected that conformation dependent  $^1\text{H}$  chemical shifts originating from the aromatic rings will play a major role in nucleic acids. The reliability of  $^1\text{H}$  chemical shift calculations in the field of nucleic acids has only recently been tested [50]. In this study the experimentally determined and calculated chemical shifts were compared for about 20 well determined DNA structures. The correspondence between calculated and experimental shifts was shown to be quite good (rmsd = 0.17 ppm). In view of these encouraging results we will discuss  $^1\text{H}$  chemical shift calculations in nucleic acids in some detail.

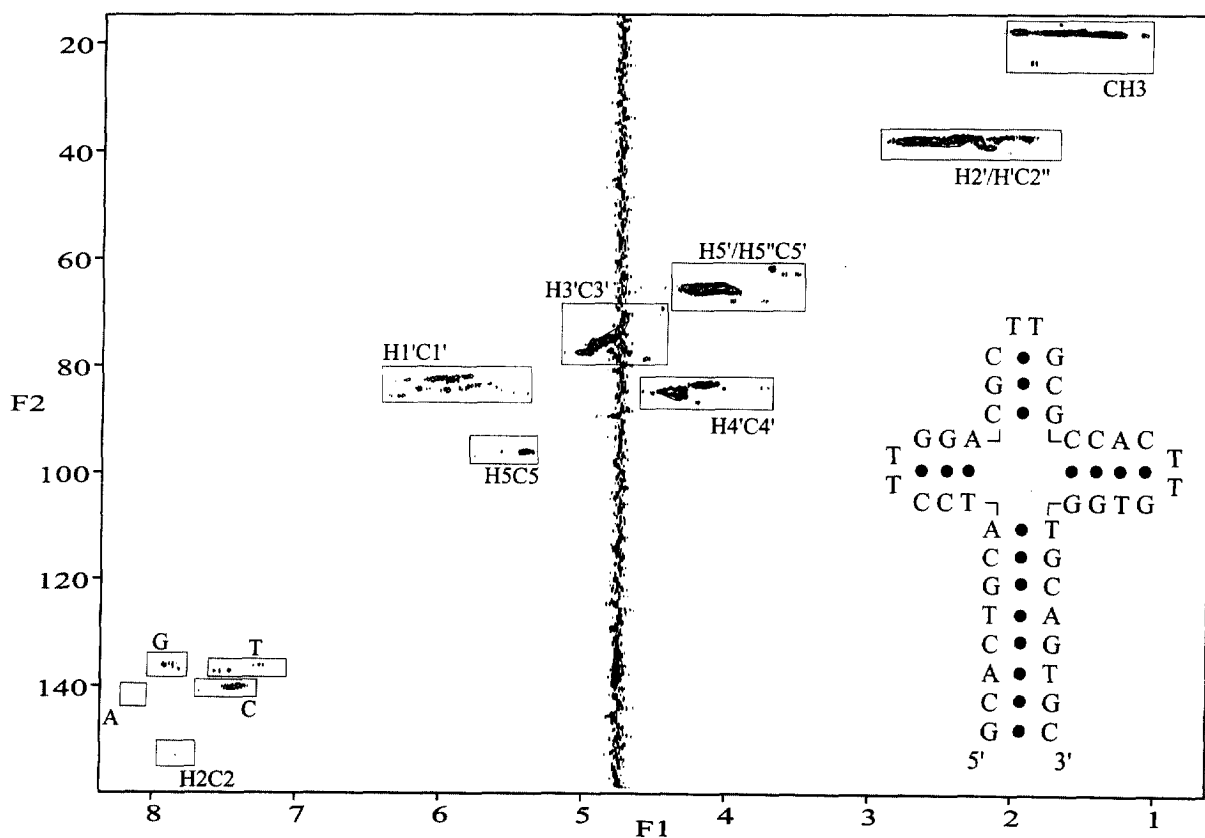


Fig. 11. Natural abundance ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HMQC of a DNA four-way junction. The different spectral regions for different carbon proton correlations are indicated in boxed regions.

### 6.1. Chemical shifts: qualitative aspects

Chemical shifts have until recently mainly been used as an assignment tool and their dispersion has been used to reduce resonance overlap in multi-dimensional NMR experiments. Table 5 gives an overview of the approximate ranges for the  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  shifts in nucleic acids. As can be seen, the H2' to H5'' resonances in RNA and the H3' to H5'' resonances in DNA strongly overlap. This overlap can be reduced by the use of  $^{13}\text{C}$  correlated experiments. Indeed, as can be seen from Fig. 11, where a ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HMQC spectrum is shown of a DNA four-way junction, the H3' to H5'' resonances are separated by the  $^{13}\text{C}$  frequency resonances of the ribose ring into three groups consisting of the (H3',C3'), the (H4',C4') and the (H5'/5'',C5') cross peaks, respectively. For RNA one also finds in this region of a ( $^1\text{H}$ ,  $^{13}\text{C}$ ) correlation spectrum (HMQC, or HSQC) three groups of cross peaks, but now the groups consist of (H3',C4'), (H2'/H3',C2'/C3') and (H5'/5'',C5') cross peaks, respectively. Unfortunately, within each group the resonances still tend to cluster quite strongly, making assignment difficult. Strong overlap is also observed for the other groups of cross peaks, as can be seen from Fig. 11. The best dispersed are the (H1',C1') cross peaks. Note that the Cytosine and Thymine (H6,C6), Adenine (H8,C8) and Guanine (H8,C8) cross peaks cluster in separate, only slightly overlapping, groups, which may help in their assignment. As can be appreciated from Table 5 several of the  $^{13}\text{C}$  resonances that have no protons attached cluster in even narrower spectral ranges than those with a proton attached. In particular the C4, C2 resonances of Cytosine and Uracil overlap quite strongly. The C5 resonances in the two purines A and G fall into two groups, and show within each group a slightly better dispersion.

The  $^{15}\text{N}$  resonances of the purines and pyrimidines also group into a number of regions (see Table 5). The imino  $^{15}\text{N}$  resonances fall into two regions around 150 ppm, with G N1 resonating around 147 ppm and U or T N3 resonating around 157 ppm. The T/U N1, C N1 and A/G N9 can be found in spectral regions centering around 143, 154 and 168 ppm, respectively. The NH2  $^{15}\text{N}$  resonances for G, A and C can be found at 73, 83 and 96 ppm, respectively. The N7

and N3 in purines and the N3 of unprotonated Cytosine reside in the spectral region which ranges from 210 to 236 ppm.

The  $^{31}\text{P}$  resonances cluster quite strongly in a narrow spectral region of about 2 ppm, ranging from  $-3$  to  $-1$  ppm. As will be discussed later in more detail, unusual backbone torsion angles may lead to a shift out of this region.

### 6.2. Theory

The chemical shift of nuclear spin  $k$  in a molecule  $C$  depends on the electron density distribution surrounding it. The electron density and the chemical shift derived from it could be obtained in principle by *ab initio* quantum mechanical calculations. However, quantum mechanical chemical shift calculation of a complete biomolecule is completely out of reach even with present day computers, although recent progress in quantum mechanical theory 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 (see for example Refs. [117,127]). For chemical shift calculations it is thus operationally expedient to divide a molecule into a number of fragments, i.e. a fragment  $A$  in which the nucleus of interest resides, the conformation of which is defined with respect to a reference conformation, and a number,  $N$ , of other fragments  $B$  interacting with  $A$ . The calculated chemical shift contributions can then be divided into two categories, namely a conformation independent part,  $\delta_{\text{intrin}}$ , and a conformation dependent part,  $\delta_{\text{conf}}$ , respectively. The former represents the chemical shift of nucleus  $k$  in fragment  $A$  in its reference state and in the absence of other fragments. The latter represents chemical shift changes of nucleus  $k$  with respect to the reference state. These shift changes can result from: (a) changes in the local electronic environment,  $\delta_{k,\text{lc}A}$ , i.e. changes in fragment  $A$  with respect to its reference state due, for example, to changes in torsion angle, bond length, bond angle, etc. or (b) changes in the interaction of nucleus  $k$  in  $A$  with the other molecular fragments  $B_j$ ,  $\delta_{k,j,B}$ . For  $^1\text{H}$  nuclei the conformation dependent shift of nucleus  $k$  in  $A$  does not depend strongly on the torsion angles in  $A$ , i.e. the conformation dependent shifts can conveniently be attributed to interactions with other fragments. For  $^1\text{H}$  nuclei it is

therefore expedient to choose as molecular fragments A, the individual C–H or N–H fragments. As a result,  $\delta_{\text{intrin}}$  of the  $^1\text{H}$  nuclei in the sugar moiety of nucleic acids can appropriately be defined as belonging to the individual C–H fragments. In contrast, for hetero-nuclei the chemical shift of a particular nucleus  $k$  may be quite strongly affected by torsion angle changes around this nucleus (see, for example, Ref. [117]). For instance, for  $\delta_{\text{intrin}}$  of a  $^{13}\text{C}$  nucleus in a ribofuranose ring it would be operationally more convenient to use as a reference state the S-puckered conformation of the sugar. The chemical shift,  $\delta_k$ , of the resonance of nuclear spin  $k$  in molecule C, can then formally be written as

$$\delta_k = \delta_{k,\text{intrin}} + \delta_{k,\text{conf}} = \delta_{k,\text{intrin}} + \delta_{k,\text{lcA}} + \sum_{j=1}^N \delta_{kj,\text{B}} \quad (33)$$

The conformation dependent terms  $\delta_{kj,\text{B}}$  that result from interactions with other fragments can each in turn be divided into a number of distinct contributions,  $\delta_{\text{rc}}$ ,  $\delta_{\text{ma}}$ ,  $\delta_{\text{E}}$  and  $\delta_{\text{CT}}$ , so that

$$\delta_k = \delta_{k,\text{intrin}} + \delta_{k,\text{lcA}} + \sum_{j=1}^N \{ \delta_{kj,\text{rc}} + \delta_{kj,\text{ma}} + \delta_{kj,\text{E}} + \delta_{kj,\text{CT}} \} \quad (34)$$

Together  $\delta_{\text{rc}}$  and  $\delta_{\text{ma}}$  form the chemical shift variation resulting from magnetic anisotropy effects, with  $\delta_{\text{rc}}$  the chemical shift from ring current effects produced by aromatic rings, and  $\delta_{\text{ma}}$  the chemical shift due to local magnetic anisotropy effects from, for example, an asymmetric electron distribution on atom B interacting with nucleus  $k$  of A. Analytical expressions have been derived for the two magnetic anisotropy terms (to be discussed below). The parameters in the analytical expressions have derived either from experimental shift data (e.g. Refs. [120,121]), or from fitting to quantum mechanically calculated isoshielding curves [127]. Note that the quantum mechanically calculated isoshielding curves reflect the sum of the two magnetic anisotropy terms,  $\delta_{\text{rc}}$  and  $\delta_{\text{ma}}$ , so that the ring current and local magnetic anisotropy contributions may be difficult to separate from the isoshielding curves [127]. The analytical expressions reproduce the quantum mechanically calculated isoshielding curves very accurately, i.e. they

differ on average by not more than 0.07 ppm and at most 0.2 ppm, this larger difference being found for the imino protons of Guanine and the H2 protons of Adenine. The polarization term,  $\delta_{\text{E}}$ , is the chemical shift change resulting from polarization by an electric field of the electron density along the chemical bond extending from atom I. It can also be described by a simple analytical expression with the parameters that have been obtained from comparison with experimental data (see below). The charge transfer term,  $\delta_{\text{CT}}$ , has been introduced by Giessner-Prettre and Pullman [127] to account for the transfer of electron density on hydrogen bonding.

For chemical shift calculations the following further considerations should be taken into account. As discussed above for shift calculations of  $^1\text{H}$  nuclei the term  $\delta_{\text{lcA}}$  can be set to zero, and when non-exchangeable protons are considered the term  $\delta_{\text{CT}}$  can also be taken to be zero. The value of the term  $\delta_{\text{intrin}}$  is not known a priori, nor is it easily calculated. However, since  $\delta_{\text{intrin}}$  is conformation independent, it is possible to derive its value from a comparison of  $\delta_{\text{conf}}$  and  $\delta_{\text{conf,exp}}$ , the conformation dependent parts of the calculated and experimental chemical shifts, respectively. The term  $\delta_{\text{conf,exp}}$  is given by

$$\delta_{\text{conf,exp}} = \delta_{\text{exp}} - \delta_{\text{ref}} \quad (35)$$

Here  $\delta_{\text{exp}}$  is the experimentally observed shift and  $\delta_{\text{ref}}$  an experimentally determined reference value, which in fact is the experimental counterpart of  $\delta_{\text{intrin}}$ . It thus represents the chemical shift that would have been observed in the absence of conformation dependent shift effects.

Shifts introduced as a result of ring currents have been discussed extensively in the literature (for reviews see Refs. [117,127]). The general form of the semi-empirical equation that describes these shifts is

$$\delta_{\text{rc}} = iBG(\vec{r}) \quad (36)$$

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 factors,  $i$ , of the rings considered are equal to the ring current intensity of that ring relative to that of benzene.  $G(\vec{r})$  is a geometric factor with  $\vec{r}$  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 [128] and by Haigh and Mallion [129]. The results generated by the two approaches are very similar. In the Johnson–Bovey method adopted by Giessner-Prettre and Pullman [127], it is assumed that the ring current shielding arises from two ring current loops of radius  $a$ , situated symmetrically at each side of and parallel to the plane of the aromatic molecule. The shielding is calculated at points characterized by the cylindrical coordinates  $\rho$  and  $z$ , expressed in units of the loop radius  $a$ . The geometry factor is then given by

$$G(r, z) = \frac{1}{a} \frac{2}{((1 + \rho)^2 + z_+^2)^{1/2}} \times \left( K(k) + \frac{1 - \rho^2 - z_-^2}{(1 - \rho)^2 + z_-^2} E(k) \right) + \frac{2}{((1 + \rho)^2 + z_+^2)^{1/2}} \times \left( K(k) + \frac{1 - \rho^2 - z_-^2}{(1 - \rho)^2 + z_+^2} E(k) \right) \quad (37)$$

where  $z_{\pm} = z \pm \langle z \rangle$ , with  $\langle z \rangle$  being the distance between the loop and the plane of the aromatic ring.  $K$  and  $E$  are complete elliptic integrals of the first and second kind. The argument  $k$  is given by

$$k = \frac{4\rho}{((1 + \rho)^2 + z^2)} \quad (38)$$

Wijmenga et al. [50] used the Johnson–Bovey method in their calculations and the parameters as derived by Giessner-Prettre and Pullman [127]. The ring current loops are then taken to be at heights of 0.5770 Å and 0.5660 Å above and below the aromatic plane, for Cytosine/Thymidine and Adenine/Guanine, respectively. The radii of the current loops are 1.3675, 1.3790, 1.3610, 1.1540, 1.3430 and 1.15440 Å, for Cytosine, Thymine, Guanine-6, Guanine-5, Adenine-6 and Adenine-5, respectively (the numbers 5 and 6 refer to the 5-membered and 6-membered rings of purines bases). The ring current intensities are 0.2570, 0.1120, 0.3, 0.6530, 0.9 and 0.66, for Cytosine, Thymine, Guanine-6, Guanine-5, Adenine-6 and Adenine-5, respectively. The empirical parameter  $B$  is adjusted to a value of  $2.13 \times 10^{-6}$  Å, so that the ring current shifts of the protons of benzene equal  $-1.5$  ppm, when  $\langle z \rangle = 0.61$  Å, and  $i = 1$ .

The chemical shift of a nucleus  $i$  of atom A induced by the local magnetic anisotropy at an atom B is, expressed in cartesian coordinates, given by [127]:

$$\delta_{\text{ma}} = \frac{1}{3r^5} \sum_{\alpha\beta} (3r_{\alpha}r_{\beta} - r^2\delta_{\alpha\beta})(1.967R_{\alpha\beta} - 5.368Q_{\alpha\beta}) \quad (39)$$

Here,  $r$  is the distance between atoms A and B, and  $\alpha, \beta = x, y, z$ , are the Cartesian coordinates. Furthermore,  $R_{\alpha\beta}$  and  $Q_{\alpha\beta}$  are the diamagnetic and paramagnetic parts, respectively, of the  $\alpha\beta$  element of the magnetic susceptibility tensor of atom B. These contributions have been calculated and tabulated by Ribas Prado and Giessner-Prettre [130]. They were used without modification by Wijmenga et al. [50] in their calculations.

The effect of an electric field on proton chemical shifts is usually calculated using the expression originally derived by Buckingham [131]:

$$\delta_E = AE_{\parallel} + BE^2 \quad (40)$$

Here  $E_{\parallel}$  is the projection of the electric field  $\vec{E}$  along the X–H bond at the proton considered. Application of different computational methods [131] indicated that  $A$  may vary between 2.5 and  $3.0 \times 10^{-12}$  e.s.u. Wijmenga et al. [50] chose for nucleic acids the same value ( $2.9 \times 10^{-12}$  e.s.u.) as chosen by Giessner-Prettre and Pullman [127]. The value of  $B = 0.74 \times 10^{-18}$  e.s.u. they took from Marshall and Pople [132]. The electric field at the proton considered,  $k$ , can be derived, using Coulomb's law, as the vector sum of the fields emanating from the electric monopoles (partial charges),  $q_j$ , predicted to be present at the different atoms  $j$ :

$$\vec{E} = \sum_{j=1}^N \frac{q_j \vec{r}_{kj}}{4\pi\epsilon_0\epsilon r_{kj}^3} \quad (41)$$

Here  $r_{kj}$  is the distance between atoms  $k$  and  $j$ ,  $\epsilon_0$  is the dielectric constant of a vacuum and  $\epsilon$  is the relative dielectric constant of the medium. For the partial charges, Wijmenga et al. [50] tried different sets, but ultimately the partial charge set for the CHARMM force field and present as a parameter set in XPLOR was used. Furthermore, in the calculations of the electric field a distance dependent dielectric constant was used,  $\epsilon(r) = 4r$ , an approach often used in molecular mechanics calculations for nucleic acids.



### 6.3. $^1\text{H}$ shifts

The accuracy of proton chemical shift calculations in nucleic acids has been tested by comparing the calculated and experimentally observed chemical shifts [50] for a set of approximately 20 well determined NMR structures (involving DNA helices, hairpins, a quadruplex and a circular dumbbell). This resulted in a database of over 2000 chemical shifts. The chemical shifts of the non-exchangeable protons were calculated, as described above, according to Eqs. (34)–(41), with  $\delta_{\text{icA}}$  and  $\delta_{\text{CT}}$  set to zero, and without adjustment of the parameters for ring current, local magnetic anisotropy and electric field derived earlier by Giessner-Prettre and Pullman [127]. Values for  $\delta_{\text{ref}}$  were obtained by using it as the only adjustable parameter in the comparison of the conformation dependent parts of the calculated and experimental observed shifts,  $\delta_{\text{conf}}$  and  $\delta_{\text{conf,exp}}$ , respectively. As can be seen from Fig. 12(A) a very good correspondence is found, the overall rmsd is 0.17 ppm Fig. 12(B). As is evident from Fig. 12(C), (D) and (E), the two main and about equally important contributors are the ring current and local magnetic anisotropy, whereas the effect of the electric field component is surprisingly small.

### 6.4. Structurally important $^1\text{H}$ shifts

The comparison of calculated and observed shifts in Ref. [50] revealed the presence of a number of characteristic chemical shift effects in various DNA structures, some of which had been observed previously [133]. The comparison showed the physical background of these characteristic shifts. The most important dependencies of the shifts of the different non-exchangeable protons in a nucleotide on their surrounding can be summarized via an estimated chemical shift,  $\delta_{\text{est}}$ , which is split for each proton into a number of terms:

$$\delta_{\text{est}} = \delta_{\text{ref}} + \delta_{\text{ib}} + \delta_{3'\text{b}} + \delta_{5'\text{b}} \quad (42)$$

Here,  $\delta_{\text{ref}}$  is the reference shift discussed earlier. The term  $\delta_{\text{ib}}$  is the chemical shift effect induced by the own base when the  $\chi$ -angle is  $240^\circ$ , which is the value usually found in DNA for this torsion angle. This value was taken from the  $\chi$ -angle dependence of the calculated chemical shift in a single nucleotide. The terms  $\delta_{3'\text{b}}$  and  $\delta_{5'\text{b}}$  represent the chemical shift

induced by the 3'- and 5'-neighboring base, respectively, in a B-helix environment. Their values were estimated in Ref. [50] from the difference of  $\delta_{\text{exp}} - \delta_{\text{ref}} - \delta_{\text{ib}}$  from zero for each type of neighbor. Table 6 gives an overview of these structural shifts; in a double helical environment, the sugar proton shifts are to a large extent determined by the own base and sequential effects are small, except for the H1' protons. In particular the shifts of the H2' and H2'' protons are separated into two groups, one for purines and one for pyrimidines. We note in passing that the H2' proton shift depends strongly on the glycosidic torsion angle  $\chi$ . For  $50^\circ < \chi < 130^\circ$  a strong downfield shift occurs with a maximum of 1.2 ppm at  $\chi = 100^\circ$ . Such a downfield shift of the H2' resonance is very strong evidence for a *syn*  $\chi$ -angle (seen in the quadruplex DNA structures for example), since ring current effects from neighboring bases are generally upfield and small, except for special cases. For H3' to H5'/H5'' the shift also solely depends on the own base, but the effect of the base is smaller and decreases going from H3' to H5'/5''. The H1' proton shift is affected both by the own base and the 3'-neighboring base. The own base again separates the shifts of the H1' protons into a purine and a pyrimidine group. Each group in turn is separated into two groups depending on whether the 3'-neighboring base is a purine or pyrimidine residue (see Table 6).

The shift of the base protons can depend in principle on both the 3'- and the 5'-neighboring bases. In accordance with the B-helix geometry the H5 and the methyl protons are found to experience a rather strong shift effect from the 5'-neighboring base. The H2 proton can show quite large upfield shifts. These stem from a variety of sources but mostly from the 3'-neighboring base, which again is in accordance with the B-helix geometry. The H2 proton shifts are the worst correlated, for a variety of reasons. It may result from the less well determined position in the test structures, but there are also strong indications that the ring current intensity used may be slightly overestimated, and finally at large shift values (around 1.5–2.0 ppm) the ring current effect may have to be corrected for so-called dispersion effects [50]. Finally, we note that, as expected for a B-helix geometry, the H8 and H6 protons are affected by both the 3'- and 5'-neighboring bases (see Table 6).

The above results are not only useful as structural

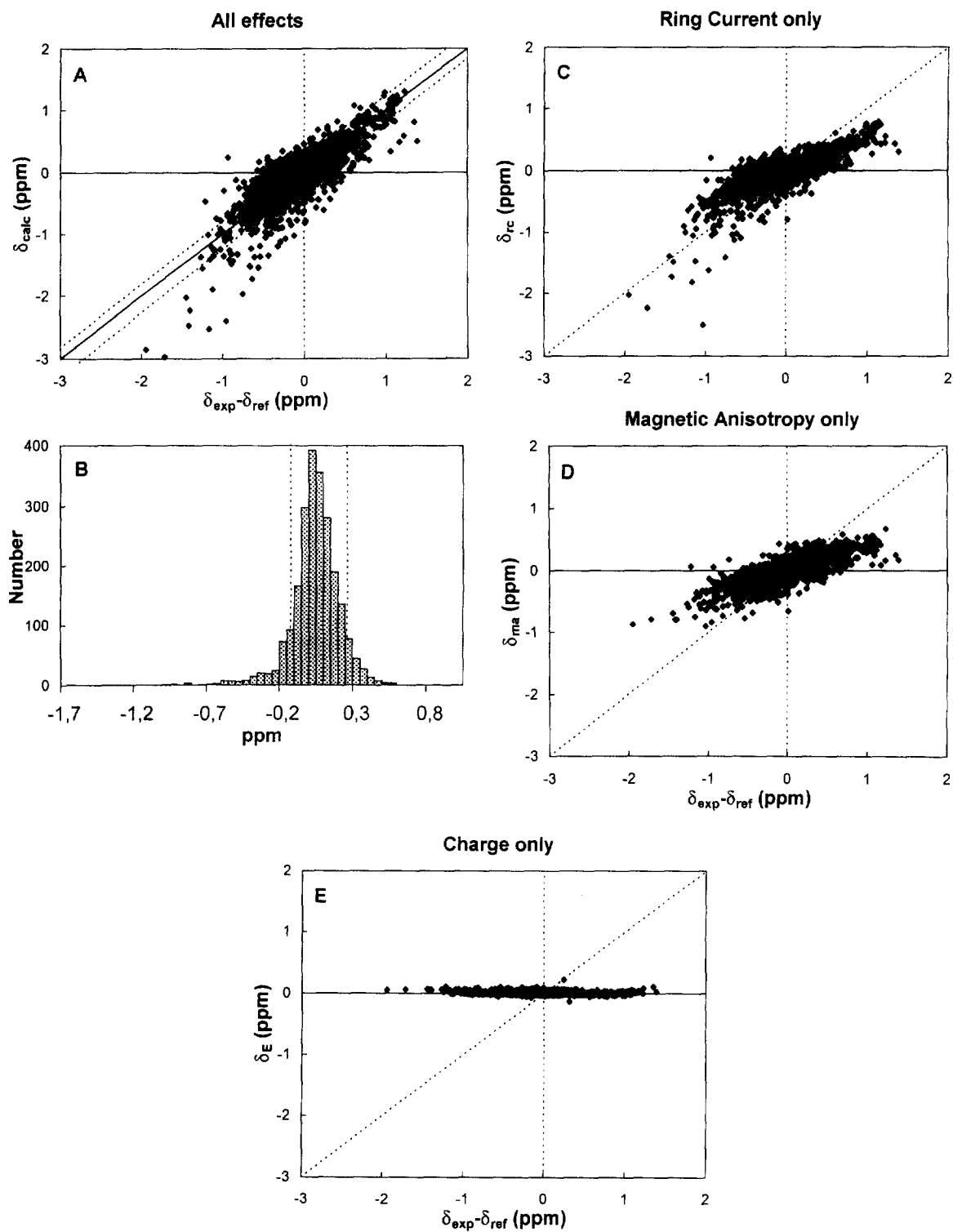


Fig. 12.

Table 6  
Structural  $^1\text{H}$  chemical shifts (ppm) in DNA<sup>a</sup>

	$\delta_{\text{ref}}^b$				$\delta_{\text{ib}}^b$				$\delta_{3'\text{b}}^b$				$\delta_{5'\text{b}}^b$			
	A	G	C	T	A	G	C	T	A	G	C	T	A	G	C	T
H1'	5.19	5.25	5.48	5.80	1.25	1.0	0.65	0.45	−0.6	−0.6	−0.3	−0.3				
H2'	2.28	2.28	2.28	2.28	0.3	0.3	−0.2	−0.2								
H2''	2.47	2.47	2.47	2.47	0.3	0.2	0.1	0.0								
H3'	4.72	4.72	4.72	4.72	0.2	0.2	0.0	0.0								
H4'	4.17	4.17	4.17	4.17	0.1	0.1	0.0	0.0								
H5'	4.02	4.02	4.02	4.02	0.15	0.15	0.0	0.0								
H5''	3.98	3.98	3.98	3.98	0.15	0.15	0.0	0.0								
H8	8.50	8.11											c	c	c	c
H6			7.70	7.64									c	c	c	c
H5			6.21										−0.8	−0.8	−0.5	−0.5
H5				2.00									−0.5	−0.5	−0.3	−0.3
H2 <sup>d</sup>	8.60								−1.5	−1.0	−0.8	−0.8				

<sup>a</sup> Structural chemical shift effects in DNA [50]. The chemical shift can be estimated using these structural chemical shift effects from,  $\delta_{\text{est}} = \delta_{\text{ref}} + \delta_{\text{ib}} + \delta_{3'\text{b}} + \delta_{5'\text{b}}$ , where  $\delta_{\text{ref}}$  is the reference shift, i.e. the chemical shift in the absence of ring current, magnetic anisotropy and electric field effect,  $\delta_{\text{ib}}$  is the shift from the own base, it is calculated assuming that the glycosidic torsion angle is  $240^\circ$  and the sugar is S-puckered,  $\delta_{3'\text{b}}$  is the experimentally observed shift resulting from the presence of a 3'-neighboring base in DNA B-type helices, and  $\delta_{5'\text{b}}$  is the experimentally observed shift resulting from the presence of a 5'-neighboring base in DNA B-type helices. The uncertainty, i.e. expected standard deviation, in the estimation of  $\delta_{\text{est}}$  is about  $\pm 0.2$  ppm giving the overall correspondence between calculated and experimentally observed shifts.

<sup>b</sup> The chemical shift of the H2 protons is mainly affected by the 3'-neighboring base, although some effect of the 5'-neighboring base can also be observed.

<sup>c</sup> 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  $\delta_{3'\text{b}}$  and  $\delta_{5'\text{b}}$  needs to be taken. For the different base sequence combinations one finds:  $\delta_{\text{exp}}$  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  $\delta_{3'\text{b}} + \delta_{5'\text{b}}$  equals  $-0.15$  to  $0.2$  ppm for N = A, G, C and T;  $\delta_{\text{exp}}$  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  $\delta_{3'\text{b}} + \delta_{5'\text{b}}$  equals  $-0.45$  to  $0.5$  ppm for N = A, G, C and T;  $\delta_{\text{exp}}$  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  $\delta_{3'\text{b}} + \delta_{5'\text{b}}$  equals  $-0.35$  to  $0.4$  ppm for N = A, G, C and T;  $\delta_{\text{exp}}$  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  $\delta_{3'\text{b}} + \delta_{5'\text{b}}$  equals  $-0.30$  to  $0.35$  ppm for N = A, G, C and T.

indicators, but they can also be quite helpful in resonance assignment. The structural shifts described above, and given in Table 6, make it possible to assign proton resonances of the sugar protons as well as of the base protons to the middle one of a triplet of residues, for example RYR, YYR, etc. where R and Y stand for purine and pyrimidine, respectively. This knowledge was sufficient to obtain unambiguous sequential proton assignments for a hairpin of 16 residues [50].

### 6.5. $^{15}\text{N}$ and $^{13}\text{C}$ shifts in RNA and DNA

The review paper by Giessner-Prettre and Pullman [127] contains, in addition to proton shifts, a wealth of information with regard to heteronuclear shifts. At the time of writing the structural information was rather limited, which hindered a comparison between experimental and calculated shifts. Since then a large body of experimental chemical shifts have been collected of structurally well characterized

Fig. 12. Correlations between calculated shifts and observed structural shifts for non-exchangeable protons in DNA (Ref. [50]). (A) 'All effects': calculated shifts 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 only': calculated shifts 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 [130] were used without adjustment; the same applies for the parameters for the electric field term (see text).

molecules, in particular of proteins. For proteins it has now been found that  $^{13}\text{C}$  shifts can be calculated with good accuracy, as evidenced by a comparison with the experimental chemical shifts of known structures [117]. These studies show that the  $^{13}\text{C}$  shifts are mainly affected by the local geometry and the electric field effect which can be quite long range in nature. These results also put the older chemical shift calculation data for heteronuclei in nucleic acids as described by Giessner-Prettre and Pullman into a new perspective. They may be expected to be more reliable than originally thought, but various aspects still have to be tested experimentally; for example, the dependence of the shifts of the carbon resonances in the ribose ring on sugar pucker; the dependence of the shifts of carbon resonances of the base moiety on electric field, hydrogen bonding and changes in the glycosidic torsion angle  $\chi$ . Ghose et al. [134] have calculated the C8 chemical shift as a function of the glycosidic torsion angle  $\chi$  and found very significant variations (up to 5 ppm). Apart from these more subtle effects of local geometry both the  $^{13}\text{C}$  and  $^{15}\text{N}$  base resonances can shift up to 70 ppm on protonation, for example, the Adenine N1 and C2 resonances [48].

#### 6.6. $^{31}\text{P}$ shifts

Gorenstein and coworkers (see, for example, Ref. [90]) have extensively studied the relationship between the  $^{31}\text{P}$  chemical shift and backbone torsion angles  $\epsilon$  and  $\zeta$ . They observed that for DNA a correlation exists between the  $\epsilon$  and  $\zeta$  torsion angles. Regular B-helix DNA can exist in a BI conformation defined by  $\epsilon$  and  $\zeta$  in the gauche range and the trans range, respectively, or in a minor BII conformation defined by  $\epsilon$  and  $\zeta$  in the gauche range and the trans range, respectively. Experimentally, the  $^{31}\text{P}$  shift shows a linear correlation with changes in  $\epsilon$ . Quantum mechanical chemical shift calculations were performed to demonstrate that the change in  $^{31}\text{P}$  shift is linearly related to changes in the torsion angle  $\zeta$ . From these data it was inferred that a change in  $\epsilon$  affects a corresponding change in  $\zeta$ . Giessner-Prettre and Pullman [127] have critically discussed these  $^{31}\text{P}$  changes. Alternative calculations have demonstrated that  $^{31}\text{P}$  shifts may just as easily result from bond angle changes or be due to counterions bound close to the phosphate group. Thus, although  $^{31}\text{P}$  shifts are

likely to correspond to torsion angle changes they may also be the result of other effects. In fact, in hair-pin loop structures changes in  $^{31}\text{P}$  shift are observed, but not all  $^{31}\text{P}$  changes can be attributed to changes in  $\epsilon$ ,  $\zeta$  and  $\alpha$  (see, for example, Ref. [11]). At present therefore the consensus is to take the conservative approach. The torsion angles  $\epsilon$ ,  $\zeta$  and  $\alpha$  are assumed to be in the regular helix ranges, trans, gauche – and gauche –, respectively, only when  $^{31}\text{P}$  is not shifted (see, for example, Refs. [36,135]).

### 7. Assignment methods

Assignment of resonances is the first essential step in an NMR study of biomolecules aimed at determining their three-dimensional structure. The recent development of  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling has made it possible to devise heteronuclear through-bond assignment methods for nucleic acids that are so successful in the field of NMR of proteins. As a result the possibilities for structural studies by NMR have been considerably enhanced. We will discuss in detail the novel through-bond assignment methods for labeled nucleic acids. In addition, we describe NOE-based assignment approaches that employ  $^{13}\text{C}$  and  $^{15}\text{N}$  edited NOESY spectra. For these heteronuclear methods we pay particular attention to the practical aspects of setting up the experiments, i.e. optimal delay settings, etc. We will, however, first briefly outline the approaches that are used in the case of unlabeled compounds.

#### 7.1. Assignment without isotopic labeling

For unlabeled compounds the  $^1\text{H}$  sequential resonance assignment is naturally based on a combination of through-bond  $^1\text{H}$  to  $^1\text{H}$  and NOE connectivities augmented with  $^1\text{H}$  to  $^{31}\text{P}$  connectivities. Although heteronuclear NMR experiments, involving  $^{13}\text{C}$  or  $^{15}\text{N}$  at natural abundance, can also be performed, they are of limited value because of their inherent lower sensitivity (see below). Since the assignment strategy for unlabeled nucleic acids has been described in detail on various occasions (see, for example, Wijmenga et al. [62]), we will here only briefly outline the main steps and concentrate on a discussion of the main uncertainties associated with such an NOE-based assignment.

The assignment for unlabeled nucleic acids consists of essentially four steps, as sketched in Fig. 13(A):

1. Assignment of imino and amino resonances via sequential NOE imino to imino and amino to imino contacts to establish the base pairing, i.e. the secondary structure (see Fig. 2 for an overview of the relevant short distances).
2. Partial assignment of non-exchangeable base and sugar proton resonances via NOE imino (and amino) to H2, H6, H8, H5(M) and H1' contacts (see Fig. 2 for an overview of the short distances). In A.T and A.U base pairs the NOE cross peak between imino (T,U) is very intense, because of the short intra-base pair distance,  $d(\text{AH2};\text{T}/\text{UNH3})$  (see Fig. 2), so that A H2 resonances can easily be identified from the imino resonances. In A-type helices the H2 protons have short cross-strand distances to H1' protons, as shown in Fig. 2, which allows an easy means of identification of H1' resonances (but see also, for example, Ref. [136]). Unfortunately, in B-type helices such short distances do not exist, so that this route does not really provide a means for further assignment of non-exchangeable proton resonances. Generally, NOE cross peaks are observed between imino resonances to H6, H5 (M) resonances in pyrimidines, which result from magnetization transfer via spin diffusion through the amino resonances.
3. The sugar proton spin systems (H1' to H3') and the Cytosine/Thymine/Uridine proton spin systems (H6 to H5/M) are identified in ( $^1\text{H}, ^1\text{H}$ ) COSY or ( $^1\text{H}, ^1\text{H}$ ) TOCSY spectra (via through-bond coherence transfer). In TOCSY spectra of DNA, the cross peaks are between Cytosine H6 and H5 resonances, a direct and easy way to identify these residues; the cross peaks between H6 and methyl resonances provide the identification of Thymine residues. In TOCSY spectra of RNA, the H6 to H5 cross peaks of Cytosine and Uracyl reside in the same spectral region, so that one cannot distinguish between these types of bases. These TOCSY cross peaks do however still allow them to be distinguished from the Guanine and Adenine bases. Depending on the sugar puckering one may be able to identify the sugar spin system more or less completely from ( $^1\text{H}, ^1\text{H}$ ) TOCSY spectra. In DNA the TOCSY transfer route from H1' up to H3' is always open independent of the sugar puckering, because of the presence of the H2'' proton (see Section 5). For DNA the sugar ring is usually in the S-type conformation, which is characterized by a small  $^3J_{3',4'}$ -coupling, so that TOCSY transfer can only occur from H1' up to H3'. To be able to identify the other sugar proton resonances the sugar ring should be N-puckered or consist of a mixture of N- and S-type puckers, so that transfer up to H4' can be achieved. Transfer up to H5' and H5'' may even be achieved when in addition the  $\gamma$  torsion angle is not gauche +. In RNA the sugar pucker is generally N-type, leading to small  $^3J_{1',2'}$ -couplings, so that TOCSY transfer cannot be achieved from H1' into the system, rendering the identification of H2' protons more difficult. On the other hand, for S-type sugar conformations or when the sugar puckering state is a mixture of N- and S-type puckers, transfer from H1' up to H4' can be achieved. This difference in transfer between N-type and S-type sugars allows an easy and direct means of identifying the sugar puckering in RNA. The identification of these spin systems aids in the subsequent sequential assignment of the non-exchangeable base and sugar proton resonances via sequential NOE contacts, i.e. the set of contacts involving H6/H8 and H1' resonances and/or the set involving the H6/8 and H2'/2'' resonances.
4. Assignment of  $^{31}\text{P}$  and H3', H4' and H5'/5'' resonances via ( $^1\text{H}, ^{31}\text{P}$ ) HETCOR experiments. Depending on the results obtained in step 3 these proton resonance assignments constitute either an extension or confirmation. The  $^{31}\text{P}$  to H3' and H5'/H5'' correlations in HETCOR experiments can in favorable cases, i.e. no overlap and large enough  $J_{\text{HP}}$ -couplings, provide sequential contacts in a through-bond fashion [137,138].

There are two main problems associated with sequential assignment of proton resonances in unlabeled nucleic acids. The first is resonance overlap. The second is related to the fact that the assignment is based mainly on NOE contacts, which are inherently ambiguous with regard to, for example, intra- and internucleotide contacts. In addition, NOE-based sequential assignment requires assumptions with regard to the conformation of structural

elements of the sequence studied. In contrast, sequential assignment based on through-bond connectivities does not suffer from these two drawbacks. It would seem to be of value, therefore, to critically discuss for the main assignment steps when and where these problems show up.

The first problem, resonance overlap, is mostly encountered in the sequential assignment of the H6/H8 and sugar resonances (step 3). Particularly strong overlap is seen for the sugar ring proton resonances H3' to H5'' in DNA and H2' to H5'' in RNA. A number of ways exist to resolve (at least partly) this overlap in unlabeled compounds. The first possibility is to use multi-dimensional homonuclear NMR spectroscopy, such as 3D TOCSY–NOESY, and/or 3D NOESY–NOESY. A single NOE cross peak is in a 3D TOCSY–NOESY spectrum spread out in a third dimension into a number of cross peaks. This set of cross peaks, corresponding to the H1' to H5'/5'' protons, results from TOCSY transfer through the sugar spin system, and serves as a fingerprint of the ribose system [62,75,77,114]. From such a spectrum the virtually complete assignment has been made in a 24-residue RNA duplex containing a G.A tandem base pair [75], and a large number of NOE constraints were derived, leading to a high resolution structure [76,77]. Another example is represented by a 31-residue intramolecular DNA triple helix, for which the proton resonance assignment could be obtained using, for example, 3D TOCSY–NOESY spectroscopy [12,139]. Gorenstein et al. have used 3D NOESY–TOCSY and 3D NOESY–NOESY spectroscopy to derive high resolution structures of DNA duplex molecules (see, e.g. Ref. [72]), and perform structural studies of DNA three-way junctions (see, e.g. Ref. [140]). An alternative approach to resolving assignment problems associated with resonance overlap is to build into the sequence structural elements of known fold and for which the resonances positions are known, a method which has been employed by Altona et al. with good success in their studies of DNA three- and four-way junctions (see, e.g. Refs. [42–44,141,142]).

The inherent ambiguity of NOE-based assignment shows up in each of the steps. It is the least prominent in the first step of the assignment process (Fig. 13(A)), for a number of reasons. Firstly, only one imino resonance is observed per base pair. In addition, the imino

resonances reside in a separate spectral region with relatively little overlap. Furthermore, their spectral positions within this region are well known in regular Watson–Crick base pairs and for other non-standard base pairs (Table 5). Consequently, the presence of an imino proton resonance in a particular chemical shift region is good evidence for the formation of a base pair. Sequential assignment in these structural elements can be achieved with a high level of certainty, since in canonical A- and B-helices the sequential contacts are quite well established. Uncertainties can

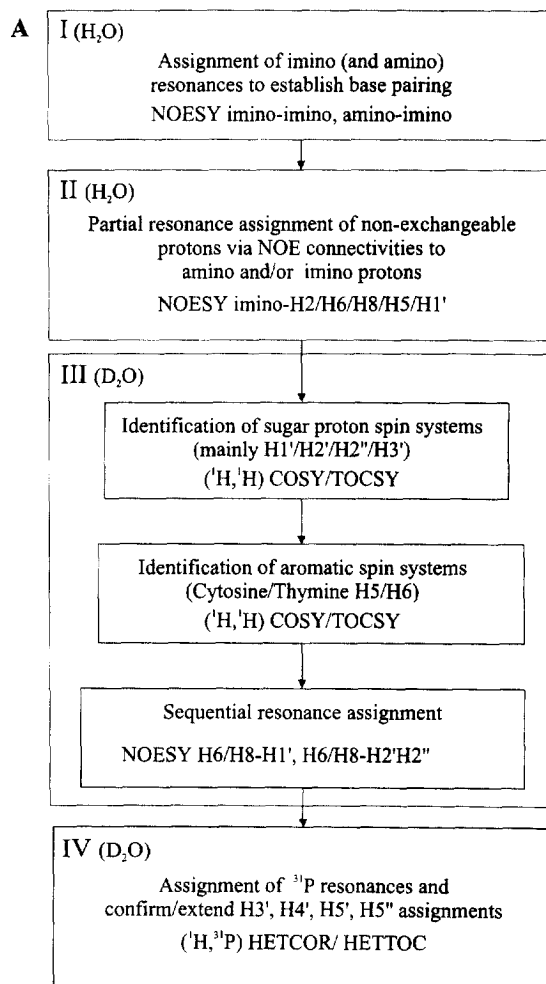


Fig. 13. Flowcharts for resonance assignment in unlabeled and labeled nucleic acids. (A) NOE-based assignment in unlabeled compounds; (B) NOE-based assignment in labeled compounds; (C) assignment via through-bond coherence transfer in labeled compounds.

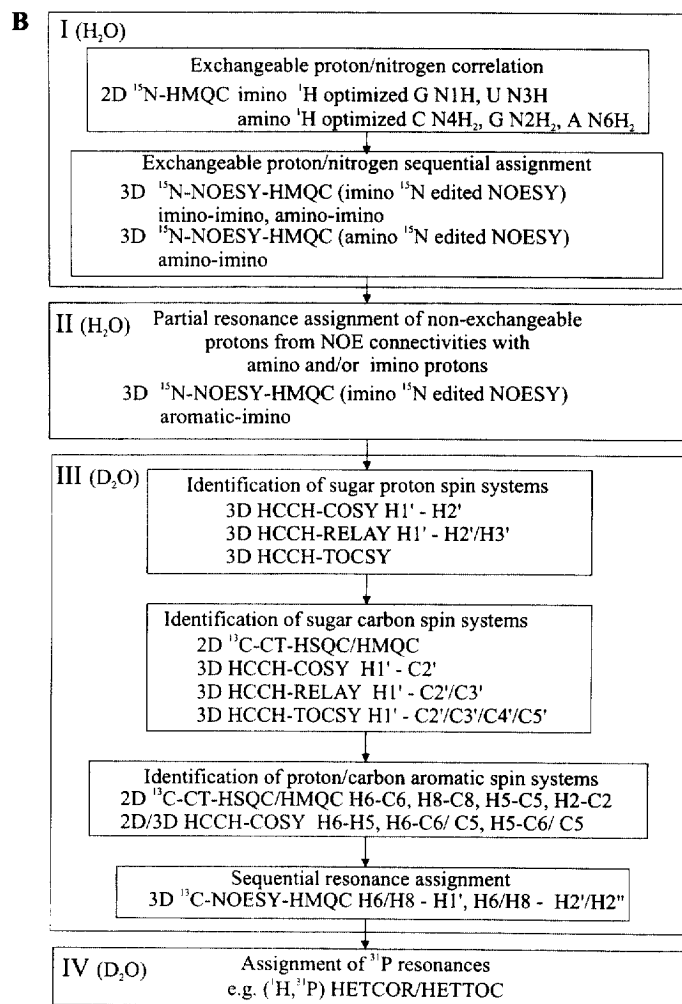


Fig. 13. (Continued).

arise for non-canonical structures such as tetramers, triple helices, and three- or four-way junctions [11,16,143]. In these structural elements unusual base pairing may be found. The imino proton resonances may then show up at unusual resonance positions and/or can be broadened. In addition, the sequential NOE contacts can be quite different from those found in a canonical helical environment.

In the second assignment step, NOE connectivities from imino and amino protons to non-exchangeable protons are used. Here again no uncertainty exists with regard to assignment if the conformation is a regular A- or B-type helix, but it is a different matter in non-canonical environments. In the latter case

uncertainties arise because assumptions must be made with regard to the conformation and the correctness of the assignment must be based on the internal consistency of NOE contacts and on a physically reasonable model. We consider as an example the H2 protons. In an A.U/T Watson–Crick the most intense and thus the most reliable is an intra base pair NOE contact from U/T NH3 to A H2. When non-standard G.A base pairs may be present, a number of possible configurations may be assumed for them [144]. In the G.A anti–anti base pair [144], the same type of intra base pair NOE contact is present as in a Watson–Crick A.T/U base pair, namely from G NH1 to A H2. On the other hand, in

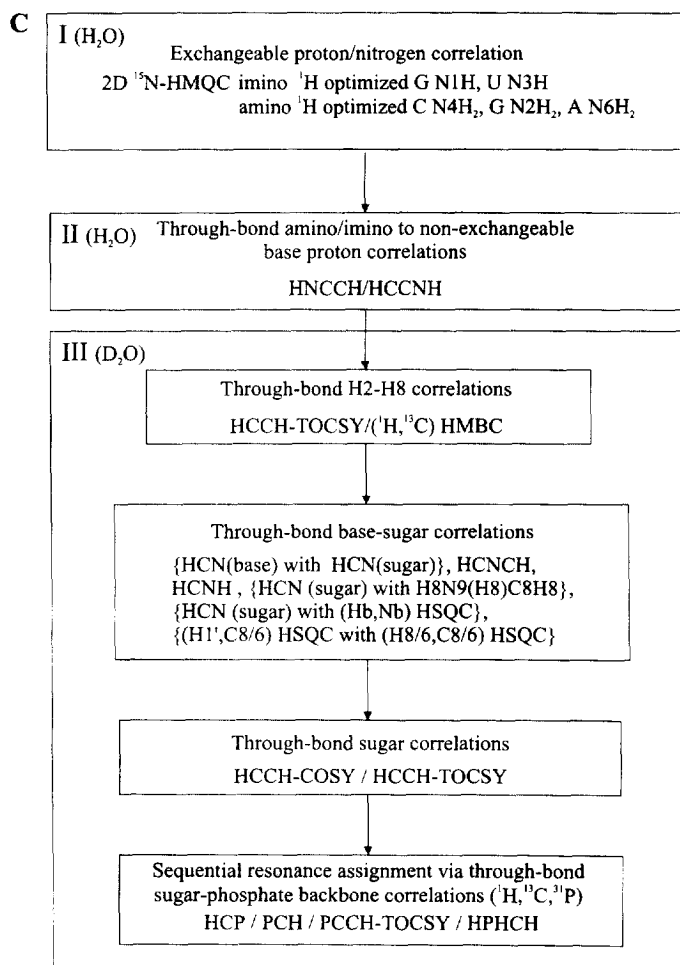


Fig. 13. (Continued).

a sheared G.A base pair this contact is absent [144]. In short, for both situations, i.e. for standard Watson–Crick base pairs and for G.A base pairs, the H2 proton is an important stepping stone on the route from imino proton resonance assignment to the assignment of the non-exchangeable proton resonances. Unfortunately, as discussed above, the H2 proton often shows a rather complex and/or limited set of NOE contacts to other non-exchangeable protons (see Section 4 and Ref. [136]). It is therefore of limited use for further sequential NOE-based assignment of non-exchangeable protons. The Adenine H8 protons have, on the other hand, an extensive set of intra- and inter-residue (sequential) NOE contacts. Through-bond H2 to H8 correlations can fortunately be obtained not only for  $^{13}\text{C}$  labeled,

but also for unlabeled compounds, as was recently shown in Ref. [144], providing a reliable link between the H2 and other non-exchangeable protons.

In the third step, sequential and intra-residue NOE contacts between H6/8 and H1'/H2'/H2' are used. Similar arguments to those above apply here with regard to uncertainties in the assignment. NOE contacts are well established and reliable in a regular A- or B-type helix environment, but not for other conformations; the correctness of the ultimate assignment must be based on internal consistency of the NOE contacts and a physically reasonable model. Fortunately, the NOE network for the H8/6/5/M and sugar ring protons is quite dense and contains a large number of NOE contacts related to conformation



independent distances (see Section 4), providing a large number of internal checks on the internal consistency of the assignment. Furthermore, as discussed above, through-bond proton–proton correlations can be obtained in the pyrimidine bases (H6 to H5 or M), so that they can be identified and distinguished from purine bases, via COSY or TOCSY experiments. The H6 proton resonances of Cytosine have a characteristic doublet structure, whereas H8 resonances have a singlet structure, so that they can often be distinguished in NOESY spectra. Through-bond proton–proton correlations in the sugar ring may serve as a fingerprint for the ribose spin system, a feature used to great advantage in 3D TOCSY–NOESY spectra. Finally, chemical shifts of the base and ribose sugar protons can help in the assignment, as they show distinct patterns depending on the type of sequence (see Section 6). Nevertheless, the only way to ascertain base to sugar ring contacts without any ambiguity is via through-bond connectivities. These are relatively easy to generate in  $^{13}\text{C}/^{15}\text{N}$  labeled compounds. They could in principle also be derived in unlabeled compounds, but their derivation suffers from the low sensitivity of ( $^1\text{H}$ ,  $^{13}\text{C}/^{15}\text{N}$ ) HETCOR experiments at natural abundance.

Through-bond correlations are used in the fourth assignment step. Thus, these assignments do not suffer from the uncertainties associated with NOEs. Here, it is the extreme overlap that often prevents full assignment.

When using an assignment approach based on NOEs, proof for the correctness of the assignment has to come from the internal consistency of the NOE contacts and from agreement with a physically reasonable model. One might at first glance expect that in the subsequent structure calculations further confirmation of the assignment could easily be obtained. This may not always be the case, however, since simulated annealing protocols can be quite forgiving with regard to a few wrong NOEs. This is important, since structural features derived by means of NMR studies often depend on a few NOE contacts. Various examples existing in the literature show that a few wrong assignments have led to wrong structural features [16,145–147]. This is a very serious problem. It is therefore of the utmost importance that correct assignments are obtained, and that the ultimately derived set of structures is checked, beyond the usual means of precision (rmsd's) and the number of

violations, to see whether the structures are physically reasonable. It also means that the arguments for the key resonance assignments should be presented.

As follows from the discussion given above, resonance assignment in unlabeled compounds can be quite difficult, due to the extensive overlap of the proton resonances of the sugar ring and the fact that NOEs are inherently ambiguous with regard to intra- and internucleotide contacts. Isotope labeling opens up the way for a suite of through-bond assignment strategies offering a variety of options for establishing unambiguous intra- and internucleotide contacts. Nevertheless, the NOE-based assignment strategy is extremely powerful and quite a number of detailed structures have been derived with this method (see, for example, Refs. [12,68,69,72,76,77]).

## 7.2. Assignment with isotope labeling

The recent possibility of  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling has signaled the introduction of alternative and improved assignment schemes, which aim to resolve the two main problems associated with resonance assignment of unlabeled nucleic acids, namely, the resonance overlap and the ambiguity associated with NOE-based assignments. In the first scheme shown in Fig. 13(A), which is a slightly adapted version of the one proposed by Nikonowicz and Pardi [148], the assignment process is improved mainly via reduction of resonance overlap, i.e. 2D NOESY experiments are replaced by  $^{15}\text{N}$ - or  $^{13}\text{C}$ -edited NOESY experiments (compare steps I, II and III in Fig. 13(A) and (B), respectively). In addition, in step III the through-bond proton–proton correlations are replaced or augmented by through-bond carbon–carbon correlations (compare step III in Fig. 13(A) and (B), respectively). In Fig. 13(C) an assignment scheme is shown, in which as many as possible of the NOE-based correlations are replaced by through-bond correlations. In the following we will first discuss the elements of the assignment scheme shown in Fig. 13(B), and subsequently, the elements of the assignment scheme shown in Fig. 13(C). Several aspects of the through-bond assignment methods have recently been reviewed [36,46–48]. We will try to give a complete overview on the one hand and on the other concentrate on matters of sensitivity and implementation. The latter can now be

done, because only recently a complete overview of all the heteronuclear  $J$ -couplings in  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled nucleic acids has appeared [49]. In the light of these results we will discuss the proposed through-bond experiments and describe their transfer functions in order to assess their relative sensitivities.

### 7.2.1. NOE-based correlation

Traditionally, sequential assignment in unlabeled nucleic acids has been based on sequential and intra-residue NOE connectivities, augmented with intra-sugar ring through-bond  $^1\text{H}$  to  $^1\text{H}$  connectivities and  $^1\text{H}$  to  $^{31}\text{P}$  connectivities (see Fig. 13(A)). Nikonowicz and Pardi [148] proposed a complete strategy for sequential resonance assignment in  $^{13}\text{C}/^{15}\text{N}$  labeled RNAs based mainly on NOE connectivities. The main steps in this scheme are illustrated in Fig. 13(B); this scheme shows some changes compared to the original scheme proposed by Pardi et al., for example the order of the assignment steps III and I/II is reversed and step IV is missing. We have adapted the original scheme in this way to emphasize the similarities with the assignment in unlabeled compounds, and to indicate the usual flow in the assignment process, which goes from imino/amino assignment to assignment of non-exchangeable protons. In the original, the assignment starts (step III in Fig. 13(B)) with identifying all protons and carbons belonging to the same sugar ring through application of a set of 2D and 3D heteronuclear HCCH–COSY NMR experiments [148,149]. Note that the sugar ring systems are identified by 2D and 3D HCCH experiments instead of (H,H) COSY or (H,H) TOCSY experiments as used in the scheme for unlabeled nucleic acids. This has the advantage that resonance overlap is reduced and, very importantly, that for the through-bond coherence transfer the large conformation independent (42 Hz)  $J_{\text{CC}}$ -couplings are used, instead of the smaller (2–8 Hz) conformation dependent  $J_{\text{HH}}$ -couplings. Next, the individual sugar rings are connected to their corresponding bases through intra-residue proton–proton NOE contacts observed in a 3D ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^1\text{H}$ ) NOESY–HMQC (see Fig. 14(A)). Sequential NOE connectivities observed in this experiment are used to assign each residue in the nucleotide sequence. The imino/amino protons are assigned by 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HMQC and 3D ( $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^1\text{H}$ ) NOESY–HMQC experiments in  $\text{H}_2\text{O}$  (step I in

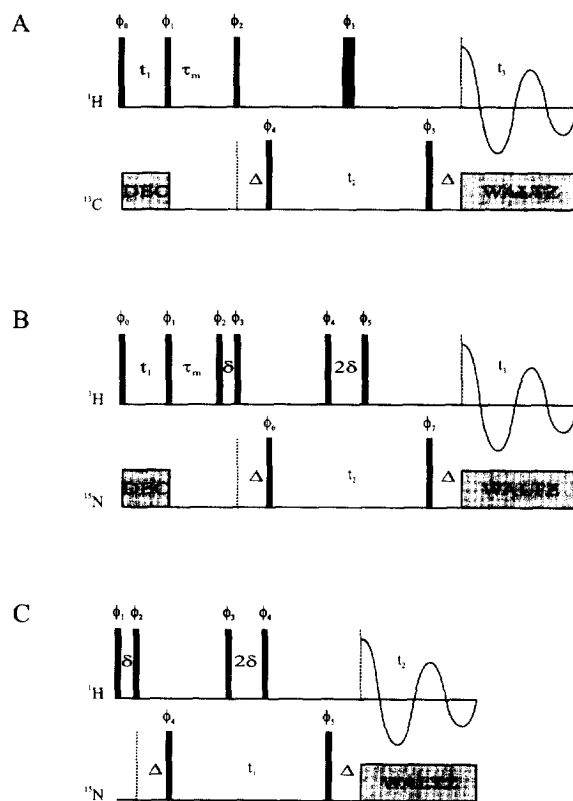


Fig. 14. The NOESY–HMQC sequences and HMQC sequence. The thin and thick filled bars indicate  $90^\circ$  and  $180^\circ$  pulses; the filled grey boxes indicate decoupling. (A) The NOESY–HMQC sequence for use in  $\text{D}_2\text{O}$ ; (B) the NOESY–HMQC sequence for use in  $\text{H}_2\text{O}$ , the water suppression is achieved via  $(1, -1, \text{echo})$  incorporated into the HMQC sequence [148]; (C) HMQC sequence for use in  $\text{H}_2\text{O}$ , with a  $(1, -1, \text{echo})$  for water suppression [148]. In the  $(1, -1, \text{echo})$  sequence, the carrier should be at the water resonance position,  $\nu_{\text{H}_2\text{O}}$ ;  $\delta$  should be set to  $(4(\nu_{\text{H}_2\text{O}} - \nu_{\text{obs}}))^{-1}$ , where  $\nu_{\text{obs}}$  is the frequency corresponding to the part of the spectrum that is to be observed, e.g. for imino resonance observation set  $\delta$  in the range 50–70  $\mu\text{s}$ . The delay  $\Delta$  is for creating an anti-phase signal, and should be approximately  $1/(2J_{\text{HX}})$ ;  $\tau_m$  is the NOESY mixing time. The minimal phase cycling is in (A):  $\phi_0 = (x, -x) + (\text{TPPI or States-TPPI})$ ;  $\phi_1 = -x$ ;  $\phi_2 = x$ ;  $\phi_3 = x$ ;  $\phi_4 = (x, x, -x, -x) + (\text{TPPI or States-TPPI})$ ;  $\phi_5 = x$ ; receiver =  $x, -x, -x, x$ ; in (B):  $\phi_0 = 8(x, -x) + (\text{States-TPPI})$ ;  $\phi_1 = -x + 45^\circ$ ;  $\phi_2 = 16x$ ;  $\phi_3 = 16(-x)$ ;  $\phi_4 = 2(2x, 2y, 2(-x), 2(-y))$ ;  $\phi_5 = 2(2(-x), 2(-y), 2x, 2y)$ ;  $\phi_6 = 4(x, x, -x, -x) + (\text{TPPI or States-TPPI})$ ;  $\phi_7 = x$ ; receiver =  $4(x, -x, -x, x)$ ; in (C):  $^1\text{H}$ :  $\phi_1 = 8x$ ;  $\phi_2 = 8(-x)$ ;  $\phi_3 = (2x, 2y, 2(-x), 2(-y))$ ;  $\phi_4 = (2(-x), 2(-y), 2x, 2y)$ ;  $^{15}\text{N}$ :  $\phi_4 = 4(x, -x) + (\text{TPPI or States-TPPI})$ ;  $\phi_5 = x$ ; receiver =  $4(x, -x)$ . The phase of  $\phi_1$  in (B) is set to  $x + 45^\circ$  to distribute the effect of radiation damping by the water signal (see Section 7.2.3).

Fig. 13(B); the pulse sequences of the two experiments are given in Fig. 14(B) and (C), respectively). These experiments also provide connectivities to non-exchangeable protons (step II in Fig. 13(B)).

### 7.2.2. Through-bond correlation

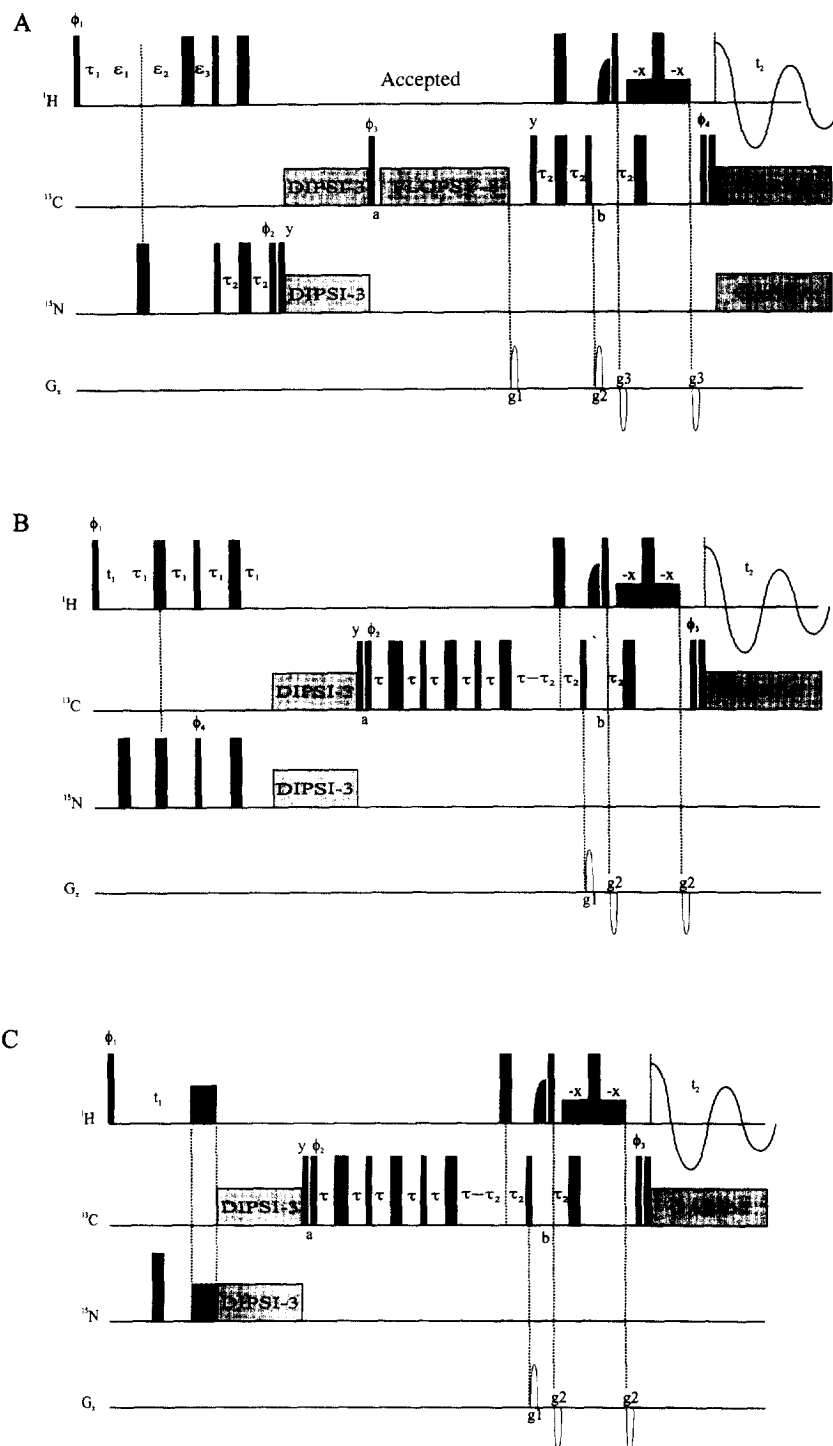
For a description of the through-bond correlation experiments we follow the assignment scheme as sketched in Fig. 13(C). In order to be able to indicate the efficiency of the various experiments we will first discuss the coherence transfer functions used.

**7.2.2.1. Coherence transfer functions.** The efficiency of a 2D or 3D NMR experiment can be calculated via coherence transfer functions. Assume an  $nD$  experiment in which  $m$  transfer steps occur of duration  $\tau_m$  and interspaced with  $p$  evolution periods. We will denote the coherence transfer efficiency for each step by  $T_m$ .  $T_m$  is the product of the transfer efficiency per se,  $F_m$ , and the effect of relaxation of the relevant coherences,  $T_{2m}$ .  $T_m$  is then given by  $T_m = F_m \exp(-\tau_m/T_{2m})$ . The term  $F_m$  can be calculated exactly. For pulse sequences using INEPT type coherence transfers it is simply a product of sin and cos terms; for TOCSY transfers analytical expressions can only be derived for simple two spin systems, whilst for larger spin systems TOCSY transfers have to be calculated numerically. The overall efficiency of the  $nD$  NMR experiment,  $T_{nD}$ , is the product of all  $T_m$  values,  $T_{nD} = T_1 \cdots T_m$ . However, in each transfer step, losses,  $T_{\text{rf}}$ , due to imperfect settings of, for example,  $180^\circ$  or  $90^\circ$  pulses and rf inhomogeneity may be incurred. We collect them together in one term,  $T_{\text{rf}}$ . With this term included the overall efficiency of an  $nD$  NMR experiment becomes  $T_{nD} = T_{\text{rf}} T_1 \cdots T_m$ . The  $T_{\text{rf}}$  term is difficult to assess exactly, although it can be reasonably estimated from the properties of a probe.  $T_{\text{rf}}$  will become smaller as the number of pulses applied in an  $nD$  NMR experiment increases. For modern probes the rf homogeneity is so good, i.e. the signal ratio is around 0.8, when comparing a  $720^\circ$  to a  $90^\circ$  pulse, that the term  $T_{\text{rf}}$  is expected to lie around 0.8 for most multi-pulse NMR experiments. With these considerations in mind the efficiency calculations excluding the effect of the  $T_{\text{rf}}$  term should still give a reasonable estimate of the actual efficiencies of an  $nD$  NMR experiment.

**7.2.2.2. Through-bond amino/imino to non-exchangeable proton correlation.** Until now three sets of pulse sequences have been proposed for the correlation of exchangeable and non-exchangeable protons via through-bond coherence transfer (step II in Fig. 13(C)). The first set is by Simorre et al. [150,151], a second set by Sklenar et al. [152], and a third by Fiala et al. [153]. The proposed sequences have large overall similarities, but differ in the parameter settings used.

The first sequence of Simorre et al. [151] (Fig. 15(A)) seeks to correlate guanosine imino protons and H8 protons in  $^{13}\text{C}/^{15}\text{N}$  labeled RNAs via an HNC–TOCSY–CH experiment. The experiment starts with imino proton coherence which is frequency labeled during  $t_1$  and then transferred to  $^{15}\text{N}$  via a refocused INEPT sequence. The in-phase N1 coherence is subsequently transferred to C6 and C2 via a (C,N) DIPSI cross-polarization sequence of 44 ms duration at a radio frequency field strength (rf field strength) of 1.9 kHz and with the carriers of  $^{15}\text{N}$  and  $^{13}\text{C}$  set at 146 and 161 ppm, respectively. The rather long mixing time is required because of the relatively small  $J$ -couplings from N1 to C6 and from N1 to C2 ( $\approx 7$  Hz, see Fig. 3). Subsequently, a 37.8 ms long FLOPSY8 (C,C) TOCSY sequence is employed, with an rf field strength of 5 kHz and the carrier at 145 ppm, to transfer the in-phase C6 or C2 coherence to C8. Since the C5 and C6 carbons reside at approximately 120 and 160 ppm, respectively, a carrier position at 145 ppm corresponds for C5 and C6 to offsets of 3750 and 2250 kHz. Considering these settings and the  $J$ -couplings in the guanosine base (see Fig. 3), possible coherence transfer routes are from C6 directly to C8 or from C6 to C4 to C8, or from C6 via C5 to C4 and then to C8. In view of the small  $J_{\text{C2C5}}$ -coupling a transfer route from C2 to C5 and then either via C4 or C6 to C8 is unlikely. The in-phase C8 coherence is in the last stage transferred to H8 via a reverse refocused INEPT sequence.

The second two sequences proposed by Simorre et al. [150], Fig. 15(B) and (C)/(D), are designed to correlate U NH and C NH<sub>2</sub> with H6 in Uridine and Cytosine, respectively. In the Uridine experiment NH proton magnetization is transferred via a refocused INEPT to in-phase N3 coherence. Next, this coherence is transferred to C4 via a heteronuclear (N,C) cross-polarization. This is done in such a way



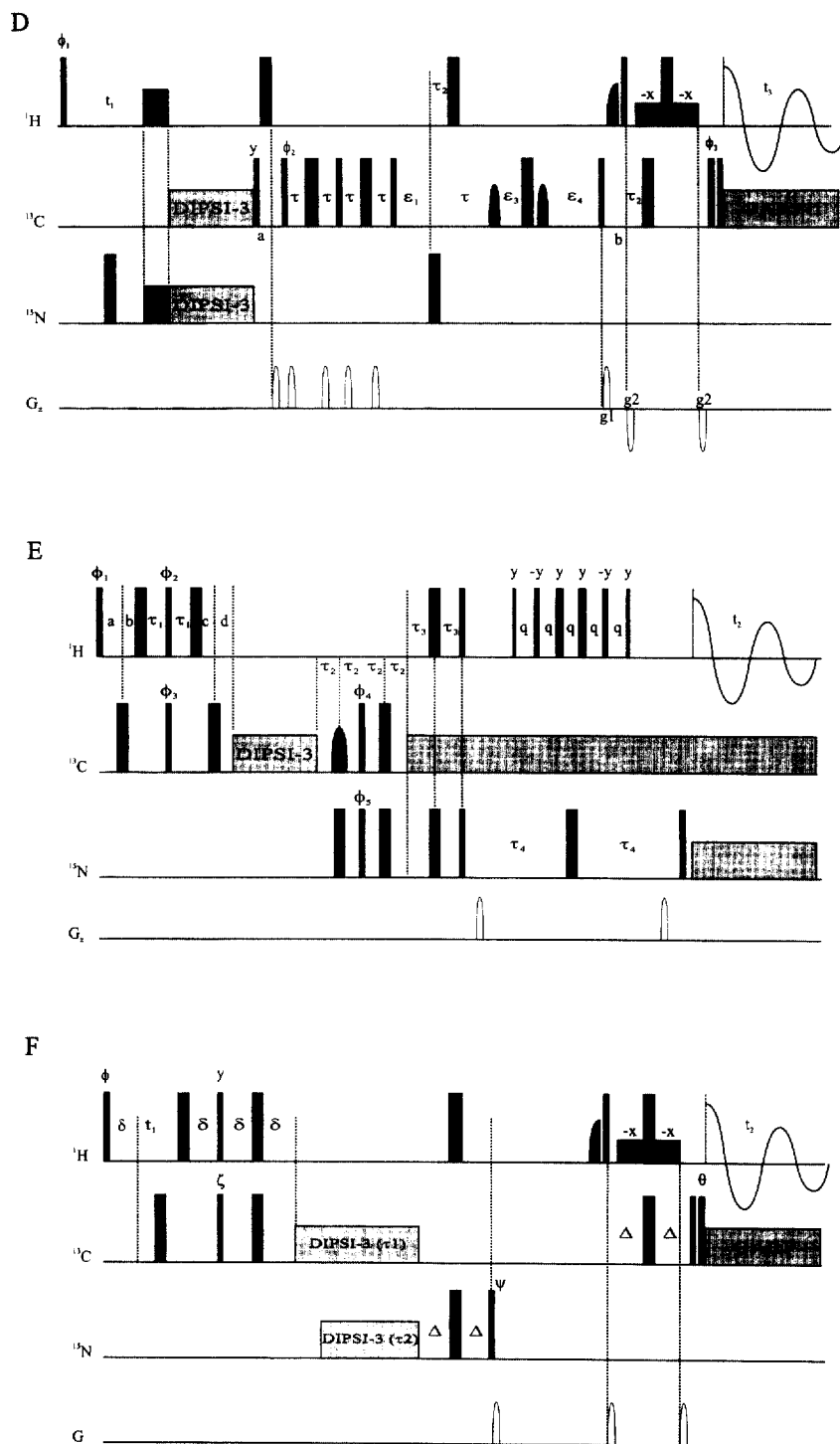


Fig. 15. (Continued).



Fig. 15. Pulse schemes for through-bond correlations of amino/imino and non-exchangeable protons. (A) The 2D H(NC)-TOCSY-(C)H experiment, used to connect imino to H8 protons in guanosines [151]. The indirect detection period  $t_1$  (see below) was partially concatenated with the first  $^1\text{H}$ - $^{15}\text{N}$  INEPT using the time-shared evolution procedure, which allows a reduction of 5 ms for the longest  $t_1$  value [234]. For the first  $t_1$  point,  $\epsilon_1 = \epsilon_2 = 0.5*(dw/2 - (4\pi)PW90(^1\text{H}) - PW180(^{15}\text{N}))$  and  $\epsilon_3 = \tau_1$ . For subsequent  $t_1$  points,  $\epsilon_1$  was incremented by  $dw/2 - t_1/(n+1)$  and  $\epsilon_2$  was decremented by  $t_1/(n+1)$ , where  $dw$  is the dwell time and  $n$  is the total number of complex points in  $t_1$ . The  $t_1$  and  $t_2$  delays were set to 2.5 and 1.25 ms respectively. The magnetization was transferred from  $^{15}\text{N}$  to  $^{13}\text{C}$  by cross polarization using a 44.9 ms DIPSI-3 sequence [235–237] at an rf field of 1.9 kHz. A 37.8 ms FLOPSY-8 sequence [238] at an rf strength of 5 kHz was used for the  $^{13}\text{C}$  homonuclear TOCSY period [239]. Unless otherwise noted, all pulses have phase  $x$ . The phase cycle was  $\phi_1 = 2(y + 45^\circ), 2(-y + 45^\circ)$ ;  $\phi_2 = y, -y$ ;  $\phi_3 = 4(y), 4(-y)$ ;  $\phi_4 = 8(x), 8(-x)$  and receiver =  $x, 2(-x), x, -x, 2(x), -x$ . Instead of phase cycling  $\phi_2$ , the two  $90^\circ$   $^{15}\text{N}$  pulses before the  $^{15}\text{N}$ - $^{13}\text{C}$  cross polarization can be eliminated and the phase of the  $^{15}\text{N}$  DIPSI-3 can be phase cycled ( $x, -x$ ) using the same receiver phase cycling as before [240]. Eliminating these two  $90^\circ$   $^{15}\text{N}$  pulses should lead to a slight improvement in sensitivity for probes with low  $^{15}\text{N}$  B1 homogeneity. The  $^{15}\text{N}$  frequency was set to 146 ppm during the  $^1\text{H}$ - $^{15}\text{N}$  and  $^{15}\text{N}$ - $^{13}\text{C}$  transfers and it was shifted to 195 ppm at point  $a$  to decouple N7 and N9 during the acquisition period. The  $^{13}\text{C}$  frequency was positioned at 151 ppm for the  $^{15}\text{N}$ - $^{13}\text{C}$  hetero-TOCSY period and was shifted to 145 ppm at point  $a$  for the  $^{13}\text{C}$ -TOCSY period and to 142 ppm at point  $b$ . To avoid a complete inversion of the water signal by the second  $^1\text{H}$   $90^\circ$  pulse, the phase of  $\phi_1$  was shifted by  $45^\circ$  and the  $^1\text{H}$  carrier was set to the water frequency. Radiation damping returns the water magnetization to the  $+z$  axis during the  $^{15}\text{N}$ - $^{13}\text{C}$  cross polarization and  $^{13}\text{C}$ -TOCSY periods. At point  $b$ , water flip-back is achieved with a 2.9 ms selective E-BURP pulse [241,242]. During the last INEPT periods, a WATERGATE sequence is applied to suppress the residual water signal using two 1.55 ms soft square pulses [243]. The phases of the selective E-BURP and soft square pulses were adjusted with a small angle phase shifter for optimal solvent suppression. During the detection period,  $^{13}\text{C}$  and  $^{15}\text{N}$  GARP1 decoupling was used at rf fields of 1.6 and 1.14 kHz, respectively. All gradients were applied along the  $z$  axis, with  $g_1 = 12$  G/cm,  $g_2 = 24$  G/cm and  $g_3 = 32$  G/cm. The gradient times for  $g_1, g_2$  and  $g_3$  were 300, 300 and 450  $\mu\text{s}$ , respectively. Each gradient was followed by a recovery time of 200  $\mu\text{s}$ . (B) The uridine optimized 2D H(NCCC)H experiment employed a refocused INEPT sequence for the  $^1\text{H}$ - $^{15}\text{N}$  transfer [150]. The  $^1\text{H}$  carrier is placed at the center of the imino proton region (13.2 ppm for uridine) at the beginning of the experiment and shifted to the water frequency at point  $a$ . A 90 ms DIPSI-3  $^{13}\text{C}$ - $^{15}\text{N}$  hetero-TOCSY mixing period was used at an rf field strength of 1.9 kHz, using  $^{13}\text{C}$  and  $^{15}\text{N}$  carrier frequencies of 169 and 160 ppm, respectively. The  $^{13}\text{C}$  carrier was shifted to 139 ppm at point  $a$ . Water magnetization returns to the  $+z$  axis by radiation damping during the  $^{13}\text{C}$ - $^{15}\text{N}$  hetero-TOCSY period. After this  $^{13}\text{C}$ - $^{15}\text{N}$  hetero-TOCSY period, the magnetization is transferred from C4 to H6 using a string of concatenated  $^{13}\text{C}$ - $^{13}\text{C}$  and  $^{13}\text{C}$ - $^1\text{H}$  INEPT type transfers. The second half of the  $^{13}\text{C}$ - $^1\text{H}$  reverse INEPT transfer has been adapted to include a WATERGATE solvent peak suppression [243]. In this scheme the proton  $90^\circ$  pulse in the center of the reverse INEPT period is preceded by a selective water flip-back  $90^\circ$  E-BURP pulse (2.8 ms) at point  $b$  [241,244]. In the final refocused INEPT interval, the proton  $180^\circ$  pulse is flanked by two square 1.67 ms  $90^\circ$  pulses of phase  $-x$ , and the carbon  $180^\circ$  pulse is placed at the interval  $t_1 = 1/(4*J_{\text{CH}})$  following the proton  $90^\circ$  pulse. The phases of the selective square  $90^\circ$  pulses and the  $90^\circ$  E-BURP pulse are further optimized for maximum solvent peak suppression, using a small angle phase shifter. The transfer delays were  $\tau = 3.6$  ms,  $\tau_1 = 2.5$  ms and  $\tau_2 = 1.25$  ms.  $G_z$  denotes the  $z$  axis pulsed-field gradients with the following values:  $g_1 = 300$   $\mu\text{s}$  at 24 G/cm and  $g_2 = 450$   $\mu\text{s}$  at 32 G/cm. A 400  $\mu\text{s}$  recovery time was added after each gradient. Quadrature detection in the proton  $t_1$  period was obtained with the TPPI-States method [245]. The 16-step phase cycle is  $\phi_1 = x, -x$ ;  $\phi_2 = y, y, -y, -y$ ;  $\phi_3 = 4(x), 4(-x)$ ;  $\phi_4 = 8(x), 8(-x)$  and receiver =  $2(x, -x, -x, x), 2(-x, x, x, -x)$ . (C) The cytidine optimized 2D H(NCCC)H [150], where the  $^{13}\text{C}$ - $^{15}\text{N}$  refocused INEPT period in the pulse sequence given in (B) is replaced with a  $^1\text{H}$ - $^{15}\text{N}$  hetero-TOCSY period. A 7.5 ms half DIPSI-3 (R, -R supercycle) sequence [237] was used for the  $^1\text{H}$ - $^{15}\text{N}$  hetero-TOCSY transfer at an rf field of 1.9 kHz with the  $^{15}\text{N}$  carrier set to 96.7 ppm and the  $^1\text{H}$  carrier before point  $a$  set to 7.6 ppm. Other changes compared to the pulse sequence given in (B) were that the  $^{13}\text{C}$ - $^{15}\text{N}$  hetero-TOCSY period was set to 45 ms ( $J_{\text{NC}} \approx 20$  Hz for cytidine instead of  $\approx 10$  Hz for uridine) and an eight-step phase

cycle was used with:  $\phi_1 = x, -x$ ;  $\phi_2 = y, y, -y, -y$ ;  $\phi_3 = 4(x), 4(-x)$ ; and receiver =  $x, -x, -x, x$ . (D) The cytidine optimized pseudo 3D H(NCC)CH experiment [150], where both the amino proton resonances and the C6 resonances are frequency labeled during the  $t_1$  evolution periods. The parameters for the cytidine optimized pseudo 3DH(NCC)CH were the same as those for the cytidine optimized 2D H(NCC)CH, except that the  $^{13}\text{C}$  carrier was positioned at 129 ppm at point  $a$  and two 1.03 ms frequency-shifted selective IBURP  $180^\circ$   $^{13}\text{C}$  pulses [241,246] were applied at 104 ppm. The first selective pulse is used to decouple C5 from C6 during the time-shared  $^{13}\text{C}$  evolution period and the second pulse at point  $b$  compensates for off-resonance effects caused by the first selective  $180^\circ$  pulse [242]. The C6 evolution period  $t_1$  is time-shared with both the  $2^*t_1$  period that refocuses the C5–C6 anti-phase component and the  $2^*t_2$  period that creates the C6–H6 anti-phase component [234]. The time periods  $\varepsilon_1, \varepsilon_3$  and  $\varepsilon_4$  were set to 0, 0 and  $t$ , respectively, for the initial  $t_1$  interval. For evolution of the C6 chemical shift during the  $t_1$  period, the increments of these intervals are  $\delta\varepsilon_1 = dw/2$ ,  $\delta\varepsilon_3 = (dw/2 - t/(n+1))$  and  $\delta\varepsilon_4 = -t/(n+1)$ , where  $dw$  is the dwell time and  $n$  is the total number of points in the  $t_1$  dimension. The  $^1\text{H}$   $180^\circ$  pulse at point  $a$  is used to invert the water magnetization and the five gradient pulses subsequently eliminate radiation damping and keep the water magnetization aligned along the  $-z$  axis until the next  $^1\text{H}$   $180^\circ$  pulse flips the water signal back to the  $+z$  axis. (E) The 3D ( $^1\text{H}, ^{13}\text{C}, ^{15}\text{N}$ ) triple resonance pulse sequence HC(CN)H TOCSY used to establish guanine imino-H8, C8 and adenine amino-H8, H2, C8, C2 correlations in  $^{13}\text{C}, ^{15}\text{N}$  uniformly labeled RNA oligonucleotides [153]. The last six proton pulses function as a selective  $180^\circ$  pulse in WATERGATE; their lengths are in the ratio 3:9:19:9:9:3 and were applied with a power of 8.9 kHz. The DIPSI-3 [237] spin lock was applied at a power of 3.0 kHz. GARP [247] decoupling was applied at a power of 2.2 kHz for  $^{13}\text{C}$  and 930 Hz for  $^{15}\text{N}$ . All  $^{13}\text{C}$  pulses were applied at 150 ppm except for the selective refocusing C6 pulse, represented by the thick rounded bar, for which a 1.5 ms REBURP [241] was frequency shifted by phase modulation [248] to cover the region between 150 and 170 ppm. The  $^1\text{H}$  frequency was set to 10.5 ppm and the  $^{15}\text{N}$  frequency to 120 ppm. Rectangular gradient pulses of 1.0 ms duration and a strength of 8 G/cm were used. Unless indicated otherwise,  $90^\circ$  pulses were applied with phase  $x$  and  $180^\circ$  pulses were applied with phase  $y$ . The phase cycle was  $\phi_1 = x + \text{States-TPPI}$ ;  $\phi_2 = 4(y), 4(-y)$ ;  $\phi_3 = x, -x + \text{States-TPPI}$ ;  $\phi_4 = 8(x), 8(-x)$ ;  $\phi_5 = x, x, -x, -x$ ; receiver:  $x, -x, -x, x, 2(-x, x, x, -x), x, -x, -x, x$ . Delays were  $a = \tau_1 + t_1/2$ ,  $b = t_1/2$ ,  $c = \tau_1 + t_2/2$ ,  $d = t_2/2$ ;  $\tau_1$  was set to 1.1 ms,  $\tau_2$  to 6.5 ms,  $\tau_3$  to 1.4 ms and  $\tau_4$  to 2.7 ms. The value of 150  $\mu\text{s}$  for  $q$  was found by optimization for the best water suppression. (F) 2D HCCNH TOCSY. (G) 3D HCCNH TOCSY experiments [152]. The INEPT evolution delays are set to match  $1/4J$ , both for  $\text{H} \rightarrow \text{C}$  ( $J \approx 18\text{--}220$  Hz) and  $\text{N} \rightarrow \text{H}$  ( $J \approx 90$  Hz) transfers ( $\delta = 1.25$  ms and  $\Delta = 2.5$  ms). In the 2D experiment, the proton carrier is centered at the water frequency. In the 3D experiment, the proton carrier is placed at the center of an aromatic region at  $\approx 7.6$  ppm for the  $t_1$  evolution period and switched to the water position for the detection during  $t_3$  in order to optimize the resolution and the size of the data matrix. DIPSI-3 mixing [237] along the  $y$  axis is used for the C–C and C–N TOCSY with an rf field of 2.9 kHz. The  $^{13}\text{C}$  and  $^{15}\text{N}$  carriers are set to 150 and 153 ppm respectively. Water flip-back WATERGATE is applied during the  $^{15}\text{N}$ – $^1\text{H}$  INEPT step, using typically a 4 ms EBURP2  $90^\circ$  pulse and 1.3 ms selective  $90^\circ$  pulses flanking the non-selective  $180^\circ$  pulse. One scrambling gradient is applied during the  $^{15}\text{N}$ – $^1\text{H}$  INEPT, when the magnetization of interest is converted into H $z$ N $z$  order. Two identical gradients are used in the WATERGATE sequence. In the 3D experiment the bipolar gradient echo [249] with an amplitude of 0.5 G/cm is used to suppress radiation damping during  $t_2$ . The phase cycle is:  $\phi$  incremented in order to obtain States–TPPI [245] quadrature detection in  $t_1$ ;  $\zeta = y, -y$ ;  $\psi = y, y, -y, -y$ ;  $q = x, x, x, x, -x, -x, -x, -x$ ; receiver =  $+, -, -, +$ . The phase of the  $90^\circ$   $^1\text{H}$  mixing pulse in the  $^1\text{H}$ – $^{13}\text{C}$  INEPT and the receiver phase are inverted after four scans. In addition, all phases of the WATERGATE and of the receiver are changed by  $180^\circ$  after eight scans for the 16-step cycle. The phase  $\psi$  is incremented according to the States–TPPI protocol to obtain quadrature detection in  $t_2$  of the 3D experiment, asynchronous GARP decoupling [247] is used to suppress the heteronuclear spin–spin interactions during the acquisition.

that transfer to C2 is minimized, i.e. an rf field strength of 1.9 kHz is chosen with the  $^{13}\text{C}$  carrier at 169 ppm, approximately, the C4 resonance position which is about 19 ppm offset from the C2 position at 150 ppm; the  $^{15}\text{N}$  carrier is set to 160 ppm. The mixing period is 90 ms. Next, the C4 magnetization is transferred by two concatenated refocused INEPT sequences to C6. Here the choice was refocused INEPT instead of TOCSY because of the large difference in chemical shift between C5 and C4, C6. Finally, the C6 coherence is transferred to in-phase H coherence via a reverse INEPT sequence. To correlate the  $\text{NH}_2$  protons with H6 in Cytosine Simorre et al. [150] propose to use cross-polarization for the transfer from  $\text{NH}_2$  protons to  $\text{NH}_2$  nitrogen as well as for the subsequent transfer to C4 (Fig. 15(C) and (D)).

The pulse sequence by Fiala et al. [153] (Fig. 15(E))

aims to correlate exchangeable and non-exchangeable protons in purines. The approach is conceptually similar to that described above, except that the transfer direction is from H8 to the imino proton instead of the reverse and that an INEPT sequence instead of a cross-polarization sequence is used in all hetero  $\text{C} \rightarrow \text{N}$  nuclear transfer steps. For the (C,C) TOCSY from C8 to C6, a mixing time of 55 ms and an rf field strength of 3 kHz were chosen, and the  $^{13}\text{C}$  carrier was placed at 150 ppm. The selective  $180^\circ$  carbon pulse for the INEPT transfer from C6 to N1 was a 1.5 ms REBURP, positioned at 160 ppm to cover the region from 150 to 170 ppm. As a result the  $J_{\text{C6C5}}$ -coupling does not influence the C6 to N1 transfer (see below). The  $^{15}\text{N}$  carrier was positioned at 120 ppm inbetween the N1 and  $\text{NH}_2$  resonances.

The third set of pulse sequences that has been published to correlate exchangeable and non-exchangeable protons is by Sklenar et al. [152]. The 2D and 3D versions of these sequences are shown in Fig. 15(F) and (G), respectively. They are very similar to those proposed by Simorre et al. [150,151]. The main difference is that the transfer here is from H6/H8 to the imino protons instead of the reverse. In addition, the authors use the same pulse sequence and settings for both uridine and guanosines. For both the C–C transfer and the C–N transfer the rf field strength used is 2.9 kHz, and the carriers are set at 150 ppm for  $^{13}\text{C}$  and at 153 ppm for  $^{15}\text{N}$ . The (C,C) TOCSY mixing period is taken to be 19 ms and the (C,N) TOCSY mixing period 58 ms. The authors suggest that with the carrier at 150 ppm the low rf field will eliminate C5 from the  $J_{\text{CC}}$ -coupling network both in G and in U. They state that in G the main route would be C8–C4–C6. However, as follows from the  $J$ -coupling data in Fig. 3, a route directly from C8 to C6 is also possible. In U the elimination of C5 from the network renders the (C,C) TOCSY here ineffective and the coherence remains on C6. The long 70 ms mixing period in the hetero TOCSY, transfers the C6 coherence in U to N1H coherence via the route C6–N1–C2–N3.

To estimate the relative sensitivities of the three groups of experiments we will compare the different sequences in terms of their transfer efficiencies, and discuss in detail the G N1H to H8 transfer. The overall efficiency of the three groups of experiments can, for the G N1H(imino) to H8 transfer, be expressed as

$$Trt = Tr_{\text{INEPT}}(\text{H8} \rightarrow \text{C8}) Tr_{\text{CCTOCSY}}(\text{C8} \rightarrow \text{C6}) Tr_{\text{cpCN}}(\text{C6} \rightarrow \text{N1}) Tr_{\text{INEPT}}(\text{N1} \rightarrow \text{N1H}) \quad (43)$$

(Simorre et al., [150,151])

$$Trt = Tr_{\text{INEPT}}(\text{H8} \rightarrow \text{C8}) Tr_{\text{CCTOCSY}}(\text{C8} \rightarrow \text{C6}) Tr_{\text{INEPT}}(\text{C6} \rightarrow \text{N1}) Tr_{\text{INEPT}}(\text{N1} \rightarrow \text{N1H}) \quad (44)$$

(Fiala et al., [153])

$$Trt = Tr_{\text{INEPT}}(\text{H8} \rightarrow \text{C8}) Tr_{\text{CCTOCSY}}(\text{C8} \rightarrow \text{C6}) Tr_{\text{cpCN}}(\text{C6} \rightarrow \text{N1}) Tr_{\text{INEPT}}(\text{N1} \rightarrow \text{N1H}) \quad (45)$$

(Sklenar et al., [152])

Here,  $Tr_{\text{INEPT}}(\text{H8} \rightarrow \text{C8})$  and  $Tr_{\text{INEPT}}(\text{N1} \rightarrow \text{N1H})$  represent the efficiency of the transfer of in-phase

proton to in-phase X-nucleus coherence. For the calculations we assume that  $Tr_{\text{INEPT}}(\text{H8} \rightarrow \text{C8})$  and  $Tr_{\text{INEPT}}(\text{N1} \rightarrow \text{NH})$  are 100% efficient, which seems reasonable in view of the large  $J$ -couplings involved (200 Hz for  $J_{\text{H8C8}}$  and 90 Hz for  $J_{\text{HN}}$ ). The term  $Tr_{\text{CCTOCSY}}(\text{C8} \rightarrow \text{C6})$  represents the efficiency of the (C,C) TOCSY transfer,  $Tr_{\text{cpCN}}(\text{C6} \rightarrow \text{N1})$  stands for the efficiency of the cross-polarization between C6 and N1, and finally,  $Tr_{\text{INEPT}}(\text{C6} \rightarrow \text{N1})$  represents the efficiency of transfer of in-phase C6 coherence to in-phase N1 coherence via INEPT. Comparison of the transfer functions shows that all sequences have the (C,C) TOCSY transfer step, the H8 to C8 INEPT transfer step, and the N1 to N1H INEPT transfer step in common. The differences are found in the C6 to N1 transfer step. For the sequences proposed by Simorre et al. and Sklenar et al. the transfer functions are identical, both using (C,N) cross-polarization for this transfer step, while Fiala et al. employ INEPT for this transfer step. We will now consider these steps in more detail in view of the more detailed knowledge of the  $J$ -couplings in the bases.

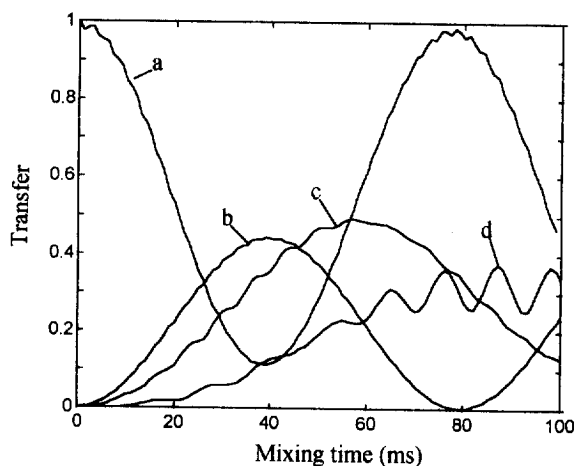


Fig. 16. Transfer of coherence during (C,C) TOCSY in guanosine via numerical simulation using MLEV-17 for three different rf field strengths, demonstrating the influence of chemical shift offsets on the efficiency of the coherence transfer. The spin system consists of C6, C8 and C4, all coupled with 9 Hz  $J$ -couplings; the chemical shift offsets are 1200 Hz for C6, 200 Hz for C4 and –1200 Hz for C8. At mixing time zero all coherence is assumed to be present on C6. (a) The coherence on C6 as a function of the TOCSY mixing time (rf field 10 kHz). (b) The coherence transferred from C6 to C8 and to C4 as a function of the TOCSY mixing time (rf field 10 kHz). (c) As in (b), except rf field 5 kHz. (d) As in (b), except rf field 3 kHz.



Fig. 16 shows the efficiency of C6  $\rightarrow$  C8 coherence via (C,C) TOCSY in Guanosine bases. As can be seen from Fig. 3, in principle all carbons in the base are connected via  $J_{CC}$ -couplings. However, the  $J$ -coupling to C2 is small. Furthermore, the C5 spin resonates at approximately 100 ppm, so that it is excluded from the network (but see discussion in Section 7.2.2.3), when a sufficiently weak rf field is used. The  $J_{CC}$ -coupling network consists then of the C6, C8 and C4 spins, connected with  $J$ -couplings of approximately 9 Hz. The (C,C) TOCSY transfer through the C6, C8, C4 spin network was simulated numerically assuming an MLEV-17 pulse sequence [77,86], for three different rf field strengths (10, 5 and 3 kHz); the offsets of C6, C8 and C4 were taken to be 1200,  $-1200$  and 200 Hz, respectively (estimated from their usual chemical shift values as given in Table 5, and assuming a carrier position of 150 ppm). The C6 to C8 transfer efficiency reaches an optimum of 0.42 at 40 ms (10 kHz rf field), 0.47 at 60 ms (5 kHz), and 0.35 at 90 ms (3 kHz). These values are obtained when  $T_2$  relaxation is assumed to be absent. In Table 7 the optimum values for (C,C) TOCSY transfer are given when  $T_2$  relaxation is taken into account; one finds for  $T_2$  relaxation times

of 3 s and 0.03 s, optimum transfer efficiencies of **0.42** at **40 ms**, **0.11** at **25 ms**, respectively, for a 10 kHz rf field. The simulations also demonstrate that the main transfer route is directly from C6 to C8, and not only via the route C8–C4–C6 as suggested by Sklenar et al. [152]. We note that the offset of the C5 spin, which has been excluded from our calculations, corresponds to  $-4500$  Hz. Obviously, at rf field strengths of 10 kHz and 5 kHz, offset effects arising from C5 are expected to affect the C6 to C8 transfer. On the other hand, for an rf field strength of 3 kHz one expects that the C5 spin can be disregarded. Sklenar et al. and Fiala et al. employed a DIPSI pulse sequence of 3 kHz, while Simorre et al. employed a FLOPSY sequence of 5 kHz. Thus, particularly in the latter case, considerable offset effects are expected involving the C5 spin. To estimate the effect of the presence of C5 in the spin network in an exact manner would require an extensive set of numerical simulations, which is beyond the scope of this review. Such an analysis would be of interest, however, since it might very well lead to improved sequences and/or parameter settings for the (C,C) TOCSY transfer. Qualitatively, it is to be expected that fully excluding the C5 spin from the network should improve the

Table 7  
Transfer efficiencies in HNCCH and HCCNH-type experiments

	Efficiency <sup>a</sup> ( $T_2$ 3 s)	$\tau$ (ms)( $T_2$ 3 s)	Efficiency <sup>f</sup> ( $T_2$ 30 ms)	$\tau^f$ (ms)( $T_2$ 30 ms)
$Tr_{INEPT}$ (C6 $\rightarrow$ N1)	0.20	56	0.034	32
$Tr_{cpCN}$ (C6 $\rightarrow$ N1) <sup>b1</sup>	0.20	56	0.034	32
$Tr_{cpCN}$ (C6 $\rightarrow$ N1) <sup>b2</sup>	0.14	50	0.034	32
$Tr_{CCTOCSY}$ (C8 $\rightarrow$ C6) <sup>c</sup>	0.46	55	0.12	30
$Tr_{CCTOCSY}$ (C8 $\rightarrow$ C6) <sup>d</sup>	0.46	40	0.17	25
$Tr_{CCTOCSY}$ (C8 $\rightarrow$ C6) <sup>e</sup>	0.15	19	0.08	19*
$Trt$ (Fiala et al.)	0.092	90	0.004	62
$Trt$ (Simorre et al.)	0.092	90	0.006	57
$Trt$ (Sklenar et al.)	0.021	69	0.003	51

<sup>a</sup> Transfer efficiencies calculated for the time  $\tau_m$

<sup>b1</sup> The optimum transfer with the settings Simorre et al. (experimental mixing time 44.3 ms) (see text).

<sup>b2</sup> Optimum transfer Sklenar et al. (experimental mixing time 58 ms) (see text).

<sup>c</sup> The transfer was simulated using an MLEV-17 sequence of 5 kHz with the carrier at 150 ppm (see text) (optimum of 0.48 at 60 ms); the efficiency is given at 55 ms which corresponds to the mixing time used by Fiala et al.

<sup>d</sup> The numbers correspond to the efficiency calculated for MLEV-17 sequence using an rf field of 10 kHz with the carrier at 150 ppm (see text) (optimum of 0.46 at 40 ms); the efficiency is given at a mixing time of 40 ms, which roughly corresponds to the mixing time used by Simorre et al. (37.8 ms) (see also text).

<sup>e</sup> Transfer found for the shorter mixing time of 19 ms used by Sklenar et al.; the rf field is assumed to be 5 kHz (optimum of 0.48 at 60 ms) (see text).

<sup>f</sup> Efficiency at the optimum value for the mixing time, which is the value given in the next column, except for those indicated with a \*.

transfer efficiency. At first sight one would expect that this could be achieved by just lowering the rf field strength, but as can be seen from Fig. 16, at 3 kHz the transfer becomes negatively affected by the off-sets of C6 and C8. Thus, an improved sequence should consist of a shaped (C,C) TOCSY sequence, which only covers the region of C6, C4, C8 and C2.

The INEPT transfer from C6 to N1 (see the sequence proposed by Fiala et al. [153]) is given by

$$T_{C6x, C6yN1z} = \sin(\pi J_{N1C6} 2\tau_2) \cos(\pi J_{C6C8} 2\tau_2) \cos(\pi J_{C6C5} 2\tau_2) \cos(\pi J_{C6C4} 2\tau_2) \quad (46)$$

$$T_{N1zC6x, N1y} = \sin(\pi J_{N1C6} 2\tau_2) \cos(\pi J_{N1C2} 2\tau_2)$$

$$Tr_{INEPT} = T_{C6x, C6yN1z} \cdot T_{N1zC6x, N1y}$$

Fiala et al. employ a selective  $^{13}\text{C}$  pulse, which covers the region of 150 to 170 ppm, i.e. the spectral region where C6, C2 and C4 resonate. Consequently, the terms involving  $J_{C6C5}$ ,  $\cos(\pi J_{C6C5} 2\tau_2)$ , and  $J_{C6C8}$ ,  $\cos(\pi J_{C6C8} 2\tau_2)$ , equal 1. The transfer efficiencies,  $Tr_{INEPT}$ , in G are shown in Fig. 17, assuming  $T_2$  values of 3 s and 30 ms, respectively. An optimum of about 20% is reached at  $\tau_2 = 14$  ms with a selective pulse on  $^{13}\text{C}$  and  $T_2 = 3$  s; this number drops to 3.4% at 8 ms when  $T_2 = 30$  ms (see also Table 7); without the selective  $^{13}\text{C}$  pulse, i.e. with  $J_{C6C8}$  active, these are

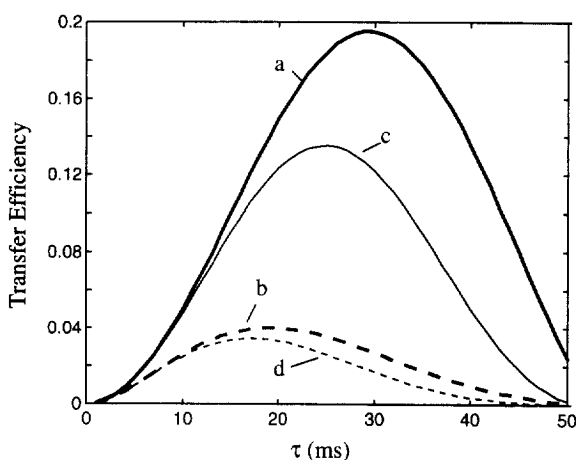


Fig. 17. Efficiency of  $N1x \rightarrow C6x$  coherence transfer via INEPT in HCCNH experiment (see text); (a)  $T_2 = 3$  s,  $J_{C6C8}$  inactive; (b)  $T_2 = 30$  ms,  $J_{C6C8}$  inactive; (c)  $T_2 = 3$  s,  $J_{C6C8}$  active; (d)  $T_2 = 30$  ms,  $J_{C6C8}$  active.  $J_{CC}$ -couplings are as given in Fig. 3.

14% at  $\tau_2 = 12.5$  ms, and 3.4% at 8 ms, respectively (see also Table 7). We note that leaving the passive  $J_{C6C5}$ -coupling present would not have a significantly negative effect on the efficiency.

In the pulse sequences of Simorre et al. and Sklenar et al., a DIPSI cross-polarization was used to achieve the  $C \rightarrow N$  coherence transfer. Simorre et al. use a weak field of 1.9 kHz and place the  $^{13}\text{C}$  and  $^{15}\text{N}$  carriers exactly on C6 (161 ppm) and on N1H (146 ppm), respectively. They use a mixing time of 44 ms. With these settings the C8 spin may be assumed to be absent from the spin system effective during the  $C$  to  $N$  transfer, i.e. it consists of C6, C2, N1 and C4. The efficiency may then be estimated to be similar to that of the INEPT transfer with a selective  $^{13}\text{C}$  180° positioned at C6 as used by Fiala et al., i.e. optimum transfer of 20% at 56 ms ( $T_2 = 3$  s). Sklenar et al. used a field of 2.9 kHz, with the carrier at 150 and 153 ppm for  $^{13}\text{C}$  and  $^{15}\text{N}$ , respectively; the mixing time for  $C \rightarrow N$  transfer was optimized for guanosines to 58 ms. With these settings C8 may not be excluded from the spin network, and the cross-polarization transfer then corresponds to an INEPT transfer where C8 is included in the spin network, i.e. optimum transfer of 14% at 50 ms ( $T_2 = 3$  s). Since the transfer based on cross-polarization is analogous to INEPT, we used the values derived from INEPT in Table 7.

The transfer efficiencies are summarized in Table 7. As can be seen, no essential differences exist for the three sequences when  $T_2$  is 3 s, although the sequences of Simorre et al. and Fiala et al. are slightly more efficient, because the C8 spin is excluded from the network on  $C$  to  $N$  transfer. For larger systems ( $T_2 = 30$  ms) the efficiency drops dramatically. Interestingly, when  $T_2$  is 30 ms, the sequence of Simorre et al. outperforms the other sequences. This is a consequence of the faster carbon to carbon transfer, which in turn results from the stronger rf field strength used by Simorre et al.

**7.2.2.3. Through-bond H2–H8 correlation.** To obtain through-bond H2–H8 correlations in  $^{13}\text{C}/^{15}\text{N}$  labeled RNAs (Fig. 13(C), step III), two pulse sequences have been published; one by Marino et al. [154] (Fig. 18(A)), and one by Legault et al. [155] (Fig. 18(B)). The two pulse sequences are quite similar and are closely related to the HCCH–TOCSY experiment originally

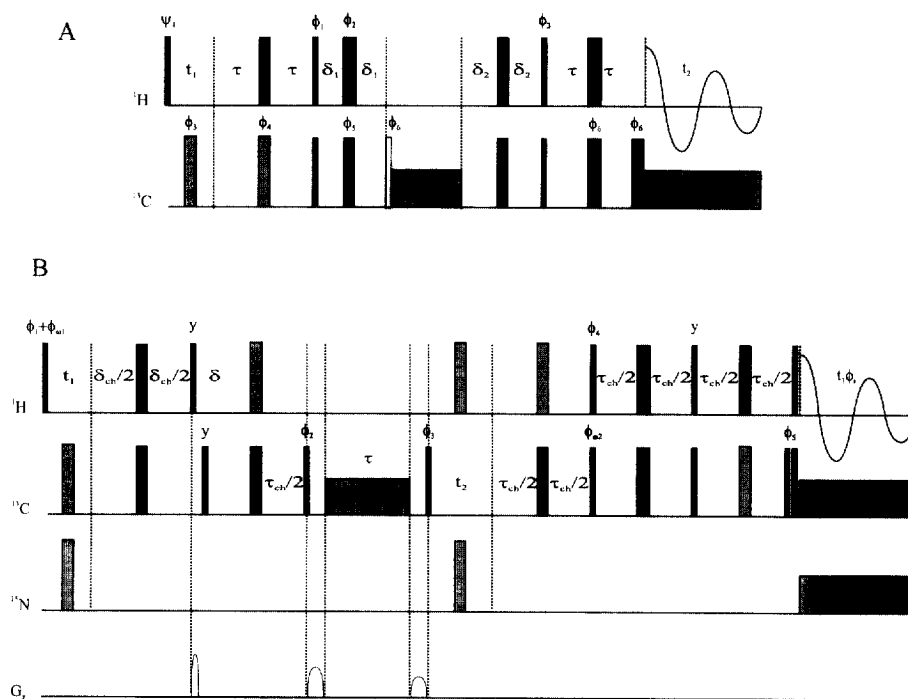


Fig. 18. Pulse schemes for HCCH-TOCSY experiments to determine adenine H2 to H8 correlations in isotopically labeled RNAs. The narrow black rectangles indicate  $90^\circ$  pulses, the wider black rectangles  $180^\circ$  pulses, the grey filled wide rectangles indicate  $180^\circ$  composite pulses (see below), and the narrow unfilled rectangle indicates a trim pulse. All pulses are applied along the x axis, unless indicated otherwise. (A) The 2D HCCH-TOCSY [157,158] experiment [154]. The 16-step phase cycle is as follows:  $\phi_1 = y, -y$ ;  $\phi_2 = 4(x), 4(y), 4(-x), 4(-y)$ ;  $\phi_3 = 8(x), 8(-x)$ ;  $\phi_4 = 2(x), 2(-x)$ ;  $\phi_5 = 2(x), 2(y), 2(-x), 2(-y)$ ;  $\phi_6 = 4(x), 4(-x)$ ; Acq. =  $2(x, -x, -x, x), 2(-x, x, x, -x)$ . The  $180^\circ \phi_3$  and  $180^\circ \phi_4$  are of the composite type ( $90x180y90x$ ) and a 1.7 ms spin lock trim pulse is applied with phase  $\phi_6$ . The  $^{13}\text{C}$  carrier is positioned in the center of the aromatic region (150 ppm), and a 100 ms spin lock period using DIPSI-2 with an rf field of 3.5 kHz (covering the range of C2, C4, C6, C8) is applied along the y axis (C5, which is 30 ppm upfield of the carrier, is not excited). The  $\tau$ ,  $\delta_1$  and  $\delta_2$  delays were set to  $1/4J_{\text{CH}}$ , 1.25 ms where  $J_{\text{CH}} \sim 200 \text{ Hz}$  for aromatic resonances. Quadrature detection is obtained with the TPPI-States method [245] using  $\psi_1 = 16(x), 16(y)$ . The receiver reference phase and  $\psi_1$  are incremented by  $180^\circ$  for each  $t_1$  increment. GARP decoupling [247] of  $^{13}\text{C}$  and  $^{15}\text{N}$  was applied during the acquisition, as well as during  $t_1$  for  $^{15}\text{N}$ . (B) The 3D ( $^1\text{H}, ^{13}\text{C}, ^{15}\text{N}$ ) HCCH-TOCSY experiment [155]. The phase cycle is:  $\phi_1 = x, -x$ ;  $\phi_2 = 2(x), 2(-x)$ ;  $\phi_3 = 4(y), 4(-y)$ ;  $\phi_4 = 8(x), 8(-x)$ ;  $\phi_5 = 16(x), 16(-x)$ ; and  $\phi_6 = \phi_1 + \phi_2 + \phi_3 + \phi_4 - \pi/2$ . The composite  $180^\circ$  pulses are of the type ( $90x240y90x$ ). Complex data were collected in  $t_1$  by States-TPPI with FIDs for  $\phi_{\omega_1} = (x, y)$  [245] being stored separately and in  $t_2$  with FIDs for  $\phi_{\omega_2} = (x, -x)$  [250] being stored separately. The 3D hypercomplex data set was processed to achieve a quadrature detected, phase sensitive display along the  $^{13}\text{C} t_2$  dimension with sensitivity enhancement [250]. A FLOPSY-8  $^{13}\text{C}$ - $^{13}\text{C}$  spin lock [238] was applied for 65.4 ms ( $\tau_m$ ) using a 2.53 kHz rf field strength. For optimum efficiency in  $^{13}\text{C}$  isotropic mixing, the  $^{13}\text{C}$  carrier was placed midway between the adenine C2 and C8 resonances. All gradients were square shaped and applied along the z axis: G1 = 8.0 G/cm for 0.5 ms, G2 = 6.5 G/cm for 1.0 ms and G3 = 4.7 G/cm for 0.7 ms. G1, G2 and G3 were followed by 1.5, 2.0 and 0.7 ms recovery times, respectively. Other relevant acquisition parameters are:  $\text{SW}_{\omega_1}(^1\text{H}) = 1 \text{ kHz}$ ,  $t_1 \text{ max} = 27.00 \text{ ms}$ ,  $\text{SW}_{\omega_2}(^{13}\text{C}) = 4.2 \text{ kHz}$ ,  $t_2 \text{ max} = 7.38 \text{ ms}$ ,  $\text{SW}_{\omega_3}(^1\text{H}) = 4.45 \text{ kHz}$ ,  $t_3 = 57.53 \text{ ms}$ ,  $\delta_{\text{ch}} = 2.0 \text{ ms}$ ,  $\tau_{\text{ch}} = 2.4 \text{ ms}$ ,  $\gamma B_2(^{13}\text{C} \text{ decouple}) = 1.79 \text{ kHz}$ , and  $\gamma B_3(^{15}\text{N} \text{ decouple}) = 1.23 \text{ kHz}$  with GARP-1 [247].

proposed for proteins [156–158]. In the version of Legault et al., improved signal-to-noise was obtained by replacing the final INEPT step by a sensitivity-enhanced version. Nowadays, versions of the 2D HCCH-TOCSY and 3D HCCH-TOCSY exist which have further improved sensitivity; as compared to the standard sequence these sequences exhibit an

improvement in sensitivity by a factor of 1.4 for the 2D versions and close to a factor of 2 for the 3D version (see also discussion in Section 7.2.2.5, and Refs. [158–166]). Thus, it would be advisable, for optimal sensitivity, to use these enhanced sequences instead of the ones illustrated in Fig. 18.

In both the published pulse sequences (Fig. 18) H2

coherence is transferred to in-phase C2 coherence via a refocused INEPT sequence. Subsequently, the in-phase C2 coherence is transferred further to in-phase C8 coherence, via the carbon–carbon  $J$ -coupling network, using a (C,C) DIPSI-2 or FLOPSY-8 sequence. In the last transfer step the in-phase C8 coherence is transferred into in-phase H8 coherence via a refocused INEPT sequence (Marino et al., Fig. 18(A)) or via a sensitivity-enhanced refocused INEPT sequence (Legault et al., Fig. 18(B)). The main differences between the two H2–H8 correlation experiments lie in the type of isotropic mixing sequence used, the settings of the rf field strength, carrier position, and the length of the TOCSY mixing time. The question is then, for which settings is optimal C  $\rightarrow$  C transfer achieved? A similar problem is encountered in the HNCCH and HCCNH experiments, where the coherence transfer is through the complex carbon network of the guanosine base (see Section 7). The carbon spins in the Adenosine base also form quite a complex network, but with somewhat different characteristics compared to the guanosine base (see Fig. 3). As can be seen from Fig. 3, C6, C4 and C8 have  $J$ -couplings of approximately 10 Hz; C5 has in turn a large  $J$ -coupling to C6 and C4 (about 80 Hz), while C2 has a  $J$ -coupling of 12 Hz to C5. C2 may be further  $J$ -coupled to C6 or C4, but these  $J$ -couplings must be smaller than 3 Hz [49]. Furthermore, C6, C4, C8 and C2 reside in a relatively narrow spectral region (162–135 ppm), while C5 resides around 120 ppm (Table 6). The problem is thus to obtain optimal transfer through this system of carbon spins from C2 to C8. Marino et al. position the  $^{13}\text{C}$  carrier at the center of the aromatic region at 150 ppm, apply a DIPSI-2 mixing sequence of 100 ms duration with an rf field strength of 3.5 kHz; Pardi et al. use a FLOPSY-8 sequence of 64.5 ms duration and rf field strength of 2.53 kHz with the carrier midway between C2 and C8 (at 145 ppm). With these settings the C2, C4, C6 and C8 are covered by both the 3.5 kHz and 2.53 kHz rf fields. The C5 spins, at 120 ppm, have offsets of 4500 Hz and 3750 Hz in the settings of Marino et al. and Legault et al., respectively. These offsets are larger, but not very much larger, than the rf field strengths. Consequently, the C5 spin cannot be neglected, leading to a coherence transfer, which is governed by a complex interplay of offset effects with field strength and

type of mixing sequence. Numerical simulations are required to analyze and determine the ultimate outcome of the transfer. Such an in-depth analysis of the coherence transfer would be of interest, since it might very well turn out that the settings used are not optimal.

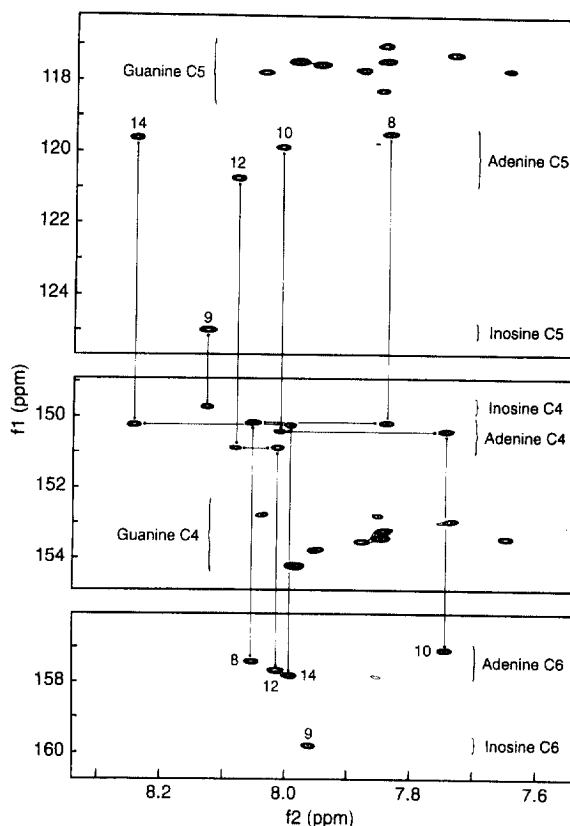


Fig. 19. Sections of the HMBC spectrum of a 1.5 mM sample of 5'-d(CGGCCG-GAAGAGA-CGGCCG)-3' in 99.98%  $\text{D}_2\text{O}$ , pH 6.8, at 25°C [144]. The spectrum was acquired in about 80 h on a Varian UNITY plus spectrometer operating at 750 MHz, although it should be noted that the obtained signal-to-noise ratio allows the duration of the experiment to be reduced by a factor of two or three. The proton carrier was placed at the position of the water resonance and the spectral width in the F2 dimension was set to 7500 Hz. For the indirect dimension the spectral width was set to 1163 Hz, centered at 139 ppm. States-TPPI was used for phase-sensitive detection of the indirectly observed frequency [245]. The spectrum was recorded with 2048 increments in  $t_2$  and 1024 increments in  $t_1$ . Suppression of the HDO signal was achieved by presaturation during the relaxation delay. During detection,  $^{13}\text{C}$  GARP decoupling was executed. The upper panel shows the C5–H8 correlations for adenine and guanine residues. In the middle panel the C4–H8 and C4–H2 cross peaks, used to correlate the adenine H8 and H2 resonances, are given. The lower panel shows the C6–H2 correlations for adenine and inosine residues.

An alternative means of obtaining H2 to H8 correlations is to use an HMBC experiment to obtain long-range H to C correlations. As can be gleaned from Fig. 3, H2 has sizable  $^3J_{\text{HC}}$ -couplings to C4 and C6, and H8 to C5 and C4, so that in an HMBC, H2 to C4, C5 and C6 cross peaks can be observed, together with H8 to C5 and C4 cross peaks. The common  $J$ -coupling partner, C4, provides a means of obtaining through-bond H2 to H8 correlations. As far as we are aware, this type of experiment has not been used in labeled RNA, but has been used in DNA (unlabeled) to obtain H2 to H8 correlations [144]. Fig. 19 shows part of the HMBC spectrum of a DNA hairpin, which contains a number of G.A base pairs, demonstrating the utility of the experiment for obtaining H2 to H8 correlations. The experiment has the added advantage that it also provides other long-range correlations. For example, one finds for the Guanosine bases H8 to C5 and C4 correlations. As can be seen in Fig. 19, the G C5 resonances have a spectral position, which differs from that of A G5, and the H8 to C5 cross peaks can be used to distinguish Guanosine and Adenosine bases.

**7.2.2.4. Through-bond base–sugar correlation.** The NMR experiments that have been proposed for correlating base and sugar protons (step III, Fig. 13(C)) are usually given a name that indicates the magnetization transfer route; for example, HCN indicates a triple resonance experiment, where H, C and N are correlated. With these naming conventions, the NMR experiments proposed for through-bond base–sugar correlation can be put under the headings of HCN experiments [167,168], HCNCH experiments [169,170], HbNb(Hb)CbHb experiments where b stands for base [168], and HCNH experiments [171]. Fig. 20(A), (B) and (C) show the pulse sequences proposed by Sklenar et al. [167,169], Fig. 20(D) the sequence proposed by Tate et al. [171], and Fig. 20(E) the sequence proposed by Farmer et al. [168,170]. In the HCN experiment either H1' is correlated with C1' and N9/N1 or H6/H8 is correlated with C8/C6 and N9/N1 [167,168]. In the HCNCH type of experiment the HCN sequence is extended into a relay experiment. The original proton magnetization, say H1', is transferred to N9/N1 as in the HCN part, and then relayed from N9/N1 via C6/C8 to H6/H8

[168–170]. Farmer et al. [168] also proposed for purines an experiment, HbNb(Hb)CbHb, where the H8 magnetization is transferred directly to N9 and back to H8, utilizing the  $J_{\text{H8N9}}$ -coupling. In order to turn it into a 3D experiment the magnetization is transferred to C8 and back before acquisition. Tate et al. [171] proposed for purine residues an experiment which also utilizes the direct H8 to N9  $J$ -coupling. In their proposed HCNH experiment the H1' magnetization is transferred via C1' to N9 as in the first part of the HCN or HCNCH experiments, and then directly via the  $J_{\text{H8N9}}$ -coupling to H8. Sklenar et al. [172] have proposed that the correlation of sugar and base protons can also be obtained via  $^{15}\text{N}$  HSQC spectrum.

It is interesting to compare the various approaches and consider their sensitivities in view of the detailed knowledge that now exists of the complex network of  $J$ -couplings in the bases (see Section 5). To gain a correct insight into the relative performance of the pulse sequences the magnetization transfer functions are required. The transfer functions governing the HCNCH pulse sequence are given by

Pyrimidine:

$$Tr_{\text{H1'C1'}} = \sin(\pi J_{\text{H1'C1'}} 2\tau_1) \quad (47)$$

$$Tr_{\text{C1'N9}} = \sin(\pi J_{\text{C1'N9}} 2\tau_2) \cos(\pi J_{\text{C1'C2'}} 2\tau_2) \\ \sin(\pi J_{\text{C1'H1'}} 2\tau_1)$$

$$Tr_{\text{N9C6}} = \sin(\pi J_{\text{C1'N9}} 2\tau_3) \sin(\pi J_{\text{N9C6}} 2\tau_3) \cos(\pi J_{\text{N9C2}} 2\tau_3)$$

$$Tr_{\text{C6H6}} = \sin(\pi J_{\text{C1'N9}} 2\tau_4) \sin(\pi J_{\text{C6H6}} 2\tau_5) \cos(\pi J_{\text{C6C5}} 2\tau_4) \\ \cos(\pi J_{\text{C6C2}} 2\tau_4)$$

$$Tr_{\text{H6C6}} = \sin(\pi J_{\text{H6C6}} 2\tau_5)$$

Purine:

$$Tr_{\text{H1'C1'}} = \sin(\pi J_{\text{H1'C1'}} 2\tau_1) \quad (48)$$

$$Tr_{\text{C1'N9}} = \sin(\pi J_{\text{C1'N9}} 2\tau_2) \cos(\pi J_{\text{C1'C2'}} 2\tau_2) \\ \sin(\pi J_{\text{C1'H1'}} 2\tau_1)$$

$$Tr_{\text{N9C8}} = \sin(\pi J_{\text{C1'N9}} 2\tau_3) \sin(\pi J_{\text{N9C8}} 2\tau_3) \cos(\pi J_{\text{N9C4}} 2\tau_3) \\ \cos(\pi J_{\text{N9C5}} 2\tau_3)$$

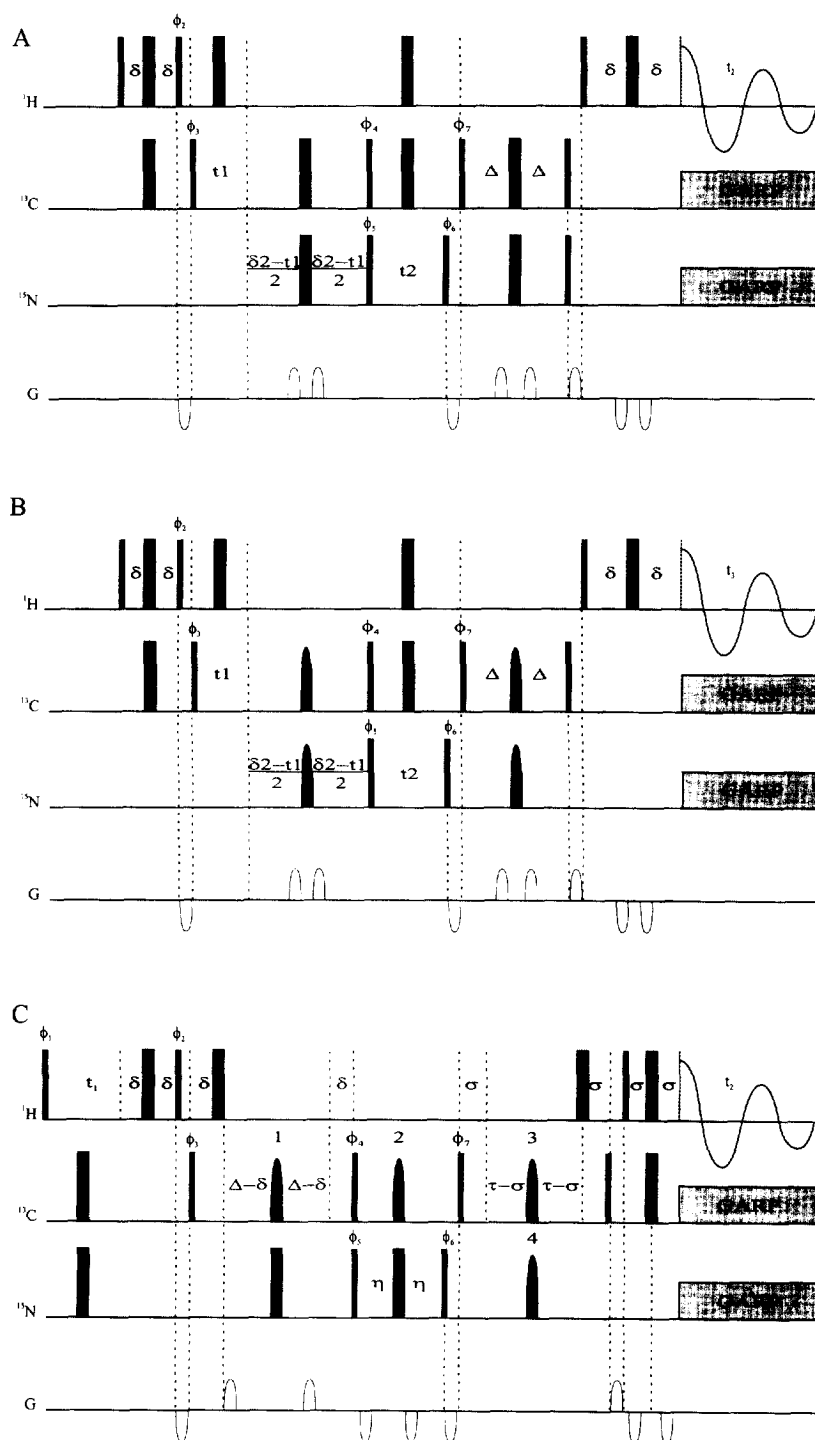


Fig. 20.

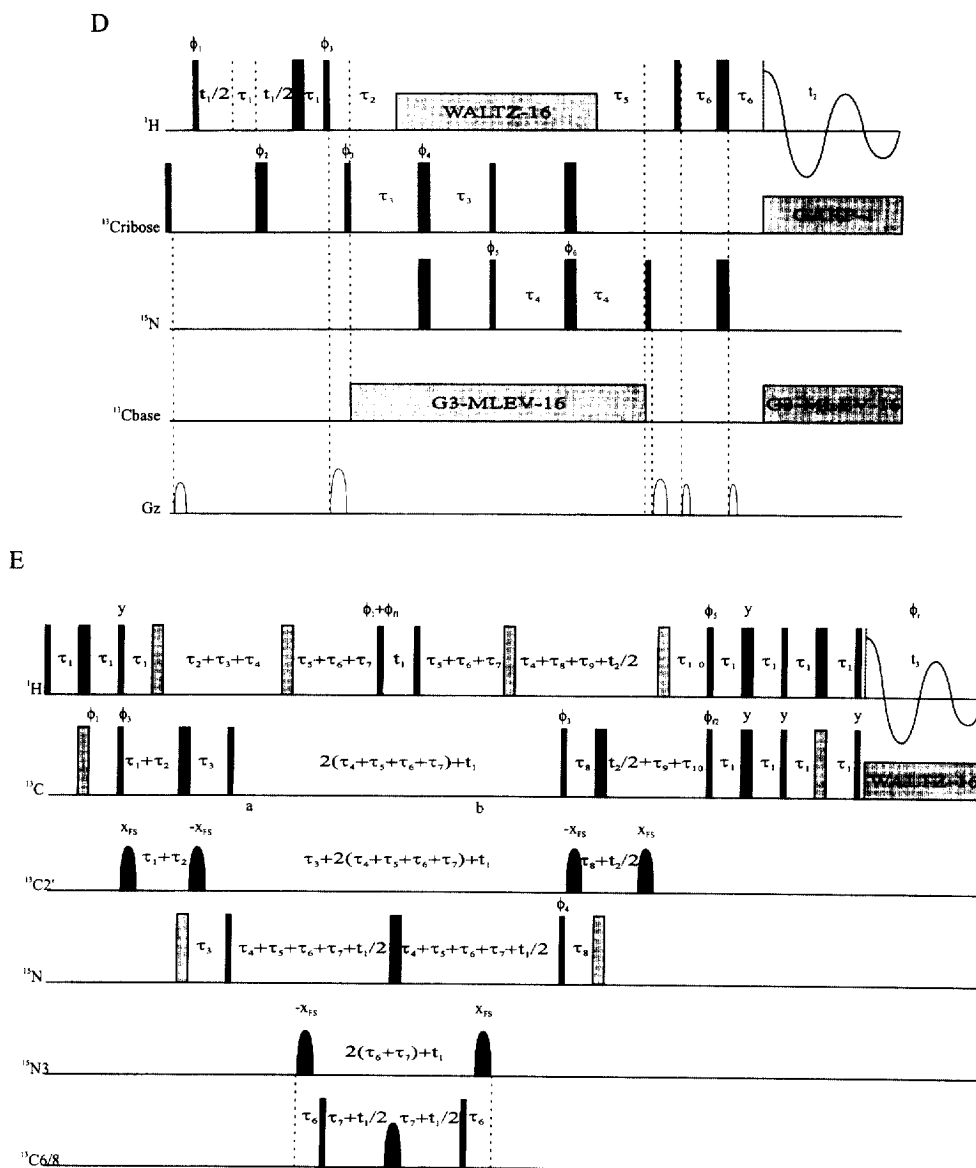


Fig. 20. Pulse schemes for experiments that correlate the base to the sugar. (A) Non-selective version of the HCN experiment [167] to detect  $H1'-C1'-N9/N1$  connectivities; (B) semi-selective version of the HCN experiment [167] to detect the  $H8-C8-N9$  and  $H6-C6-N1$  connectivities; (C) 2D triple resonance HCNCH experiment to determine through-bond  $H1'-H6/H8$  connectivities via the route  $H1'-C1'-N9/N1-C6/C8-H6/H8$  [169]; (D) 2D triple resonance HCNH experiment to determine through-bond  $H1'$  to  $H8$  connectivities via the route  $H1'-C1'-N9-H8$  [171]; and (E) 3D triple resonance HCNCH experiment proposed by Farmer et al. [170]. Narrow black rectangles indicate  $90^\circ$  pulses, wide black rectangles  $180^\circ$  pulses, grey filled rectangles composite  $180^\circ$  pulses of the type  $(90, 240, 90, \dots)$ , black filled rounded rectangles semi-selective  $180^\circ$  pulses (B, C and E), and unfilled rounded rectangles pulsed field gradients along the  $z$  axis. All pulses are applied along the  $x$  axis unless otherwise indicated. For (A),  $\delta = 1/4 J_{H1'C1'} = 1.6$  ms,  $\delta_2 = 42$  ms,  $\Delta = 21$  ms. The  $^{13}C$  and  $^{15}N$  carriers and sweep widths are set to cover the range of  $C1'$  (88–96 ppm) and  $N9/N1$  (142–176 ppm) chemical shifts. For (B),  $\delta \sim 1/4 J_{C8H8} \sim 1/4 J_{C6H6} \sim 1.25$  ms,  $\delta_2 = 38$ –40 ms,  $\Delta = 19$ –20 ms,  $180\text{ sel}1 = 4$  ms, REBURP refocusing pulse [241] centered at 140.5 ppm,  $180\text{ sel}2 = 2$  ms, IBURP2 inversion pulse [241] centered at 160 ppm. For both experiments, phase cycling is:  $\phi_2 = y, -y$ ;  $\phi_3 = 4(x), 4(-x)$ ;  $\phi_4 = 8(y), 8(-y)$ ;  $\phi_5 = x, x, -x, -x$ ;  $\phi_6 = 32(x), 32(-x)$ ;  $\phi_7 = 16(-y), 16(y)$ ; receiver  $a, -a, -a, a$  where  $a = [x, -x, -x, x, 2(-x, x, x, -x), x, -x, -x, x]$ . In addition,  $\phi_3$  and  $\phi_5$  phases are cycled to obtain the States-TPPI  $t_1$  and  $t_2$  quadrature detection [245]. The 1 ms gradient pulses (800  $\mu$ s gradient pulse shaped to a 1%

truncated sine envelope and 200  $\mu$ s for the magnetic field recovery) are used with amplitudes  $-6/3/3/-6/3/3/12/-3/-3$  G/cm. The asynchronous GARP decoupling [247] is used to suppress the heteronuclear spin–spin interactions during  $t_3$  acquisition. For (C), the delay intervals are set to  $\delta = 1.6$  ms,  $\sigma = 1.25$  ms,  $\Delta = 19.6$  ms,  $\eta = 18$  ms,  $\tau = 19$  ms. The 180° semi-selective pulses (indicated with black round bars) as used at 500 MHz are (1)  $^{13}\text{C}$  4 ms REBURP refocusing pulse [241] covering the range of C1' (88–96 ppm), (2)  $^{13}\text{C}$  4 ms IBURP2 inversion pulse [241] with additional cosine modulation [251,252] for simultaneous inversion of C1' (88–96 ppm) and C6/C8 (136–144.5 ppm), (3)  $^{13}\text{C}$  4 ms REBURP refocusing pulse [241] covering the range of C6 and C8 (136–144.5 ppm), and (4)  $^{15}\text{N}$  2 ms IBURP2 pulse [241] centered at 160 ppm covering the range of N1 and N9 (142–176 ppm). Phase cycling:  $\phi_1 = x$ ;  $\phi_2 = y$ ,  $-y$ ;  $\phi_3 = x, x, -x, -x$ ;  $\phi_4 = 16(x), 16(-x)$ ;  $\phi_5 = 4(x), 4(-x)$ ;  $\phi_6 = 8(x), 8(-x)$ ;  $\phi_7 = 16(x), 16(-x)$ ; receiver =  $a, -a, -a, a$ , with  $a = x, -x, -x, x$ . In addition,  $\phi_1$  is phase cycled to obtain States–TPPI [245]  $t_1$  quadrature detection. The 1 ms gradient pulses (800  $\mu$ s gradient pulse shaped to a 1% truncated sine envelope and 200  $\mu$ s for the magnetic field recovery) are used with amplitudes  $-6/3/3/-6/3/3/12/-3/-3$  G/cm. The asynchronous GARP decoupling [247] is used to suppress the heteronuclear spin–spin interactions during  $t_2$  acquisition. For (D), the delay intervals are set to  $\tau_1 = 1.3$  ms,  $\tau_2 = 2.6$  ms,  $\tau_3 = 12.8$  ms,  $\tau_4 = 16.2$  ms,  $\tau_5 = 16.0$  ms and  $\tau_6 = 8.0$  ms. All carbon pulses are generated using a simple synthesizer without frequency switching. The coherence time delays,  $\tau_5$  and  $\tau_6$ , were experimentally optimized to compensate for the sensitivity loss by the relaxation effect. A delay time as short as 16 ms was found to be a good value, although it was substantially shorter than the theoretical optimal  $\tau_5$  of 42 ms. The frequency offset for carbon pulses is at the center of the C1' carbons, 84 ppm. The rf field strength for all carbon pulses is 19.2 kHz. For off-resonance selective decoupling a G3–MLEV16 expansion is used [253], where each G3 inversion pulse [254] is phase modulated to shift its inversion center to  $+7.8$  kHz [248,255] which is around the center of the base carbon, 146 ppm, excluding the C5 carbon. This G3–MLEV16 selective decoupling is achieved at a field of 2.9 kHz, and under these conditions the selective decoupling has a bandwidth of  $\pm 1.5$  kHz and has little perturbation on the deoxyribose ring carbons. The  $^1\text{H}$  pulses are at a field strength of 29.5 kHz, with an offset on the water resonance. For  $^1\text{H}$  decoupling WALTZ16 [256] is used with a 2.7 kHz field strength. All nitrogen pulses are applied at a field strength of 6.6 kHz, with an offset at the midpoint between the N9 and N7 nitrogen resonances, at 180 ppm. During acquisition, GARP [247] is applied from the nitrogen channel at a field strength of 0.86 kHz and, simultaneously, the same  $^{13}\text{C}$  decoupling as described above is applied from the carbon channel. The durations and strengths of the gradients are G1 = (1.0 ms, 8.0 G/cm), G2 = (4.0 ms, 28.2 G/cm), G3 = (3.0 ms, 18.3 G/cm) and G4 = G5 = (1.0 ms, 5.0 G/cm). A delay of at least 150  $\mu$ s is inserted between the gradient pulse and the subsequent application of an rf pulse to avoid the eddy current effects. All gradients are applied along the  $z$  axis and are rectangular. The phase cycle is  $\phi_1 = x$ ;  $\phi_2 = y$ ,  $-y$ ;  $\phi_3 = x$ ;  $\phi_4 = 2(x), 2(y), 2(-x), 2(-y)$ ;  $\phi_5 = 8(x), 8(-x)$ ;  $\phi_6 = x$ ; receiver =  $x, 2(-x), 2(x), 2(-x), x, -x, 2(x), 2(-x), 2(x), -x$ . The quadrature detection in  $t_1$  is accomplished by States–TPPI [245] of  $\phi_1$ . For (E), the phase cycle is  $\phi_1 = x, -x$ ;  $\phi_2 = 2(x), 2(-x)$ ;  $\phi_3 = 8(x), 8(-x)$ ;  $\phi_4 = 4(x), 4(-x)$ ;  $\phi_5 = 16(x), 16(-x)$ ; and  $\phi_r = \phi_1 + \phi_2 + \phi_3 + \phi_4 + \phi_5$ . Complex data were collected in  $t_1$  by States–TPPI [245] with FIDs for  $\tau_{11} = x, y$  stored separately. Phase modulated data were collected in  $t_2$  with FIDs for  $\tau_{12} = x, -x$  corresponding to N- and P-type signals, respectively [250], being stored separately. The 3D hypercomplex data set was processed as described by Palmer et al. [250] to achieve quadrature detected phase-sensitive display along the C1' dimension with sensitivity enhancement. The  $^{13}\text{C}$  carrier was placed at the center of the C1' resonances, except during the time between points a and b when the  $^{13}\text{C}$  carrier was placed at the C6/C8 resonances. All pulses on lines labeled with a particular spin group, i.e.  $^{13}\text{C}2'$  or  $^{15}\text{N}3$  or  $^{13}\text{C}6/8$ , were Gaussian shaped (64 steps, 5 $\sigma$  cutoff) and are selective for that spin group. The subscript FS appended to certain pulse phases denotes a frequency shifted pulse achieved by phase modulation [246]. Typical settings at 600 MHz are  $\tau_1 = 1.48$  ms,  $\tau_2 = 8.52$  ms,  $\tau_3 = 20.0$  ms,  $\tau_4 = 16.40$  ms,  $\tau_5 = t_1/4$ ,  $\tau_7 = 2.40$  ms,  $\tau_6 = \tau_4 - \tau_5 - \tau_7$ ,  $\tau_8 = \tau_3 - (t_2/2)$ ,  $\tau_9 = \tau_2 - (t_2/2)$ , and  $\tau_{10} = \tau_1 + (t_2/2)$ ;  $t_{90}(^1\text{H}) = 7.5$   $\mu$ s,  $t_{90}(^{13}\text{C}) = 13.4$   $\mu$ s,  $t_{180}(^{13}\text{C}2') = 896$   $\mu$ s,  $t_{90}(^{13}\text{C}6) = t_{180}(^{13}\text{C}6) = 704$   $\mu$ s,  $\gamma B_2 = (\text{C1}' \text{ decouple}) = 1.29$  kHz with WALTZ16 [256],  $t_{90}(^{15}\text{N}) = 37$   $\mu$ s,  $\text{sw}(\text{H}6) = 650.0$  Hz,  $t_1 \text{max}(\text{H}6) = 35.58$  ms,  $\text{sw}(\text{C1}') = 600.0$  Hz,  $t_2 \text{max} = 25.00$  ms,  $\text{sw}(^1\text{H}) = 3$  kHz,  $t_3 \text{max}(\text{H}1') = 120.0$  ms.

$$Tr_{\text{C8H8}} = \sin(\pi J_{\text{N9C8}} 2\tau_4) \cos(\pi J_{\text{N7C8}} 2\tau_4) \cos(\pi J_{\text{C8C6}} 2\tau_4) \\ \cos(\pi J_{\text{C8C4}} 2\tau_4) \sin(\pi J_{\text{C8H8}} 2\tau_5)$$

$$Tr_{\text{N9H8}} = \sin(\pi J_{\text{C1}'\text{N9}} 2\tau_2) \sin(\pi J_{\text{N9H8}} 2\tau_3) \\ \times \cos(\pi J_{\text{N9C4}} 2\tau_2) \cos(\pi J_{\text{N9C8}} 2\tau_2)$$

$$Tr_{\text{H8C8}} = \sin(\pi J_{\text{H6C6}} 2\tau_5)$$

$$Tr_{\text{N9H8}} = \sin(\pi J_{\text{N9H8}} 2\tau_4) \cos(\pi J_{\text{N7H8}} 2\tau_4)$$

The transfer functions of the HCNH experiment for Purines, proposed by Tate et al., are given by

$$Tr_{\text{H1}'\text{C1}'} = \sin(\pi J_{\text{H1}'\text{C1}'} 2\tau_1) \quad (49)$$

$$Tr_{\text{C1}'\text{N9}} = \sin(\pi J_{\text{C1}'\text{N9}} 2\tau_2) \cos(\pi J_{\text{C1}'\text{C2}'} 2\tau_2) \\ \times \sin(\pi J_{\text{C1}'\text{H1}'} 2\tau_1)$$

The transfer functions of the experiment proposed by Farmer et al. [168], the HbNb(Hb)CbHb experiment for purine, where Hb is H8, Nb is N9, and Cb is C8, are given by

$$Tr_{\text{N9H8},v} = (\sin(\pi J_{\text{H9H8}} 2\tau_1) \cos(\pi J_{\text{N7H8}} 2\tau_1))^2 \quad (50)$$

$$Tr_{\text{H8C8},v} = (\sin(\pi J_{\text{H8C8}} 2\tau_2))^2$$

Here,  $v$  denotes that the transfer efficiency given is for the out- and back route.



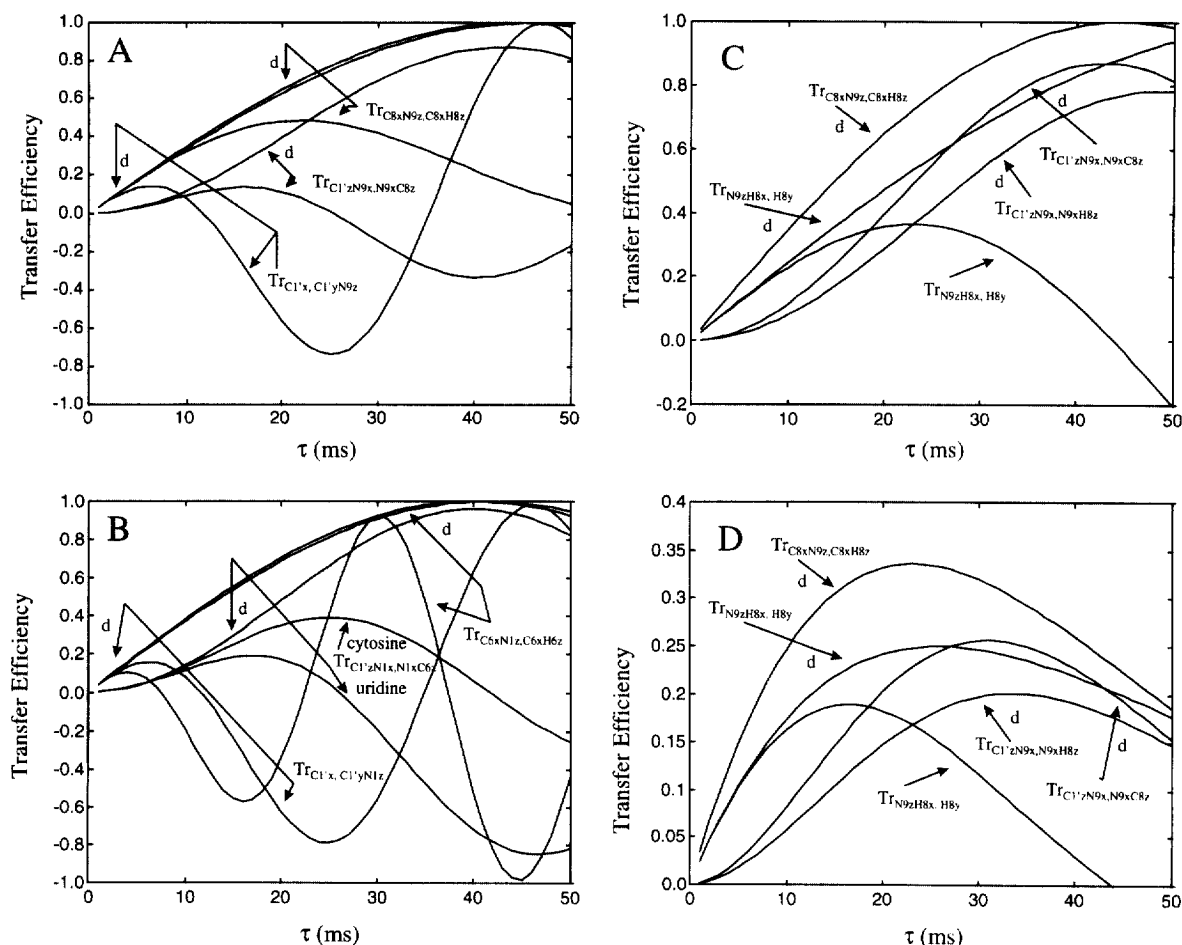


Fig. 21. Transfer efficiency in sugar-to-base correlation experiments. The coherence transfers are given as a function of the transfer time  $\tau$  (ms) (see text). The transfer steps considered are indicated by their name, e.g.  $Tr_{C8\alpha N9z,C8\alpha H8z}$ . Transfer curves are shown either in the presence or in the absence of additional passive  $J$ -couplings; the latter are indicated by the letter 'd', which is placed close to the corresponding curve. The  $T_2$  relaxation is assumed to be 3 s, unless otherwise indicated. (A) Transfer curves for the different transfer steps in the HCNCH experiment in purines. (B) Transfer curves for the different transfer steps in the HCNCH experiment in pyrimidines. For uridine and cytosine the 'decoupled' transfer curves are identical. (C) Transfer curves of the final transfer steps (see text) in the HCNCH and the HCNH experiments in purines, indicated as  $Tr_{C1'zN9x,N9xC8z}$ ,  $Tr_{C8\alpha N9z,C8\alpha H8z}$  and  $Tr_{C1'zN9x,N9xH8z}$ ,  $Tr_{N9zH8x,H8y}$ , respectively. For the transfers in the HCNCH experiment only the 'decoupled' versions are shown. For the corresponding transfers in the HCNH experiment both the 'coupled' and 'decoupled' versions are shown. (D) As in (C), except a  $T_2$  relaxation time of 30 ms is assumed (see text).

In Fig. 21(A) and (B), the HCNCH transfer functions are shown for purines and pyrimidines, respectively, using realistic  $J$ -coupling constants. The presence of the additional passive  $J$ -couplings in non-selective transfers adversely affects the transfer efficiency. Farmer et al. [168] and Sklenar et al. [167] have therefore introduced selective refocusing in the various transfer steps. Considering in somewhat more detail the HCNCH pulse sequence by Sklenar et

al. [167], which is designed to perform well for both purine and pyrimidine residues, we find selective refocusing of the C1's in the C1'–N1/N9 step, which removes the passive coupling with C2'. As can be seen in Fig. 21(A) and (B) the selective decoupling of C2' is not essential for sensitivity. In the subsequent step, C1'zN9/N1x  $\rightarrow$  N9/N1y  $\rightarrow$  N9/N1xC6/C8z, an IBURP is employed, which simultaneously inverts C1' and C6/C8, so that C2

resonating at 176 ppm becomes decoupled from N1, C4 resonating at 151 ppm and C5 resonating at 120 ppm become decoupled from N9. The selective decoupling is essential for this transfer step since a considerable increase in the transfer efficiency is achieved (Fig. 21(A) and (B)). In purine C4 resonates quite close to C8 and C6, so that the selective IBURP on C8/C6 may also touch the C4 spins, and consequently, the  $J_{N9C4}$ -coupling may remain slightly effective. Further note that in U the  $J_{N1C2}$ -coupling is 19.2 Hz and in C 12 Hz. This transfer step will have a much lower efficiency in C than in U, when C2 is not decoupled from N1. In the third step, N9/N1zC6/C8x  $\rightarrow$  C8/C6y, a simultaneous REBURP and IBURP pulse is given on C6/C8 and N9/N1, respectively, which removes the coupling of C8 to C4 and C6 as well as to N7, and in pyrimidines the couplings of C6 to C5 and C2. As Fig. 21(A) and (B) show, the selective decoupling thus employed significantly increases the transfer magnitude in this transfer step. *It is important to note that the pulse sequence proposed by Sklenar et al. [167] functions for pyrimidines as well as for purines. The overall efficiency of the HCNCH will be quite high, since the efficiency is close to 1.0 for each transfer step, at least when  $T_2$  is large (the optimal transfer times lie for all transfer steps around 40 ms (see Fig. 21(A) and (B))).*

In this first step of the HCNCH experiment no adverse effect is felt from the presence of the passive  $J_{C1'C2}$ -coupling. Sklenar et al. [167] therefore proposed using two HCN 3D experiments to achieve the H1' to H6/H8 correlation, i.e. an HCN (sugar), which starts from H1' and transfers via C1' to N9/N1 and back, and an HCN (base), which starts from the base proton, H6/8, and transfers via C8/C6 to N1/N9 and back. In this way the H1'–C1'–N9/N1 correlation can be non-selective and is efficient for both purine and pyrimidines. The HCN (base) experiment, where H8–C8–N9 or H6–C6–N1 are correlated, employs selective pulses as in the HCNCH experiment. The main advantage of such an approach is that each of these experiments is more sensitive than the relayed experiment.

It is also interesting to compare the efficiency of the HCNCH experiment with the HCNH experiment proposed by Tate et al. [171] and the HbNb(Hb)CbHb experiment proposed by Farmer et al. [168]. In all of these experiments the  $J_{H8N9}$ -coupling is used to

transfer from N9 to H8. Since the HCN step in the HCNH is the same as in the HCNCH experiment, we compare in Fig. 21(C) only the efficiencies in the final transfer steps. As can be seen, the transfer  $C1'zN9x \rightarrow N9y \rightarrow N9xC8z$ , indicated as  $Tr_{C1'zN9x,N9xC8z}$ , and the transfer  $C1'zN9x \rightarrow N9y \rightarrow N9xH8z$ , indicated as  $Tr_{C1'zN9x,N9xH8z}$ , have comparable efficiencies, with the latter being slightly less effective. In the HCNCH experiment the refocusing step  $C8xN9z \rightarrow C8y \rightarrow C8xH8z$ , indicated as  $Tr_{C8xN9z,C8xH8z}$ , also has an optimal efficiency close to 1, while in the HCNH experiment the corresponding refocusing step,  $N9zH8x \rightarrow H8y$ , indicated as  $Tr_{N9zH8x,H8y}$ , has a much lower efficiency due to the remaining passive  $J_{H8N7}$ -coupling. If a selective IBURP on N9 had been used as was done by Farmer et al. [168] in the HbNb(Hb)CbHb experiment, the efficiency of this step would again be close to 1 (see Fig. 21(C)). Thus, the efficiency of the HCNH is not expected to be better than the efficiency of HCNCH. The HbNb(Hb)CbHb will have a better efficiency since the HCN step and the relay step are absent.

To further establish how well the transfer function calculations reproduce the experimental observations, we compare the optimal delay settings as suggested by the authors of the various pulse sequences with the optimal settings as deduced from the transfer functions. For this the effect of  $T_2$  relaxation needs to be considered. In Fig. 21(D), we therefore show the same transfer functions as in Fig. 21(A)–(C), but now calculated assuming a  $T_2$  of 30 ms. For the C1'N9 refocusing step Sklenar et al. and Farmer et al. [167,168] found 40 ms to be the optimal transfer time. This value lies inbetween the optimal values, 40 ms and 26 ms, calculated for  $T_2 = 3$  s and 30 ms, respectively. Tate et al. [171] did not apply C2'-decoupling during this transfer step in the HCNH experiment, and neither did Sklenar et al. [167] in the HCN (sugar) experiment. The former used 25 ms for the transfer time, which corresponds to the first optimum in the transfer function (Fig. 21(A)), while Sklenar et al. [167] used a value of 42 ms, which corresponds to the second maximum. Apparently, the DNA studied by Tate et al. [171] has shorter  $T_2$  values than the RNA studied by Sklenar et al. and Farmer et al. [167,168] in their experiments. In the relay step,  $C1'zN9x \rightarrow N9y \rightarrow C8zN9x$ , Sklenar et al. and Farmer et al. [167,168] use values for the transfer

time of 36 ms and 30–40 ms, respectively. As can be seen from Fig. 21(A)–(D), these values again lie inbetween the optimal values for the calculated transfers when  $T_2$  is 3 s and 30 ms, respectively. Tate et al. [171] used a transfer time of 32 ms for the transfer  $C1'zN9x \rightarrow N9y \rightarrow H8zN9x$ , which is very close to the optimal value for this transfer when  $T_2$  is 30 ms Fig. 21(D). For the refocusing of  $H8zN9x$ , Tate et al. [171] used a transfer time of 16 ms. This is again close to the optimal value for the calculated transfer when  $T_2$  is 30 ms (Fig. 21(D)). Thus, the calculated transfer curves explain nicely the experimentally obtained results. In fact, they could provide an alternative way to estimate the  $T_2$  governing the systems under study.

Since the calculated transfer functions explain well the observed experimental results, it is worth considering the relative merits of each approach in terms of sensitivity. For this we compare the overall efficiencies, which are calculated as the product of the transfer efficiencies of the individual transfer steps and compiled in Table 8. As can be seen, the overall transfer efficiency drops dramatically from close to 1, when  $T_2 = 3$  s, to much smaller values, ranging from 0.006 to 0.2, when  $T_2$  is 30 ms; for high values of  $T_2$  most of the experiments have quite similar efficiencies, but for  $T_2 = 30$  ms considerable differences appear. The HCN experiments, as well as the HbNb(Hb)CbHb experiment, are then significantly more efficient than the relay experiments, although the HCNCH experiment remains efficient enough to be performed. The advantage of the HCN and HCNCH experiments is that they have similar efficiencies for both purine and pyrimidine residues.

The (Hb,Nb) HSQC suggested by Sklenar et al. [172] has an efficiency similar to that of the HbNb(Hb)CbHb experiment (when  $T_2 = 3$  s, see Table 8). For  $T_2$  values of 30 ms the efficiency remains reasonable for purines, but becomes quite small for pyrimidines, which is due to the small size of the  $J_{H6N1}$ -couplings (see Table 8).

We finally consider the possibility of using the  $J_{H1'C6/8}$ -coupling and the  $J_{H1'C2/4}$ -coupling for obtaining base–sugar correlation via long-range HSQC or HMQC experiments. The  $J_{H1'C6/8}$ - and  $J_{H1'C2/4}$ -couplings depend on the  $\chi$  torsion angle and have values of 4–5 Hz and 2 Hz, respectively, when the  $\chi$  torsion angle is in the usual *anti* range [49]. The HSQC or HMQC experiments could be expanded

Table 8

Overall transfer efficiency<sup>a</sup> in base-to-sugar through-bond correlation

	Efficiency (%; $T_2$ 3 s)	Efficiency (%; $T_2$ 30 ms)
HCNCH <sup>a</sup>	82	2.5
HCNH <sup>b</sup> (Tate et al.)	20	1.2
HCNH <sup>c</sup> (Tate et al. optimal)	73 (10)	1.7 (0.2)
HbNb(Hb)CbHb <sup>d</sup>	90 (10)	6.3 (0.6)
HCN (sugar) <sup>e</sup>	98	10
HCN (base) <sup>e</sup>	96	10
(HbNb) HSQC <sup>f</sup>	98 (98)	6.3 (0.6)
(H1'C8/6) HSQC <sup>g</sup>	98	2.3
(H1'C4/2) HSQC <sup>h</sup>	98	0.6

<sup>a</sup>The overall transfer efficiencies are obtained as the product of the efficiencies of the individual steps. In parentheses are the values for pyrimidines when different from purine values. Optimal delay settings for (a) HCNCH,  $T_2 = 3$  s:  $\tau_1 = 1.5$  ms,  $\tau_2 = 42$  ms,  $\tau_3 = 40$  ms,  $\tau_4 = 42$  ms,  $\tau_5 = 1.2$  ms;  $T_2 = 30$  ms:  $\tau_1 = 1.5$  ms,  $\tau_2 = 25$  ms,  $\tau_3 = 31$  ms,  $\tau_4 = 26$  ms,  $\tau_5 = 1.2$  ms; (b) HCNH,  $T_2 = 3$  s:  $\tau_1 = 1.5$  ms,  $\tau_2 = 25$  ms,  $\tau_3 = 50$  ms,  $\tau_4 = 25$  ms,  $\tau_5 = 1.2$  ms;  $T_2 = 30$  ms:  $\tau_1 = 1.5$  ms,  $\tau_2 = 25$  ms,  $\tau_3 = 32$  ms,  $\tau_4 = 16$  ms,  $\tau_5 = 1.2$  ms; (c) HCNH,  $T_2 = 3$  s:  $\tau_1 = 1.5$  ms,  $\tau_2 = 48$  ms,  $\tau_3 = 50$  ms (50 ms),  $\tau_4 = 50$  ms (50 ms);  $T_2 = 30$  ms:  $\tau_1 = 1.5$  ms,  $\tau_2 = 25$  ms,  $\tau_3 = 32$  ms (32 ms),  $\tau_4 = 26$  ms (26 ms); (d) HbNb(Hb)CbHb,  $T_2 = 3$  s:  $\tau_1 = 50$  ms (50 ms),  $\tau_2 = 1.2$  ms;  $T_2 = 30$  ms:  $\tau_1 = 26$  ms (26 ms),  $\tau_2 = 1.2$  ms; (e) the overall efficiencies of HCN (sugar) and HCN (base) are the square of  $T(H1'C1') \cdot T(C1'N9)$  and  $T(H8C8) \cdot T(N9C8)$ , respectively, with the delays as in HCNCH; (f) (HbNb) HSQC  $T_2 = 3$  s:  $\tau_1 = 50$  ms (110 ms);  $T_2 = 30$  ms:  $\tau_1 = 26$  ms (20–30 ms); (g) (H1'C8/6) HSQC  $T_2 = 3$  s:  $\tau_1 = 55$  ms;  $T_2 = 30$  ms:  $\tau_1 = 26$  ms; (h) (H1'C4/2) HSQC  $T_2 = 3$  s:  $\tau_1 = 120$  ms;  $T_2 = 30$  ms:  $\tau_1 = 26$  ms.

into a 3D experiment, by adding HMQC type correlation to  $C1'$ , using the large  $^1J_{H1'C1'}$ -coupling, in much the same way as has been done for the HbNb(Hb)CbHb experiment. In combination with a (H8/6,C8/6) HSQC, a through-bond correlation can then be obtained between  $H1'$  and  $H6/8$  resonances. The experiment has a theoretical efficiency which is of the order of the efficiency for an  $H1'$  to  $C6/8$  correlation in an HCNCH experiment when  $T_2$  is 30 ms (Table 8).

In summary, from the viewpoint of sensitivity, and taking the values at  $T_2 = 30$  ms as a yardstick, the two 3D HCN experiments perform best, while the (HbNb) HSQC and the related HbNb(Hb)CbHb are also quite sensitive, followed by the HCNCH and the (H1'C6/8) HSQC. In practice, losses due to rf inhomogeneity

have to be taken into account. Limiting the number of pulses will improve sensitivity, although with improved probes, this aspect has become of lesser importance. Table 8 should therefore give a fair representation of the practical performance of the different base–sugar through-bond correlation experiments. Finally, we should consider the aspect

of resonance overlap. The best dispersed resonances are those of H1', C1', H6, H8, C6 and C8; the resonances of N9 and N1 form two well separated groups, but within each group overlap is quite strong. Ideally, one would therefore like to obtain triple correlations involving the best dispersed resonances, e.g. H1'–C8–C1'. Note that the best

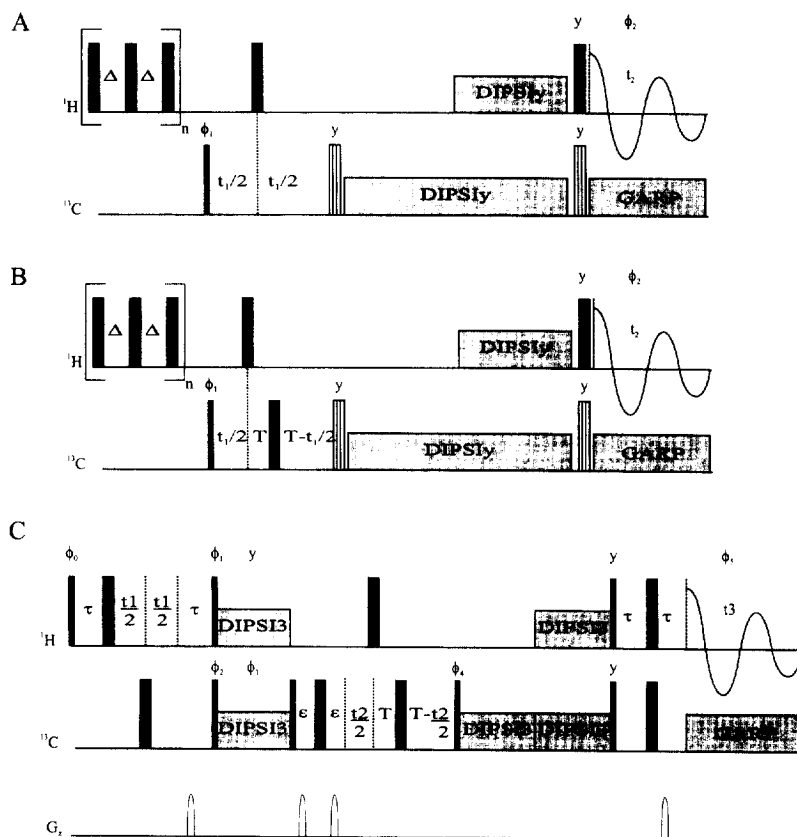


Fig. 22. Pulse sequences of the CP CCH-TOCSY (A) [76], constant time CP CCH-TOCSY (B) [76] and the 3D CP HCCH-TOCSY (C) [163] experiments to determine carbon to carbon connectivities in the ribose sugar ring. Narrow black rectangles indicate  $90^\circ$  pulses, wide black rectangles  $180^\circ$  pulses, vertically striped rectangles trim pulses bracketing the DIPSIS sequences, and unfilled rounded rectangles pulsed field gradients along the  $z$  axis. All pulses are applied along the  $x$  axis unless otherwise indicated. (A, B) Phase cycling is as follows:  $\phi_1 = x, -x$  (+ TPPI( $t_1$ ));  $\phi_2 = x, -x$ . The applied rf fields in the DIPSIS correspond typically to a  $90^\circ$  pulse of  $28 \mu\text{s}$ ; the cross-polarization period is normally set to about  $6.2 \text{ ms}$  ( $\approx 1/J_{\text{CH}}$ ); the homonuclear (C–C) mixing time can be set to approximately  $18 \text{ ms}$  in order to obtain sufficient transfer through the  $J$ -coupled  $^{13}\text{C}$  sugar ring spin system; the constant time period  $T$  is typically set to approximately  $12 \text{ ms}$  ( $\approx 1/2J_{\text{CC}}$ ). (C) All  $B_0$  gradient fields are sine-bell shaped. Phase cycle:  $\phi_0 = y; \phi_1 = x; \phi_2 = x, -x; \phi_3 = y; \phi_4 = y; \phi_5(\text{receiver}) = x, -x$ . Absorption mode spectra in the indirect dimensions are obtained by separate measurements of the N- and P-type coherences, which is achieved by simultaneously inverting the sign of the first gradient (G1) and of phase  $\phi_1$  for the  $t_1$  dimension and by simultaneously inverting the sign of the last gradient (G4) and of phase  $\phi_4$  for the  $t_2$  dimension. The values to be used for delay lengths, gradient amplitudes, gradient durations, and field strengths are discussed in detail in Wijmenga et al. [163]. In all cases (A,B,C) the heteronuclear and homonuclear cross-polarizations are carried out using synchronous DIPSIS sequences. Carbon decoupling during acquisition can be achieved via GARP sequence as indicated.

dispersion is of course obtained in the triple resonance experiments.

7.2.2.5. *Through-bond sugar correlation.* For correlation through the ribose sugar ring (step III,

Fig. 13(C)), HCCH experiments can be used. HCCH experiments have been developed in the field of protein NMR and come in a variety of forms. In all experiments the transfer from  $^1\text{H}$  to  $^{13}\text{C}$  is achieved either via a refocused INEPT [157,158] or via a

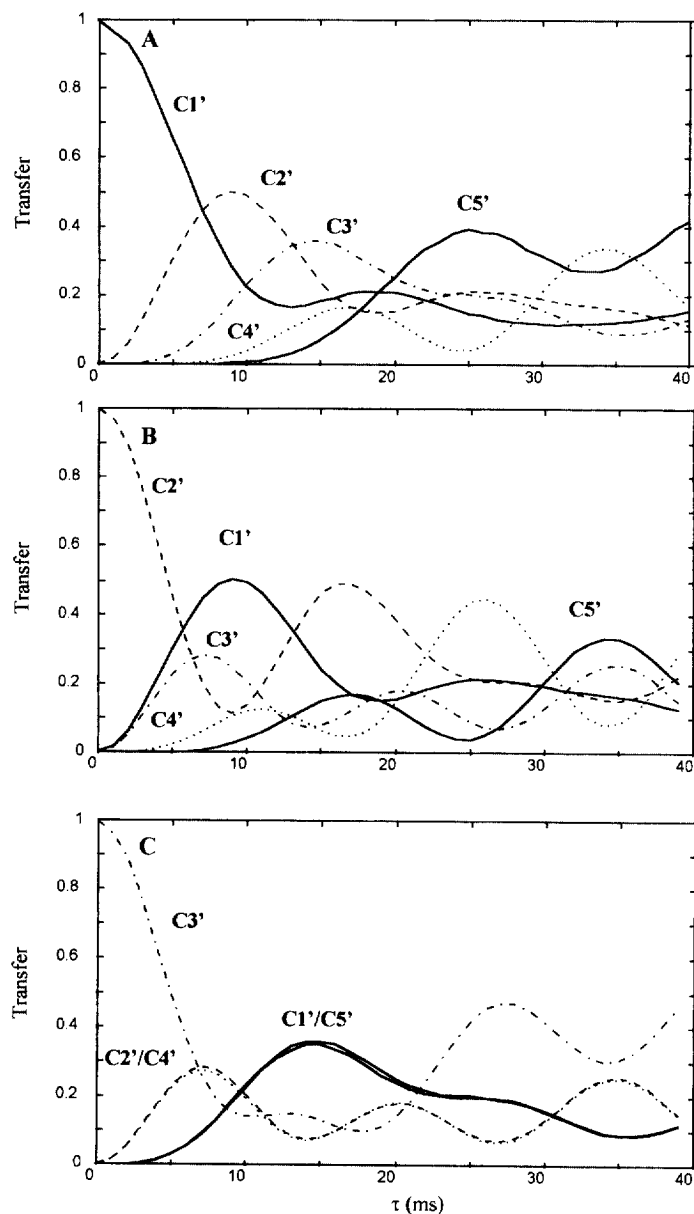


Fig. 23. (C,C) TOCSY transfer efficiencies in ribose sugar. (A) Transfer from C1' into the ring spin system; (B) transfer from C2' into the ring spin system; (C) transfer from C3' into the ring spin system. The transfer from C4' is identical to that of C2', and the transfer from C5' is identical to that of C1'. The transfer has been numerically simulated using the  $J_{\text{CC}}$ -coupling constants as given in Section 5, and offsets as given in Table 6; rf field assumed to be 20 kHz, so that offset effects are minimized.

cross-polarization sequence [156,164,165]. Since it has been demonstrated by Zuiderweg and coworkers [156,164,165] that cross-polarization is more effective than refocused INEPT, the HCCH experiments using cross-polarization are expected to have the higher sensitivity. Traditionally, phase cycling has been used to suppress artifacts or unwanted coherences, but nowadays, gradients are used to select the coherence pathways of interest and suppress artifacts. Kay et al. [159] were the first to show how gradient coherence selection can be combined with enhanced sensitivity, i.e. per transfer step the enhancement can be as much as a factor of  $2^{1/2}$ . Recently, this enhancement scheme has been incorporated into HCCH sequences, which

employ refocused INEPT sequences to affect  $^{13}\text{C}$  to  $^1\text{H}$  and  $^1\text{H}$  to  $^{13}\text{C}$  transfers [160–163]. Furthermore, doubly-enhanced 3D HCCH experiments have been published that give close to a factor of 2 higher sensitivity than the conventional 3D HCCH sequence using either refocused INEPT [161,162] or cross-polarization [163]. An important feature of these gradient-selective 3D HCCH experiments is their impressive water signal suppression. Finally, we note that the  $^{13}\text{C}$  evolution is often incorporated into a constant time period [166,173,174] in order to remove the splittings on the cross peaks stemming from the rather large  $J_{\text{CC}}$  couplings in the ribose ring, which are all approximately 40 Hz (see Section 5). Fig. 22(A)–(C)

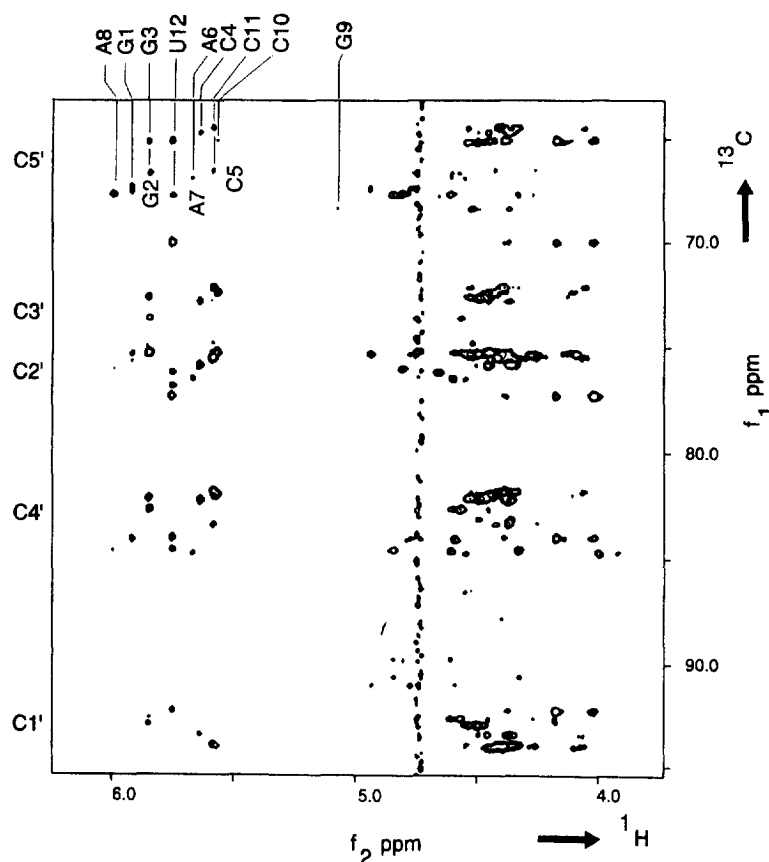


Fig. 24. Constant time CCH-TOCSY spectrum, showing the ribose sugar region (C1' to C5'). The spectrum is of a  $^{13}\text{C}$  labeled RNA hairpin with sequence 5'pppGGGC-CAAA-GCCU. The assignments are indicated in the  $\text{H1}'$  region. The spectrum was recorded with a constant time period of approximately 12 ms ( $1/2J_{\text{CC}}$ ); the (C,C) TOCSY (DIPSI-2) mixing time was 18 ms; the cross-polarization was 6.2 ms; the rf field 9 kHz.

show the pulse sequences for the 2D cpCCH and 3D cpHCCH experiments either with or without constant time  $^{13}\text{C}$  evolution.

In order to be able to select the optimal mixing times for  $^{13}\text{C}$  to  $^{13}\text{C}$  transfers, we show in Fig. 23(A)–(C) the (C,C) TOCSY transfer functions in the ribose sugar ring. As can be seen, in order to optimally transfer through the sugar ring C4' to C1', a (C,C) TOCSY mixing time of approximately 18 ms should be used, while maximal C1' to C5' transfer is achieved at 25 ms. Single step transfer, for example from C1' to C2', can be obtained when the mixing time is taken to be less than 6 to 7 ms. Thus, the HCCH experiment is ideally suited to establish which sugar ring carbons and/or protons belong to the same ribose. Because the C1', C2'/C3', C4' and C5' reside in clearly separate regions they can easily be distinguished, as shown in Fig. 24. To differentiate between C2' and C3' one can either use the HCCH experiment with short mixing times or instead a (C,C) COSY, which is the approach taken by Nikonowicz and Pardi [148].

**7.2.2.6. Through-bond sequential backbone assignment.** In the protein NMR field, the sequential resonance assignment along the backbone is now done via through-bond experiments. For labeled RNA and DNA a similar approach is possible (step III, Fig. 13(C)). A number of pulse sequences have been proposed that use various possible transfer routes to achieve sequential backbone assignment via through-bond coherence transfer. As shown in Fig. 25, a number of routes is available for this purpose. The first experiment, originally proposed by Pardi et al. [137] before the advent of  $^{13}\text{C}$  labeling, utilizes  $J_{\text{H}3'/\text{P}3'}$  and  $J_{\text{H}5'/5''/\text{P}1'}$ . With the advent of  $^{13}\text{C}$  labeling it became possible to achieve backbone assignment by through-bond coherence transfer via the generally rather large  $J_{\text{C}4'/\text{P}1'}$  and  $J_{\text{C}4'/\text{P}1'}$  in a 3D HCP experiment [56,76,175,176]. Heus et al. [175] proposed an assignment scheme that uses  $J_{\text{C}4'\text{P}}$  to establish the sequential connectivity, and  $J_{\text{C}5'\text{P}}$  to establish the 5'–3' directionality. A CCH–TOCSY experiment is used to establish the C5'*i* to C4'*i* identity [175]. The HCP pulse sequence is shown in Fig. 26(A). In this pulse sequence one starts with generating H4' coherence, which is transferred to C4', and subsequently is transferred to P, and back.

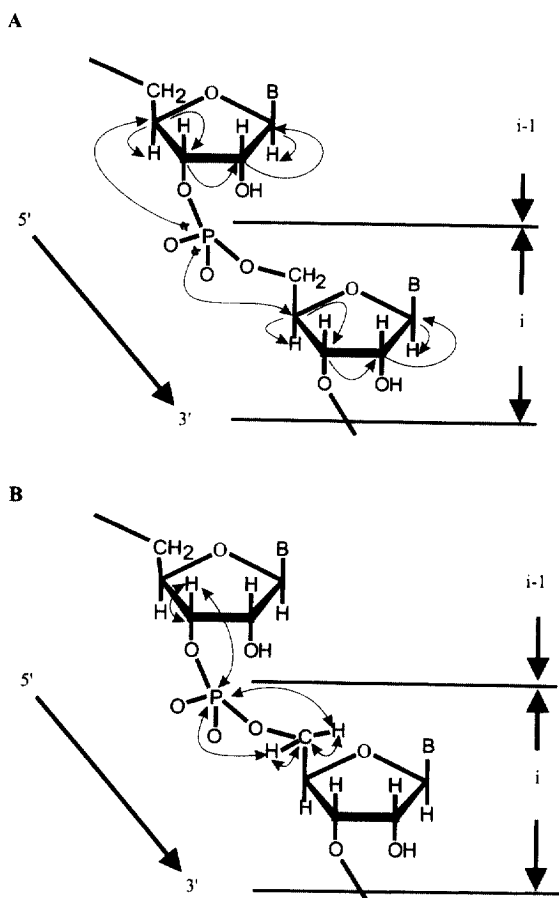


Fig. 25. Transfer routes for sequentially correlating residues via through-bond coherence transfer. (A) Transfer routes in HCP/PCH experiments and PCCH–TOCSY experiments; (B) transfer routes in HPHCH experiments.

Alternatively, as shown in the PCH pulse sequence in Fig. 26(B), one can start on P and transfer the P coherence via C4' to H4'. The sensitivity of these experiments is quite similar. Finally, the experiments can be converted into versions with gradients for coherence selection and with sensitivity enhancement. Since the H4' resonances as well as the P resonances tend to overlap quite strongly, it is better to transfer the magnetization to more dispersed resonances such as the H1' resonances. It is possible for this purpose to convert the PCH sequence into a PCCH–TOCSY sequence, in which the initially generated  $^{31}\text{P}$  coherence is

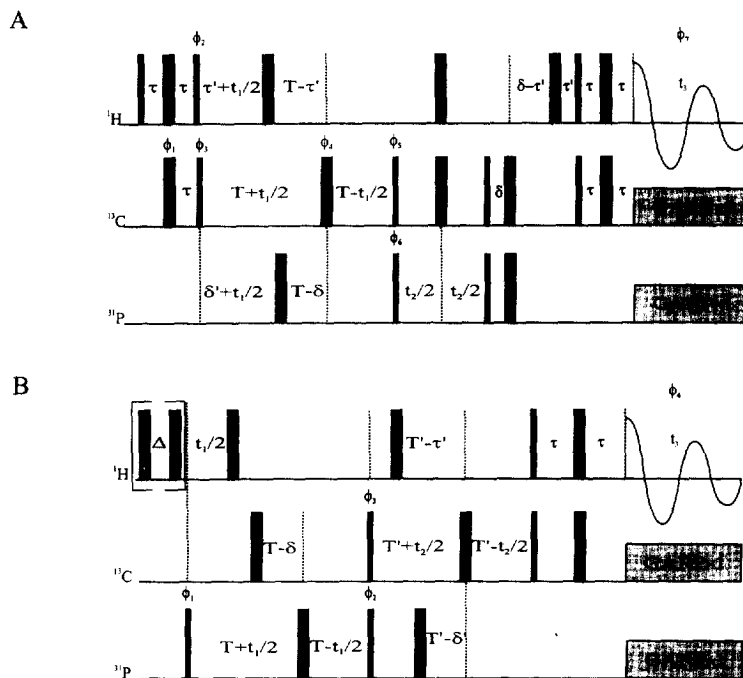


Fig. 26. Pulse sequence of the (A) HCP [76,175] and (B) PCH [76] experiments. Narrow black rectangles indicate  $90^\circ$  pulses, and wide black rectangles  $180^\circ$  pulses. All pulses are applied along the  $x$  axis unless otherwise indicated. (A) Phase cycling is as follows:  $\phi_1 = x, -x$ ;  $\phi_2 = y, -y$ ;  $\phi_3 = x( + \text{TPPI}(t_1))$ ;  $\phi_4 = 4(x), 4(y)$ ;  $\phi_5 = x, -x$ ;  $\phi_6 = x, x, -x, -x( + \text{TPPI}(t_2))$ ;  $\phi_7(\text{receiver}) = x, x, -x, -x, -x, -x, x, x$ . Typical settings are  $\tau = \tau' = 1.5 \text{ ms}$  ( $\approx 1/4J_{\text{CH}}$ ),  $t_1 \text{max}/2 = T = 12.5 \text{ ms}$ , and  $T - \delta = 5 \mu\text{s}$  (see text). (B) Phase cycling is as follows:  $\phi_1 = x( + \text{TPPI}(t_1))$ ;  $\phi_2 = y, -y$ ;  $\phi_3 = y, y, -y, -y( + \text{TPPI}(t_2))$ ;  $\phi_4(\text{receiver}) = x, -x, -s, x$ . Typical settings are  $\tau = \tau' = 1.5 \text{ ms}$  ( $\approx 1/4J_{\text{CH}}$ ),  $t_1 \text{max}/2 = T = 12.5 \text{ ms}$ ,  $T - \delta = 5 \mu\text{s}$ ,  $t_2 \text{max}/2 = T' = 12.5 \text{ ms}$ ,  $T' - \delta' = 5 \mu\text{s}$  (see text), and  $\Delta = 25 \text{ ms}$ .

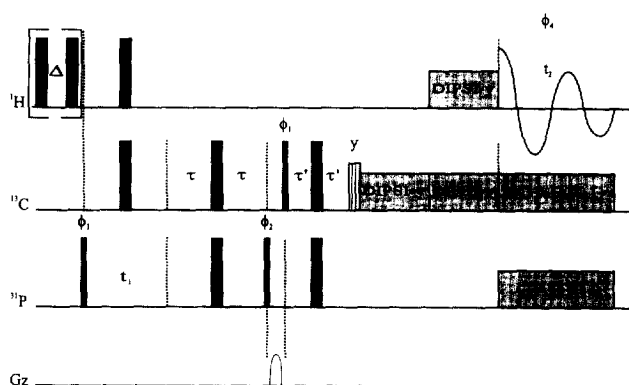


Fig. 27. Pulse sequence of the PCCH-TOCSY experiment [177]. Narrow black rectangles indicate  $90^\circ$  pulses, wide black rectangles  $180^\circ$  pulses, vertically striped rectangle trim pulse before the DIPSI sequences, and the unfilled rounded rectangle sine-bell shaped pulsed field gradient along the  $z$  axis. All pulses are applied along the  $x$  axis unless otherwise indicated. Phase cycling is as follows:  $\phi_1 = x, x, -x, -x$  (with TPPI [257]);  $\phi_2 = y$ ;  $\phi_3 = y, -y$ ;  $\phi_4(\text{receiver}) = x, -x, -x, x$ . Details concerning settings are given in the legend of Fig. 28.



transferred to C4', and then via the route C3' → C2' → C1' relayed to the less overlapping protons, the H1's [177]. The pulse sequence for the PCCH–TOCSY is shown in Fig. 27, and the experimental 2D spectrum in Fig. 28. As can be seen, a complete sequential walk can be made. A similar experiment has been proposed by Marino et al. [154].

A number of aspects may adversely influence the backbone assignment via these approaches. Firstly, the size of the  $J_{C_4'P}$ -couplings may decrease from 10 to 11 Hz, the value usually found in a regular helix, to smaller values. Secondly, the phosphorus resonances

as well as the proton (H4') and carbon (C4') resonances of the ribose tend to overlap quite strongly. Varani et al. [178] have therefore proposed performing additional HCP triple resonance experiments, namely, a HPHCH experiment (Fig. 29). Here, coherence is transferred from H3' and H5'/5'' to P for frequency labeling, then transferred back to proton coherence, after which it is transferred to C3' and/or C5' for carbon frequency labeling and finally back to proton coherence. This approach is very similar to that employed in the HbNb(Hb)CbHb experiment of Farmer et al. [168]. The advantage of using such an additional experiment is that certain cross peaks that

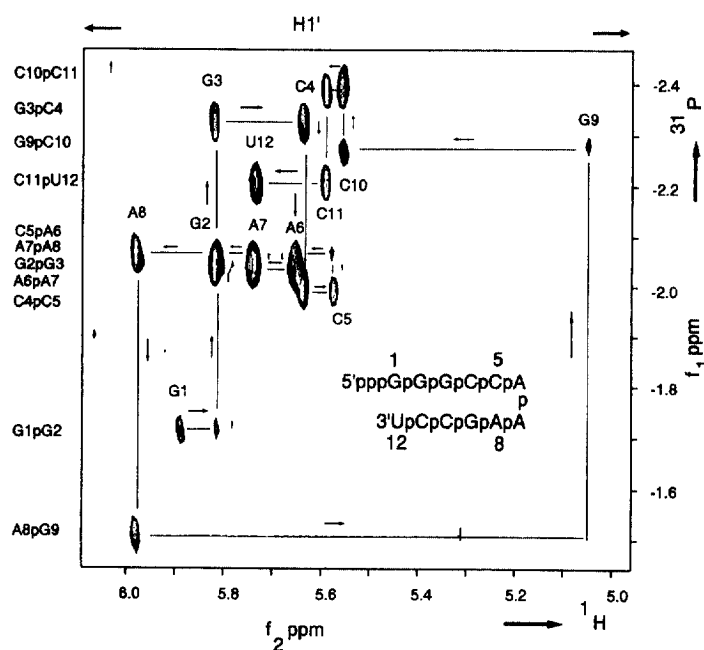


Fig. 28. H1' region of the 600 MHz 2D P(CC)H–TOCSY spectrum of a  $^{13}\text{C}$  labeled RNA hairpin with a CAAA tetranucleotide loop, shown schematically in the inset, in  $\text{D}_2\text{O}$  at  $30^\circ\text{C}$ . The arrows indicate the  $5' \rightarrow 3'$  direction of the sequential walk along the sugar–phosphate backbone. The labels at the cross peaks indicate the H1' assignment, while the  $^{31}\text{P}$  assignment is given along the left side of the spectrum at the vertical position ( $^{31}\text{P}$ ) of the cross peaks. The chemical shifts are referenced relative to TSP for the  $^1\text{H}$  and  $^{31}\text{P}$  dimensions; for  $^{31}\text{P}$  the 0 ppm chemical shift value is obtained by multiplying the  $^1\text{H}$  TSP frequency by 0.40480793, which corresponds to calibration relative to inorganic phosphate. The  $^{13}\text{C}$  labeled RNA CAAA hairpin was prepared as described by Wijmenga et al. [56]. The P(CC)H–TOCSY experiment was performed on a Bruker AMX(2)600 spectrometer, equipped with a broadband  $^{13}\text{C}/^1\text{H}$  probe. The spectrum was recorded in approximately 12 h with the acquisition settings: 512 scans for each FID of 1024 points ( $t_2$ ), 64  $t_1$  values. The delay  $\Delta$  was set to 25 ms;  $\tau$ ,  $\tau' = 12.5$  ms; 1 ms trim pulse; DIPSI3 mixing time for isotropic mixing equal to 13.0 ms and DIPSI3 mixing time for cross-polarization set to 6.5 ms, both with an rf field strength 8333.3 Hz, 1.0 s relaxation delay, 1 ms gradient pulse (800  $\mu\text{s}$  gradient pulse shaped to a 1% truncated sine envelope and 200  $\mu\text{s}$  for the magnetic field recovery) of strength 12 G/cm, low power GARP [247] decoupling (rf strength 625 Hz) of  $^{13}\text{C}$ , spectral width 486 Hz, and 5000 Hz for  $^{31}\text{P}$  and  $^1\text{H}$  respectively; carrier positions at  $-2.09$ ,  $70$  and  $4.62$  ppm for  $^{31}\text{P}$ ,  $^{13}\text{C}$  and  $^1\text{H}$  receptively. Typical processing parameters were zero-filling twice in  $t_1$  and once in  $t_2$ , and  $0.4\pi$  shifted squared sine window multiplication in  $t_1$  and  $t_2$ . The final data matrix consisted of  $128 \times 1024$  data points. Reproduced with permission from Ref. [177].

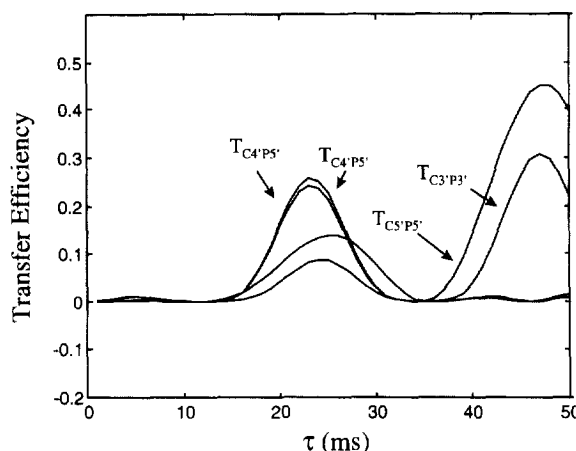


Fig. 29. Efficiency of the HCP experiment for C4' to P5', C4' to P3', C5' to P5', and C3' to P3' transfer. The transfer efficiency is calculated according to the equations in the text as a function of the C → P transfer time ( $\tau$ ); the overall efficiency for C to P transfer and back is the product of the identical transfer functions; the  $J$ -couplings are assumed to be 11 Hz for  $^3J_{C4'P5'/3'}$ , -5 Hz for  $^2J_{C5'P5'}$ , and -4 Hz for  $^2J_{C3'P3'}$ ;  $T_2$  relaxation has not been taken into account.

may be absent in the HCP/PCH experiment, due to inadvertently small  $J_{C4'P}$ -couplings, may still show up in the HPHCH experiment. Finally, Ramachandran et al. [179] have introduced a simultaneous HCP/HCN experiment which in principle combines the sequential backbone assignment via through-bond coherence transfer and sugar-to-base correlation via through-bond coherence transfer.

In order to find the optimal settings for the various experiments the coherence transfer functions have to be considered:

HCP : (51)

$$Tr_{H4'C4'} = \sin(\pi J_{H4'C4'} 2\tau_1)$$

$$Tr_{C4'IP5'i} = \sin(\pi J_{C4'IP5'i} 2\tau_2) \cos(\pi J_{C4'IP3'i+1} 2\tau_2) \\ \times \cos(\pi J_{C4'IC5'i} 2\tau_2) \cos(\pi J_{C4'IC3'i} 2\tau_2)$$

$$Tr_{C4'IP3'i+1} = \cos(\pi J_{C4'IP5'i} 2\tau_2) \sin(\pi J_{C4'IP3'i+1} 2\tau_2) \\ \times \cos(\pi J_{C4'IC5'i} 2\tau_2) \cos(\pi J_{C4'IC3'i} 2\tau_2)$$

PCH : (52)

$$Tr_{C4'IP5'i} = \sin(\pi J_{C4'IP5'i} 2\tau_2) \cos(\pi J_{C4'i-1P5'i} 2\tau_2) \\ \times \cos(\pi J_{C5'IP5'i} 2\tau_2) \cos(\pi J_{C3'i-1P5'i} 2\tau_2) \\ \times \cos(\pi J_{C2'i-1P5'i} 2\tau_2)$$

$$Tr_{C4'IP5'i} = \sin(\pi J_{C4'IP5'i} 2\tau_3) \cos(\pi J_{C4'IP3'i+1} 2\tau_3) \\ \times \cos(\pi J_{C4'IC5'i} 2\tau_3) \cos(\pi J_{C4'IC3'i} 2\tau_3)$$

$$Tr_{C4'i-1P5'i} = \cos(\pi J_{C4'IP5'i} 2\tau_2) \sin(\pi J_{C4'i-1P5'i} 2\tau_2) \\ \times \cos(\pi J_{C5'IP5'i} 2\tau_2) \cos(\pi J_{C3'i-1P5'i} 2\tau_2) \\ \times \cos(\pi J_{C2'i-1P5'i} 2\tau_2)$$

$$Tr_{C4'i-1P5'i} = \sin(\pi J_{C4'i-1P5'i} 2\tau_3) \\ \times \cos(\pi J_{C4'i-1P5'i-1} 2\tau_3) \\ \times \cos(\pi J_{C4'i-1C5'i-1} 2\tau_3) \\ \times \cos(\pi J_{C4'i-1C3'i-1} 2\tau_3)$$

$$Tr_{H4'C4'} = \sin(\pi J_{H4'C4'} 2\tau_1)$$

HP(H)CH : (53)

$$Tr_{H3'IP3'i} = \sin(\pi J_{H3'IP3'i} 2\tau_1) \cos(\pi J_{H3'IH4'i} 2\tau_1) \\ \times \cos(\pi J_{H3'IH2'i} 2\tau_1)$$

$$Tr_{H3'IC3'i} = \sin(\pi J_{H3'IC3'i} 2\tau_2)$$

$$Tr_{H5'IP5'i} = \sin(\pi J_{H5'IP5'i} 2\tau_1) \cos(\pi J_{H5'IH4'i} 2\tau_1) \\ \times \cos(\pi J_{H5'IH5'i} 2\tau_1)$$

$$Tr_{H5'IC5'i} = \sin(\pi J_{H5'IC5'i} 2\tau_2) \cos(\pi J_{H5'IH5'i} 2\tau_2)$$

Fig. 29 shows the transfer function of the HCP experiment. As can be seen, the optimal transfer is achieved for  $\tau_2$  equal to 12 ms. It can also be concluded from these curves that decoupling of C4' from C3' and C5' will not greatly improve the transfer efficiency. The transfers from C3' to P3' and C5' to P5' are considerably less effective due to the smaller  $J_{C5'P5'}$ ,  $J_{C3'P3'}$ -couplings involved, but may still be observed. The efficiency of the transfer from C4' to P3' and P5' depends on the torsion angles  $\epsilon$  and  $\beta$ ,

respectively. As discussed previously, the allowed range of the torsion angle  $\varepsilon$  is  $170^\circ$  to  $290^\circ$ . It usually lies in the range  $170^\circ$ – $240^\circ$ , but may in exceptional cases become  $g^-$ . When  $170^\circ < \varepsilon < 240^\circ$ , the transfer from C4' to P3' remains quite high for all values of  $\beta$ , as when this torsion angle lies outside its usual trans range (Fig. 30(A)). When  $\varepsilon > 240^\circ$  the efficiency drops, but then transfer from C2' to P3' increases to detectable values (Fig. 30(C)). Thus, the transfer to P3' should always be visible, either via the C4' to P3' pathway or via the P2' to P3' pathway. On the other hand, the transfer from C4' to P5' may drop to very low values when the torsion angle  $\beta$  lies outside its usual trans range (Fig. 30(B)). Here, no other carbon can take over, instead it is either the H5' to P5' transfer, that becomes quite effective ( $\beta > \approx 240^\circ$ ), or the H5'' to P5' transfer that becomes effective ( $\beta < \approx 240^\circ$ ). The latter correlations can be detected in an HP(H)CH experiment (Fig. 30(D)). Thus, when the sequential backbone assignment is performed via a combination of HCP (or PCH) and HP(H)CH experiments the torsion angle dependence of the J-couplings involved can adversely affect the presence of sequential correlation.

Another aspect is the relative efficiency of the various experiments and the effect of  $T_2$  relaxation on the efficiencies. In Table 9, the optimal transfer efficiencies are given for the various proposed experiments assuming a  $T_2$  of 3 s and of 30 ms. The most interesting conclusion that can be drawn is that sufficient transfer will occur even for relatively small  $T_2$  values, i.e. for large systems (50 nucleotides). Whether sequential backbone assignment is possible for such large systems depends therefore mainly on the aspect of resonance overlap. Unfortunately, the P resonances overlap strongly, and the same applies for the ribose H2', H3', H4' and H5'/H5''. It is therefore expected that for larger systems one has to resort to labeling only specific residues to achieve the assignment (see below).

### 7.2.3. X-filter techniques

X-filters form a very powerful tool in the field of NMR spectroscopy of biomolecules. Otting et al. [180] were the first to introduce so-called X-filters into the field of NMR of proteins; here X stands for X-nucleus, which can be  $^{13}\text{C}$  or  $^{15}\text{N}$ , for example. These methods can also be used to great advantage

in the field of NMR of  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled nucleic acids.

X-filters can be incorporated into any pulse sequence, but we will illustrate their usage in NOESY spectroscopy and consider the application given by Van Dongen et al. [143]. They used X-filters to unambiguously demonstrate the formation of a DNA·DNAXRNA triple helix, which consisted of a  $^{13}\text{C}/^{15}\text{N}$  labeled single-stranded RNA nucleotide sequence, associated with an unlabeled DNA hairpin with a B-helix stem. Fig. 31(A) shows the NOESY pulse sequence with an  $\omega_2$   $^{15}\text{N}$  filter, while (B) shows the NOESY pulse sequence with an  $\omega_1$   $^{13}\text{C}$  filter. There are a variety of possibilities in the application. Van Dongen et al. [143] used an  $\omega_2$   $^{15}\text{N}$  filter in NOESY experiment I and an  $\omega_1$   $^{13}\text{C}$  filter in NOESY experiment II. In experiment I, the  $\omega_2$   $^{15}\text{N}$  filter can be set in such a way that magnetization of protons directly bonded to  $^{15}\text{N}$  have either a positive (Ia) or negative sign (Ib) during acquisition. The experiment can be set up in such a way that Ia and Ib are recorded interleaved. When processing the data the two data sets are either added, giving the NOESY spectrum ADDI (= Ia + Ib), or subtracted giving the NOESY spectrum SUBI (= Ia – Ib). Consequently, in the NOESY spectrum SUBI only those cross peaks which correspond to protons that are directly bonded to  $^{15}\text{N}$  during acquisition will be present, while in the NOESY spectrum ADDI only those cross peaks will be present of protons not bonded to  $^{15}\text{N}$ . Similarly, for the  $\omega_1$   $^{13}\text{C}$  filtered NOESY, cross peaks which stem from protons that are directly bonded to  $^{13}\text{C}$  in the  $\omega_1$  dimension are found in the NOESY spectrum SUBII, while in the NOESY spectrum ADDII, cross peaks of protons that are not bonded to  $^{13}\text{C}$  in the  $\omega_1$  dimension are found. Thus in the imino region of the NOESY spectrum of the DNA·DNAXRNA triplex molecule the DNA and RNA NOESY spectra can be separated in the  $\omega_2$  dimension. The ADDI spectrum contains only cross peaks corresponding to NOE contacts to DNA imino (Fig. 32(A)) and amino protons (Fig. 32(B)) protons, while in a SUBI spectrum only NOE contacts to RNA imino (Fig. 32(C)) and amino protons (Fig. 32(D)) are found. In the latter imino spectrum (Fig. 32(C)) NOE cross peaks are present at 2 ppm, which is a spectral region where only H2'/2'' protons of DNA reside, thus demonstrating unambiguously that the RNA is associated with the DNA hairpin.

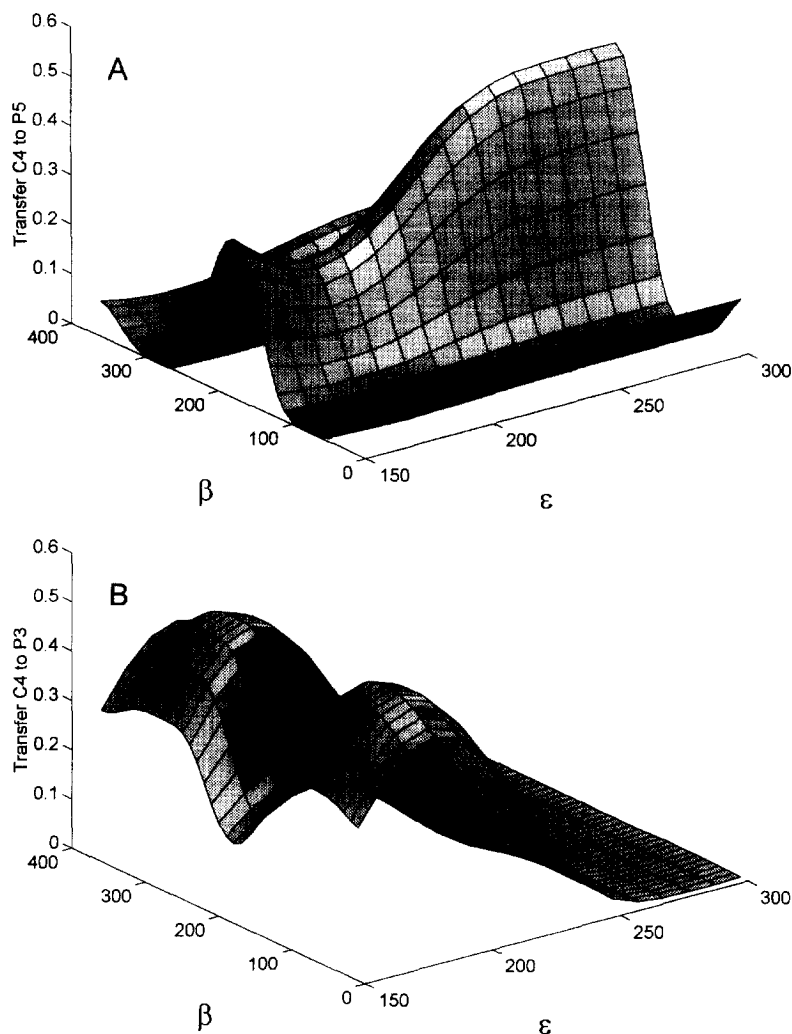


Fig. 30. Efficiency for C4'  $\rightarrow$  P3' (A), C4'  $\rightarrow$  P5' (B), C2'  $\rightarrow$  P3' (C), and H5'/H5''  $\rightarrow$  P5' (D) transfer and back as a function of the torsion angles  $\epsilon$  and  $\beta$ . The efficiency is calculated according to the transfer functions given in the text; optimal delay settings are assumed, i.e. 24 ms for the transfer time in (A) to (C), and 19 ms for the time for the H5'/H5''  $\rightarrow$  P5' transfer (D) (see also Table 9);  $T_2$  relaxation has not been taken into account.

The two X-filters were not simultaneously incorporated into the NOESY sequence because of sensitivity limitations, although this would conceptually have been more elegant, and intrinsically more powerful. In that case four FIDs could have been recorded, either with both filters off (no change in sign in  $\omega_1$  and  $\omega_2$ ) (I), or with the  $\omega_1$   $^{13}\text{C}$  filter on and the  $\omega_2$   $^{15}\text{N}$  filter off (negative sign in  $\omega_1$  and positive in  $\omega_2$ ), or with the  $\omega_1$   $^{13}\text{C}$  filter off and the  $\omega_2$   $^{15}\text{N}$  filter on (positive in  $\omega_1$  and negative sign in  $\omega_2$ ), or with the  $\omega_1$   $^{13}\text{C}$  filter on and the  $^{15}\text{N}$  filter on

(negative in  $\omega_1$  and negative in  $\omega_2$ ). From these four data sets it is possible to create four NOESY spectra, i.e. a NOESY spectrum which contains only DNA to DNA NOE contacts (I + II + II + IV), or only RNA to RNA NOE contacts (I - II - III + IV), or only DNA to RNA contacts (I + II - III - IV), or only RNA to DNA contacts (I - II + II - IV). The latter two would contain all the peaks that directly demonstrate the RNA to DNA interactions.

The X-filtered NOESY is an extremely powerful experiment and has a wide range of applications.

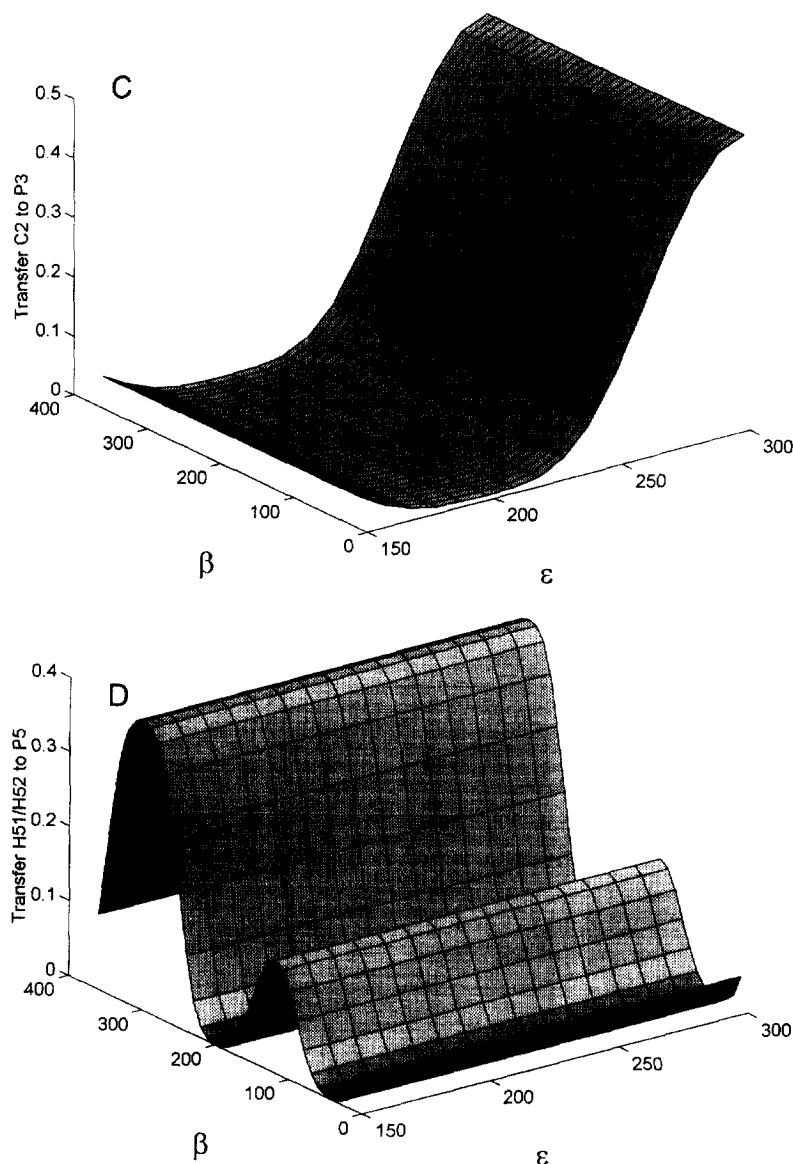


Fig. 30. Continued.

For example, it has been used in the study of DNA–protein interactions (see, e.g. Ref. [181]) or RNA–protein or peptide complexes [31,33,37,182–184]. It has been used to derive NOE contacts in a symmetric ssDNA binding protein [185–187]. It could also be of great value in deriving NOE contacts in selectively labeled RNAs or DNAs. For example, if only the A residues are labeled a NOESY spectrum can be derived that contains only the intra-residue NOE

contacts and a spectrum that contains only the inter-residue contacts.

## 8. Relaxation and dynamics

For relaxation in nucleic acids two mechanisms of interaction with the environment are important: the dipole–dipole interaction and the chemical shift

Table 9

Overview transfer efficiencies in HCP type experiments<sup>a</sup>

	C4'P	C3'/H3'P	C5'/H5'P	C4'P	C3'/H3'P	C5'/H5'P
HCP <sup>a</sup>	28	15	8	6	3	1.8
HCP (dec) <sup>b</sup>	30	45	30	7	3	3
HP(H)CH <sup>c</sup>	–	25 – (10 – 3)	45 – (10 – 1.2)	–	6 – (2 – 0.6)	18 – (2 – 0.4)
PCH <sup>d</sup>	26	15(20)	5(13)	6	3	1.3

<sup>a</sup> The overall transfer efficiencies,  $T_{xxx}$ , are given in % and calculated as the product of the transfer functions,  $T_{xxx}$ , of the separate transfer steps (eqns (51)–(53)); xxx stands for C4'P, C3'P or C5'P, and represent transfer from Cx to P. It is assumed that the H to C transfer is 100% effective for C–H systems, whereas for CH<sub>2</sub> spin systems a smaller value may be obtained. Columns two to four give the transfer efficiencies for  $T_2 = 3$  s, the last three columns give the transfer efficiencies for  $T_2 = 30$  ms; the  $J$ -coupling values used are (Hz):  $J_{H4/C4'} = 145$ ;  $J_{H5/H5'C5'} = 145$ ;  $J_{H3/C3'} = 145$ ;  $J_{C4'P3'} = 10.8$ ;  $J_{C4'P5'} = 11$ ;  $J_{C5'P5'} = -5$ ;  $J_{C3'P3'} = -4$ ;  $J_{C4'C3'} = 43$ ;  $J_{C4'C5'} = 43$ ;  $J_{C3'C2'} = 43$ ;  $J_{H5/H5''} = -14$ .

<sup>b</sup> HCP: optimal setting for delay  $2\tau_2$  is 24 ms for all transfers for both  $T_2 = 3$  s and  $T_2 = 30$  ms.

<sup>c</sup> HCP (dec), here it is assumed that C5' can be decoupled from C4'; C4' can be decoupled from C3' and C5'; C3' can be decoupled from C4' (C2' still present): optimal setting delay  $2\tau_2$  is 24 ms for C4'P, 50 ms for C3'P and C5'P transfers when  $T_2 = 3$  s; for  $T_2 = 30$  ms: all optimal delays range 16 to 18 ms.

<sup>d</sup> optimal delay settings all 18–20 ms; assumed is that the sugar conformer is N-type, and that  $\gamma$  is gauche +; the three numbers represent efficiencies under three conditions: the first number represents the efficiency when  $J_{H3'P}$  or  $J_{H5'/H5''P}$  have maximum value; the second and the third (in bold) correspond to the limiting values of  $J_{H3'P}$  or  $J_{H5'/H5''P}$  usually found in a double helix.

<sup>e</sup>  $T_2 = 3$  s: optimal setting for delays,  $2\tau_2$  and  $2\tau_3$  is about 24 ms for C4'P; for C3'P and C4'P optimal setting is 18–20 ms for the first transfer step and 45 ms for the refocus step, which are indicated by the number in parentheses;  $T_2 = 30$  ms: first delay at 16–18 ms and second at 24 ms for C4'P, C3'P and C5'P.

anisotropy (CSA). In the case of dipole–dipole interaction between two nuclei of spin  $1/2$ , rotational motion of the vector connecting the two interacting nuclei relative to the external magnetic field modulates the interactions and thus causes transitions between the energy levels of the spin  $1/2$  nuclei if the rates of the motions correspond to the nuclear transition frequencies. Similarly, modulation of the length of the vector occurring at the right frequencies can cause transitions. The latter is not important if the distance between the interacting nuclei is fixed by a chemical bond. This makes <sup>15</sup>N and <sup>13</sup>C relaxation attractive for studies of mobility because they relax primarily by dipole–dipole interaction with directly bonded protons. The chemical shift of a nucleus expresses the shielding of the external magnetic field by its environment. This shielding depends on the orientation of the chemical structure relative to the external magnetic field, i.e. it depends on the orientation of the principal axes of the chemical shift tensor relative to the external magnetic field. Rotational diffusion of the nucleic acid and internal mobility can modulate this effect and thus cause transitions between the two energy levels of the spin  $1/2$  nucleus if the rates of the motions correspond to the nuclear transition frequency. Hence, CSA relaxation depends on whether the chemical

structure in the nucleic acid to be investigated moves with frequencies corresponding to differences between energy levels. The distribution of these frequencies is usually expressed with so-called power spectral density functions (see, for example, Ref. [188]):

$$J^{AB}(\omega) = 2 \int_0^{\infty} \cos(\omega t) \langle A(0)B(t) \rangle dt$$

$$= 2 \int_0^{\infty} \cos(\omega t) G^{AB}(t) dt \quad (54)$$

The symbols  $A(t)$  and  $B(t)$  stand for the Hamiltonians relevant for fluctuating components of the relaxation, such as dipole–dipole interaction and CSA. The brackets represent the average over all molecules in the sample. The ensemble average of the product function,  $\langle A(0)B(t) \rangle$ , is also called the time correlation function,  $G^{AB}(t)$ , which for the dipole–dipole and CSA relaxation is given by:

$$G_0(t) = \frac{1}{5} \langle P_2 \{ \cos[\chi(0) - \chi(t)] \} \rangle \quad (55)$$

Here,  $\chi$  is the angle of the dipole–dipole vector or the principal axis of the CSA tensor relative to the external field. We call this vector the *relaxation vector*.

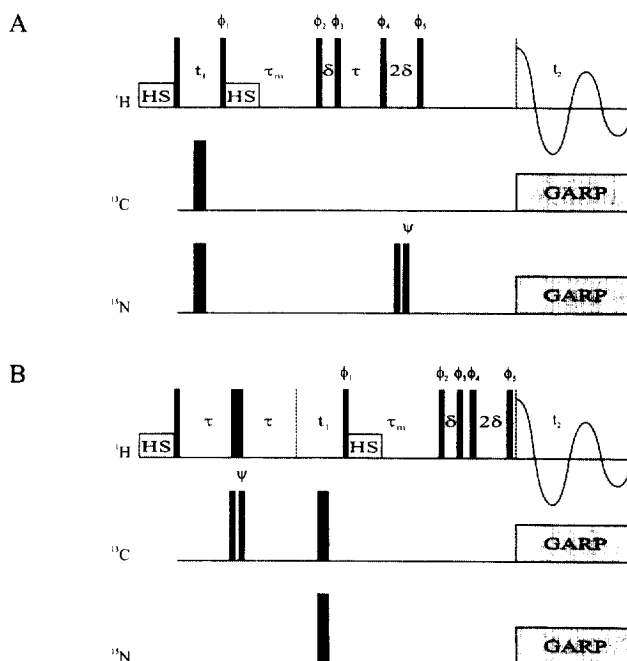


Fig. 31. Pulse schemes of the  $^{15}\text{N}$  ( $\omega_2$ ) filtered NOESY (A) and the  $^{13}\text{C}$  ( $\omega_1$ ) filtered NOESY (B) experiments. Homospoil delays of 20 ms, which were applied during the relaxation delay and the NOE mixing period, are indicated by blocks marked with HS. Phase cycle for both experiments:  $\phi_1 = 223^\circ$ ;  $\phi_2 = 4(x), 4(-x)$ ;  $\phi_3 = 4(-x), 4(x)$ ;  $\phi_4 = x, -x, y, -y, -x, x, -y, y$ ;  $\phi_5 = -x, x, -y, y, x, -x, y, -y$ ;  $\phi_6(\text{receiver}) = 2(x), 4(-x), 2(x)$ . Two separate data sets are obtained with  $\psi = x$  and  $\psi = -x$ , respectively. To achieve phase-sensitive detection in the indirect dimension, the phase of the  $^1\text{H}$  excitation pulse was alternately  $x$  and  $y$ , and the procedure was completed during processing by inverting the signs of every third and fourth data point. This approach allows the  $\text{H}_2\text{O}$  signal to be kept prior to acquisition for both the  $x$  and  $y$  phases of the  $^1\text{H}$  excitation pulse in the  $+z$  half of the  $xyz$  sphere, reducing radiation damping effects. In addition, keeping  $(1/2)$ , 2 of the  $\text{H}_2\text{O}$  signal along the  $+z$  axis prior to acquisition increases the signal intensity of the exchangeable protons [258]. The pulse schemes are adapted from Ref. [143].

In order to calculate  $G_0(t)$  a description of motion is required. We first consider the rotation of internally rigid molecules. The simplest motion is the isotropic rotational motion shown by spherical molecules, the time constant of which can be calculated from the Stokes–Einstein equation:

$$\tau_c = \frac{4\pi\eta R^3}{3kT} \quad (56)$$

Here  $\eta$  is the viscosity of the solvent,  $k$  is the Boltzmann constant,  $T$  is the absolute temperature, and  $R$  is the hydrodynamic radius. The correlation function decays exponentially for isotropic rotational motion with the time constant  $\tau_c$ :

$$G(t) = \frac{1}{5} \exp(-t/\tau_c) \quad (57)$$

so that the corresponding spectral density is given by:

$$J(\omega) = \frac{2}{5} \frac{\tau_c}{1 + (\omega\tau_c)^2} \quad (58)$$

The assumption of isotropic tumbling does not exactly hold for nucleic acids because the form of these molecules is not spherical but rather cylindrical. For a molecule with cylindrical or ellipsoidal shape two rotation correlation times exist, namely, a time constant for the rotation of the long axis,  $\tau_L$ , and one for rotation of the short axis,  $\tau_s$ . The time constants  $\tau_L$  and  $\tau_s$  can be related to the molecular radius  $R$  and the length  $L$  by [189]:

$$\tau_L = \frac{\pi\eta L^3}{18kT} (\ln(L/2R) + \delta_L)^{-1} \quad (59)$$

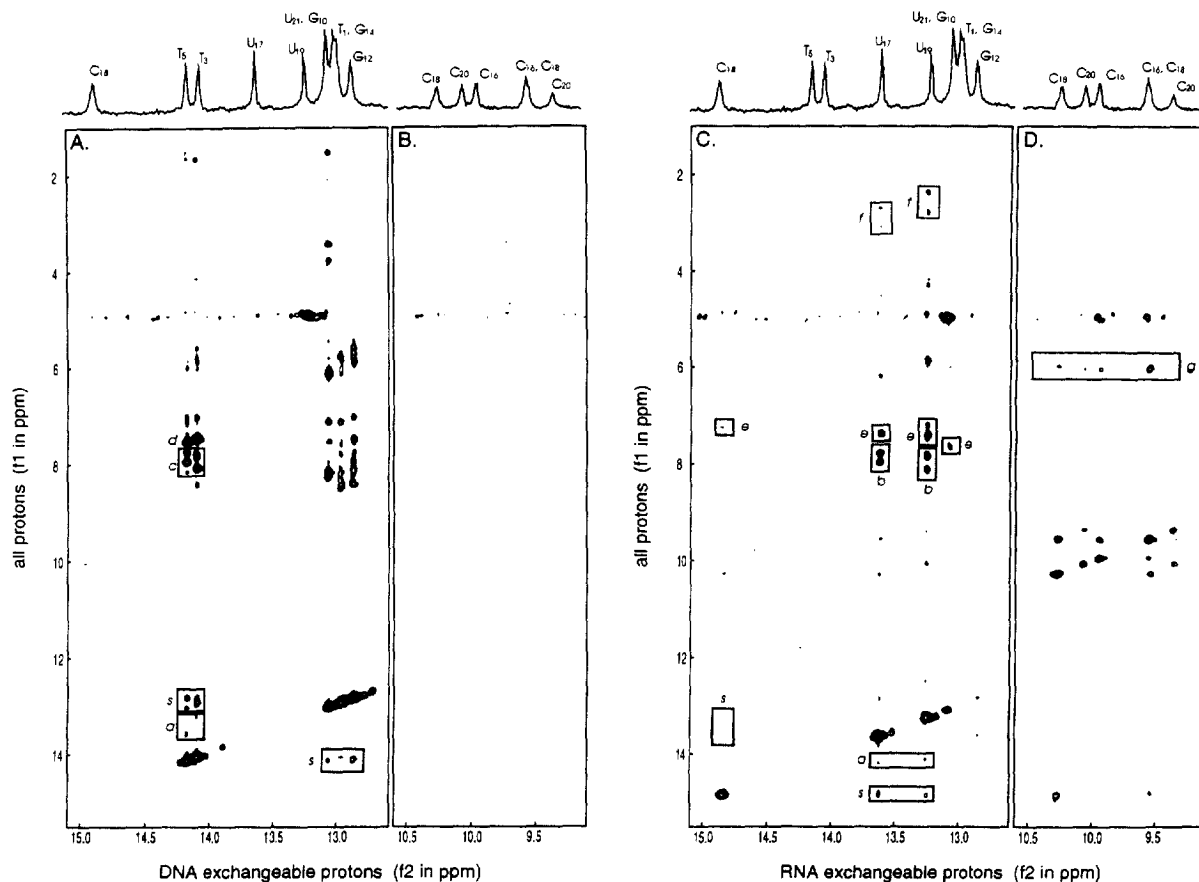


Fig. 32. Parts of the spectrum after summation (A, B) and subtraction (C, D) of the  $\psi = x$  and  $\psi = -x$  data sets from the  $^{15}\text{N}$  ( $\omega_2$ ) filtered NOESY experiment of hybrid DNA-DNAxRNA triple helix, in which the RNA is  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled. The DNA sequence is 5'TCTCTC-TTT-GAGAGA and folds as a hairpin. The RNA has sequence 5'CUCUCU and is single stranded at high pH. At low pH the RNA cytosines are protonated, and the RNA strand binds to the major groove of the DNA hairpin so that the triple helix is formed. The sum spectrum shows NOEs to NH protons (A), and NH2 protons (B) of the DNA; the difference spectrum shows NOEs to NH protons (C), and NH2 protons of the RNA and NH2 protons (D) of the RNA. In the spectral regions B and D only the NH2 protons of protonated cytosines are expected to resonate. The spectral region B does not contain any cross peaks, i.e. no DNA NH2 resonances reside in this spectral region, demonstrating that none of the DNA residues are protonated. The cross peaks in the boxed regions f in C directly demonstrate the existence of NOE contacts between DNA H2'/2'' protons and RNA imino protons, providing direct evidence for the formation of a triple helix. The other boxed peaks are discussed in detail by Van Dongen et al. [143]. Reproduced with permission from Ref. [143].

$$\tau_s = \frac{3.842\pi\eta LR^2}{6kT}(1 + \delta_s)^{-1}$$

where

$$\delta_L = -0.0662 + 0.917(2R/L) - 0.05(2R/L)^2 \quad (60)$$

and

$$\delta_s = 1.119 \times 10^{-4} + 0.6884(2R/L) - 0.2019(2R/L)^2 \quad (61)$$

This equation holds for axial ratios,  $2 < L/R < 30$ . Other equations that correlate the cylinder length and radius to the time constants exist, for example the equation of Perrin and Broersma, but they only apply for large  $L/R$  ratios [190,191].

As was shown by Woessner [192–194] the correlation function for the vector connecting two nuclei or the long axis of the chemical shift tensor will then contain three exponentially decaying



terms:

$$G_0(t) = 0.125(3 \cos^2 \phi - 1)^2 \exp(-t/\tau_1) + \frac{3}{2} \cos^2 \phi \times \sin^2 \phi \exp(-t/\tau_2) + 0.375 \sin^4 \phi \exp(-t/\tau_3) \quad (62)$$

where  $\tau_1 = \tau_L$ ,  $\tau_2 = 6\tau_L\tau_s/(5\tau_s + \tau_L)$ ,  $\tau_3 = 3\tau_L\tau_s/(\tau_s + 2\tau_L)$ . This expression also includes an explicit dependence on the angle,  $\phi$ , which the dipole–dipole vector makes with the principal axis of rotation. The spectral density function will then consist of the following three terms:

$$J(\omega) = 0.25(3 \cos^2 \phi - 1)^2 J(\omega, \tau_1) + 3 \cos^2 \phi \sin^2(\phi) J(\omega, \tau_2) + 0.75 \sin^4(\phi) J(\omega, \tau_3) \quad (63)$$

Thus, a vector along the long axis is affected only by the end-over-end tumbling. On the other hand, for a vector perpendicular to the long axis the spectral density function is given by  $J(\omega) = 0.25J(\omega, \tau_L) + 0.75J(\omega, \tau_3)$ , so that  $J(\omega)$  approaches  $0.25J(\omega, \tau_L)$  at large  $L/R$  ratios. For sequential contacts the relaxation vectors are mainly found almost parallel to the long helix axis. In that case the correlation function  $G_0(t)$  contains only one relaxation time,  $\tau_L$ . In a double helical conformation the H6 to H5 vector lies approximately perpendicular to the long axis. The correlation function  $G_0(t)$  then contains two relaxation terms with time constants  $\tau_L$  and  $\tau_3$ , respectively. In case the vector lies somewhere inbetween these extremes three relaxation times are present in  $G_0(t)$ , namely,  $\tau_L$ ,  $\tau_2$  and  $\tau_3$ . In the particular case that the relaxation vector is at ‘the magic angle’  $G_0(t)$  does not contain the relaxation term with time constant  $\tau_L$ . Consequently, in this particular case the relaxation is determined by  $J(\omega, \tau_2)$  and  $J(\omega, \tau_3)$  and thus has a rather weak dependence on helix length. For close to symmetric molecules all three relaxation times are almost identical and the orientation of the relaxation vector is not very dependent on the angle it makes with the helix axis.

Internal motions will be superimposed on the overall motion, and as a result additional correlation times will show up in the correlation functions. Woessner et al. [194] have, for example, derived the time correlation function for internal motion superimposed

on anisotropic overall rotation of an axially symmetric molecule; in their derivation they assumed that internal rotation consists of free diffusion on a cone. The time correlation function can then contain up to nine relaxation times. Thus, generally speaking, quite complex functions are obtained. The approach is then to use a particular model for internal motion and to try to interpret the relaxation parameters in terms of these models. As pointed out by Lipari and Szabo [65] there is the danger of overinterpretation of limited data and the possibility that the resulting physical picture is not unique. Models cannot be proven; they can only be eliminated in favorable cases. Lipari and Szabo [65,66] have addressed this problem by invoking their model-free approach. They show that the information content of NMR relaxation data on internal motions can be completely described by (1) a *generalized order parameter*  $S$ , which is the measure of spatial restriction of the motion, (2) an *effective correlation time*  $\tau_e$ , which is a measure of the rate of motion. Lipari and Szabo assume that overall motion and internal motion are independent. The time correlation function can then be rigorously factored into two contributions for isotropic overall tumbling, namely, one from the overall motion,  $G_0(t)$ , corresponding to  $G(t)$  discussed above, and one contribution representing internal motions,  $G_1(t)$ :

$$G(t) = G_0(t) \cdot G_1(t) \quad (64)$$

As pointed out by Lipari and Szabo, anisotropic overall motion cannot rigorously be factored into a product of contributions due to overall and internal motion even when it is assumed that these motions are independent, but numerical evidence shows that the factoring provides a very good approximation. Thus, the above equation can also be used for internal motion superimposed on anisotropic overall motion. The second step is to describe the internal motion. Internal motion can rigorously be described by a sum of exponentially decaying terms:

$$G_1(t) = \sum_i a_i e^{-t/\tau_i} \quad (65)$$

This leads to a large number of parameters in the time correlation functions that cannot be determined from experimental NMR relaxation data. Lipari and Szabo

then try to determine the simplest possible description for internal motion. From the definition of  $G_1(t)$  it follows that  $G_1(0) = 0$ ; they define  $S$  as  $G_1(\infty)$  and  $\tau_c$  as the surface under the time correlation function, i.e.

$$\tau_c = \frac{\int_0^{\infty} (G_1(t) - S^2) dt}{1 - S^2} \quad (66)$$

They then assume that the simplest approximation to  $G_1(t)$ , which is exact at  $t = 0$  and  $t = \infty$ , has the form:

$$G_1(t) = S^2 + (1 - S^2) \exp(-t/\tau_c) \quad (67)$$

The overall correlation function then becomes, for the case of isotropic overall motion:

$$G(t) = \frac{1}{5} S^2 \exp(-t/\tau_c) + \frac{1}{5} (1 - S^2) \exp(-t/\tau) \quad (68)$$

with

$$\tau^{-1} = \tau_c^{-1} + \tau_e^{-1} \quad (69)$$

This leads to a spectral density function of the form

$$J(\omega) = \frac{2}{5} \left[ \frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} + \frac{(1 - S^2) \tau}{1 + (\omega \tau)^2} \right] \quad (70)$$

The spectral density function thus contains two fitting parameters for each vector since the overall correlation time  $\tau_c$  is expected to be the same for all nuclei in the molecule.

Thus, Lipari and Szabo invoke no particular model to describe the internal motion; the approach is therefore *model-free*.  $S$  is a *model-independent measure of the spatial restriction of the internal motion*. The effective correlation time is determined by the area under the correlation function for internal motion. In that sense it is a *model-independent quantity*. In summary, Lipari and Szabo have shown that this description of internal motion in terms of two parameters,  $\tau_c$  and  $S^2$ , suffices, i.e. such a description represents the total information content of NMR relaxation measurements. The subsequent question is whether these quantities can be used to describe physical models for internal motion. As pointed out by Lipari and Szabo, the quantities  $\tau_c$  and  $S^2$  cannot be uniquely interpreted, and their values are, in principle,

consistent with an infinite number of physical pictures of motion [66]. To be more specific, it is not possible to relate  $\tau_c$  to some physical aspects of the internal motion in a model-independent way; the connection between  $\tau_c$  and the microscopic diffusion constants or rates can only be established in the context of a model; moreover, the value of  $\tau_c$  depends on both the microscopic time constants and the spatial nature of the motion. In the special case of models with only the rate of diffusion as an adjustable parameter, one could derive from  $\tau_c$  a physically significant value for it. But for the usual, physically more relevant models which have more than one adjustable rate parameter for internal motion,  $\tau_c$  can be reproduced in an infinite number of ways. On the other hand, the quantity  $S^2$  can be used to disprove certain models. It is therefore of interest to consider  $S$  for a number of models.

In the special case that the internal motion is represented by free diffusion of the relaxation vector **on** a cone, we have

$$S = P_2[\cos(\theta)] \quad (71)$$

Here  $\theta$  is the angle between the symmetry axis of the cone and the relaxation vector connecting the nuclei. Thus, it is possible for  $S$  to vanish at the 'magic angle'. Such a model would be a good representation of the motion of a methyl group. In that case the angle  $\theta$  is known a priori and  $S^2 < 0.11$ , if the free diffusion model applies. For free diffusion of the vector **in** a cone,  $S$  is given by

$$S_{\text{cone}} = \frac{1}{2} \cos \theta_0 (1 + \cos \theta_0) \quad (72)$$

Here  $\theta_0$  represents the maximum angle that the relaxation vector can rotate away from the symmetry axis of the cone. As pointed out by Lipari and Szabo, any type of internal motion can be modeled as diffusion in a cone, irrespective of whether it makes sense physically. Restricted diffusion on a cone is given by

$$S^2 = [P_2(\cos \theta)]^2 + \frac{3 \sin^2 \theta \sin^2 \gamma}{\gamma^2} \times (\cos^2 \theta + \frac{1}{4} \sin^2 \beta \cos^2 \gamma) \quad (73)$$

Here  $\theta$  is the half angle of the cone, and the angular range is given by  $\pm \gamma$ . This model is more complex than free diffusion in a cone and is multi-interpretable.

Interestingly, Lipari and Szabo derived that for the case

of free diffusion in a cone superimposed on anisotropic motion, the order parameter is exactly given by

$$S^2 = S_{\text{cone}}^2 [P_2(\cos \theta_{\text{MD}})]^2 \quad (74)$$

where  $S_{\text{cone}}$  is given as above and  $\theta_{\text{MD}}$  represents the angle the vector makes with the long axis of the cylindrical molecule. It is important to note here that for the derivation of this equation for  $S$ , Lipari and Szabo have taken for the internal correlation function all relaxation times, as well as those resulting from rotation about the long axis, except for the one representing the tumbling of the long axis. This is of interest for nucleic acids, since it demonstrates that  $S = 0$  for a relaxation vector moving on a cone, which has its symmetry axis at the magic angle with respect to the long axis; thus, when tumbling of the long axis does not contribute to the relaxation.

It has been found for some residues in proteins that the experimental parameters cannot be fitted to the simple analytical expression (Eq. (67)) and additional degrees of freedom must be introduced i.e. an additional order parameter,  $S_f$ , and time constant,  $\tau_f$ . The analytical model introduced to take care of the inconsistencies employs the following internal correlation function:

$$G_1(t) = S^2 + (1 - S_f^2)e^{-t/\tau_f} + (S_f^2 - S^2)e^{-t/\tau_s} \quad (75)$$

For anisotropic overall motion with internal motion superimposed, the overall time correlation function,  $G(t)$ , follows from Eq. (64) with  $G_0(t)$  and  $G_1(t)$  given by Eqs. (62) and (67), respectively,

$$\begin{aligned} G(t) = & \frac{1}{2}S^2[0.25(3 \cos^2 \phi - 1)^2 \exp(-t/\tau_1) \\ & + 13 \cos^2 \phi \sin^2 \phi \exp(-t/\tau_2) + 0.75 \sin^4 \phi \\ & \times \exp(-t/\tau_3)] + \frac{1}{2}(1 - S^2)[0.25(3 \cos^2 \phi - 1)^2 \\ & \times \exp(-t/\tau) + 13 \cos^2 \phi \sin^2 \phi \exp(-t/\tau_{i2}) \\ & + 0.75 \sin^4 \phi \exp(-t/\tau_{i3})] \end{aligned} \quad (76)$$

with

$$\tau^{-1} = \tau_L^{-1} + \tau_c^{-1} \quad (77)$$

$$\tau_{i2}^{-1} = \tau_2^{-1} + \tau_c^{-1}$$

$$\tau_{i3}^{-1} = \tau_3^{-1} + \tau_c^{-1}$$

This leads to a spectral density function of the form

$$\begin{aligned} J(\omega) = & S^2[0.25(3 \cos^2 \phi - 1)^2 J(\omega, \tau_1) + 13 \cos^2(\phi) \sin^2 \\ & \times (\phi) J(\omega, \tau_2) + 0.75 \sin^4(\phi) J(\omega, \tau_3)] \\ & + (1 - S^2)[0.25(3 \cos^2 \phi - 1)^2 J(\omega, \tau) \\ & + 13 \cos^2(\phi) \sin^2(\phi) J(\omega, \tau_{i2}) + 0.75 \sin^4(\phi) \\ & \times J(\omega, \tau_{i3})] \end{aligned} \quad (78)$$

For anisotropic overall motion the complete time correlation function contains six time constants, even in its simplest form. In addition, the quantity  $S$  is present as a parameter. It is therefore of interest to consider some examples. If the ratio  $L/R$  is close to 2 (spherical) all three time constants for overall motion are quite similar. It is then appropriate to describe the overall motion by one exponential term decaying with an effective average relaxation time,  $\tau_{\text{av}}$ . This average value will now depend on the orientation of the relaxation vector with respect to the cylinder axis, and contrary to spherical molecules is not the same for all nuclei. The internal motion can be included by multiplying  $G_0(t)$  by  $G_1(t)$  and a simple description similar to that for isotropic tumbling is obtained. The order parameter is then representative of the spatial restriction for internal motion. When the ratio  $L/R \gg 1$ , the overall motion consists of a long relaxation time  $\tau_L$  and two much shorter relaxation times. Again it seems appropriate to use the simplest description, and let the time correlation function consist of a long correlation time,  $\tau_c$ , representing the rotation of the long axis ( $\tau_c = \tau_L$ ), and a shorter correlation time,  $\tau_e$ , which represents the time constants resulting from rotation about the long axis as well as the time for internal motion. The description then becomes again a two parameter description, as for isotropic motion. The order parameter  $S^2$  is in this case given by Lipari and Szabo (Eq. (74)). The quantity  $S^2$  now depends on the angle,  $\theta_{\text{MD}}$ , the relaxation vector makes with the long axis, and on  $S_{\text{cone}}^2$  representing the spatial restriction of internal motion. Thus,  $S^2$  is now not directly representative of the spatial restriction of internal motion;  $\tau_c$  ( $= \tau_L$ ) should be the same for all nuclei as in the case of isotropic overall motion. For intermediate  $R/L$  ratios a description in terms of an overall time constant,  $\tau_c$ , a time constant

for fast motions,  $\tau_e$ , and an order parameter  $S^2$  should suffice. However, in contrast to the case of isotropic motion, these parameters will now have a more complex interpretation and the orientation of the relaxation vector relative to the long axis has to be taken into account. In fact, Lipari and Szabo proposed using, for the case of anisotropic overall motion, two exponentially decaying terms to describe the overall motion:

$$G_0(t) = A \exp(-t/\tau_L) + (1 - A) \exp(-t/\tau_s) \quad (79)$$

where  $\tau_L$  is the time constant for end-over-end tumbling of the long axis rotation, and  $\tau_s$  represents the shorter time constant related to overall tumbling. Internal motion is then taken into account by multiplying  $G_0(t)$  by  $G_I(t)$ :

$$G(t) = S^2 [A \exp(-t/\tau_L) + (1 - A) \exp(-t/\tau_s)] \\ + (1 - S^2) [A \exp(-t/\tau) + (1 - A) \exp(-t/\tau_{es})] \quad (80)$$

with

$$\tau^{-1} = \tau_L^{-1} + \tau_e^{-1}$$

$$\tau_{es}^{-1} = \tau_s^{-1} + \tau_e^{-1}$$

This 'extended' Lipari and Szabo approach now represents a five parameter description,  $A$ ,  $S^2$ ,  $\tau_L$ ,  $\tau_s$  and  $\tau_e$ . It has the advantage over the two parameter description that  $\tau_L$  is an interpretable quantity that should be the same for all nuclei, and  $S^2$  is the parameter that represents the spatial restriction for internal motion. On the other hand the number of parameters has increased. The quantity  $A$  is related to the orientation of the relaxation vector with respect to the long axis; in its simplest interpretation  $A = P_2(\cos\theta_{MD})^2$ . The amplitude of the term containing  $\tau_L$ , which is usually interpreted as the order parameter, equals  $S^2 A = S^2 P_2(\cos\theta_{MD})^2$ . Note further that inclusion of  $\tau_s$  in the description will lead to a better reproduction of the value of  $\tau_L$ . This description has been used to interpret nucleic acid relaxation data (see below).

Lipari and Szabo have investigated the range of validity of their description. The general trend is to predict values of  $S^2$  that are too large and values of  $\tau_e$  that are too small; the effect being stronger for smaller values of  $S^2$  and larger values of  $\tau_e$ . For internal motions slow on the NMR time-scale, i.e.  $(\omega\tau_e)^2 > 1$ , they predict for values of  $S^2 > 0.5$  an error of 15% or

less, and for  $0.3 < S^2 < 0.5$  an error of the order of 30%. For  $(\omega\tau_e)^2 < 1$  the error is smaller. Secondly, as pointed out by Lipari and Szabo, if there is an overall motion common to all parts of the molecule there cannot be a component in the correlation function that decays more slowly than the overall motion. Consequently, internal motions on a time-scale slower than the overall tumbling are not detectable. Thus, the longest time constant is always that of the overall tumbling.

As pointed out, relaxation depends on whether the chemical structure to be investigated moves with frequencies corresponding to differences between energy levels. This distribution of frequencies is usually expressed with so-called power spectral density functions. For example, the commonly measured relaxation parameters  $T_1$ ,  $T_2$ , and the NOE effect depend on spectral density functions at five frequencies. In addition, the relaxation of longitudinal spin order, anti-phase coherences, and multiple quantum coherences can be measured. With the advent of proton-detected heteronuclear experiments, sensitive 2D experiments have been developed for the measurement of (heteronuclear) relaxation parameters (see, for example, Ref. [196] for a summary). Below we summarize the relations for the relaxation rates,  $R$ , of the most relevant types of coherences and spin order:

$$R_{IN}(I_z \rightarrow N_z) = d(-J(\omega_I - \omega_N) + 6J(\omega_I + \omega_N)) \quad (81)$$

$$R_{IN}(N_z \rightarrow 2I_z N_z) = KJ(\omega_N)$$

$$R_{IN}(N_x \rightarrow 2I_z N_x) = K(\frac{2}{3}J(0) + \frac{1}{2}J(\omega_N))$$

$$R_N(N_z) = d[J(\omega_I - \omega_N) + 3J(\omega_N) + 6J(\omega_I + \omega_N)] \\ + c[J(\omega_N)] \quad (82)$$

$$R_N(N_x) = \frac{d}{2}[4J(0) + J(\omega_I - \omega_N) + 6J(\omega_I) + 3J(\omega_N) \\ + 6J(\omega_I + \omega_N)] + c[\frac{2}{3}J(0) + \frac{1}{2}J(\omega_N)] + R_{ex}^N$$

$$R_I(I_z) = d[J(\omega_I - \omega_N) + 3J(\omega_I) + 6J(\omega_I + \omega_N)] + \rho_L$$

$$R_I(I_x) = \frac{d}{2}[4J(0) + J(\omega_I - \omega_N) + 6J(\omega_N) + 3J(\omega_I) \\ + 6J(\omega_I + \omega_N)] + \rho_T + R_{ex}^I$$

$$R_{IN}(2I_z N_x) = \frac{d}{2}(4J(0) + J(\omega_I - \omega_N) + 3J(\omega_N) + 6J(\omega_I + \omega_N)) + c\left(\frac{2}{3}J(0) + \frac{1}{2}J(\omega_N)\right) + \rho_L + R_{ex}^N \quad (83)$$

$$R_{IN}(2I_x N_z) = \frac{d}{2}(4J(0) + J(\omega_I - \omega_N) + 3J(\omega_I) + 6J(\omega_I + \omega_N)) + c(J(\omega_N)) + \rho_T + R_{ex}^I$$

$$R_{IN}(2I_z N_z) = d(3J(\omega_N) + 3J(\omega_I)) + c(J(\omega_N)) + \rho_L$$

$$R_{IN}(2I_1 N_{-1}) = R_{IN}(2I_{-1} N_1) = \frac{d}{2}(2J(\omega_I - \omega_N) + 3J(\omega_N) + 3J(\omega_I)) + c\left(\frac{2}{3}J(0) + \frac{1}{2}J(\omega_N)\right) + \rho_T + R_{ex}^{NI}$$

$$R_{IN}(2I_1 N_1) = R_{IN}(2I_{-1} N_{-1}) = \frac{d}{2}(3J(\omega_I) + 3J(\omega_N) + 12J(\omega_I + \omega_N)) + c\left(\frac{2}{3}J(0) + \frac{1}{2}J(\omega_N)\right) + \rho_T + R_{ex}^{NI}$$

$$R_{IN}(2I_x N_x) = \frac{d}{2}(J(\omega_I - \omega_N) + 3J(\omega_N) + 3J(\omega_I) + 6J(\omega_I + \omega_N)) + c\left(\frac{2}{3}J(0) + \frac{1}{2}J(\omega_N)\right) + \rho_T + R_{ex}^{NI}$$

$R_{ex}$  stands here for the exchange contribution to the relaxation of transverse coherences. The longitudinal and cross relaxation rates in a NOESY experiment,  $\rho_L^\dagger$  and  $\sigma_L^\dagger$ , and the longitudinal and cross relaxation rates in a ROESY experiment,  $\rho_T^\dagger$  and  $\sigma_T^\dagger$ , are given by:

$$\rho_L^j = \sum_{i=1}^N q[J_{ji}(0) + 3J_{ji}(\omega) + 6J_{ji}(2\omega)] \quad (84)$$

$$\rho_T^j = \sum_{i=1}^N q\left[\frac{5}{2}J_{ji}(0) + \frac{9}{2}J_{ji}(\omega) + 3J_{ji}(2\omega)\right]$$

$$\sigma_L^j = q[6J_{ji}(2\omega) - J_{ji}(0)]$$

$$\sigma_T^j = q[2J_{ji}(0) + 3J_{ji}(\omega)]$$

In Eqs. (82) and (83),  $d$  is the constant for

heteronuclear dipolar relaxation:

$$d = \frac{\left(\frac{h}{2\pi}\right)^2 \gamma_I^2 \gamma_N^2}{4r_{IN}^6} \quad (85)$$

The constant  $c$  for relaxation by chemical shift anisotropy,  $\Delta$  (ppm), is given by

$$c = \frac{\omega_N^2 \Delta_N^2}{3} \quad (86)$$

The constant  $K$  is given by

$$K = \frac{\left(\frac{h}{2\pi}\right) \gamma_I \gamma_N \omega_N \Delta_N \langle P_2(\cos(\Phi)) \rangle}{r_{IN}^3} \quad (87)$$

and the constant for homonuclear dipolar interaction is

$$q = \frac{\left(\frac{h}{2\pi}\right)^2 \gamma_I^4}{4r_{ij}^6} \quad (88)$$

The terms  $\gamma_N$  and  $\omega_N$  stand for the gyromagnetic ratio and the Larmor frequency of spin N;  $r_{IN}$  is the distance between spin I and N,  $h$  is Planck's constant.

Thus, NMR relaxation parameters are expressed in spectral density functions. In order to analyze these data one can proceed in two ways: (1) the spectral density function at the sampled frequencies can be derived exactly from a number of different relaxation parameters; this procedure, called spectral density mapping, is proposed by Peng and Wagner [197,198]. The commonly measured relaxation parameters  $T_1$ ,  $T_2$  and NOE depend on spectral density functions at five frequencies. Such data cannot therefore be interpreted without a known form of the spectral density function, or, in other words, an analytical form for the spectral density functions that depends on not more than three parameters. In the concept of spectral density mapping the approach has been abandoned. Instead, six experiments are selected or designed for measuring relaxation properties. The spectral density functions at the five frequencies can then be calculated from a linear combination of relaxation rates [197,198]. Thus, values of the spectral density function are obtained at a number of frequencies without reference to any model. Subsequently, these spectral density data can be analyzed in terms of the model-free approach of Lipari and

Szabo or in terms of specific physical models. (2) The second approach is to analyze the relaxation rates directly in terms of either the Lipari and Szabo model-free approach or in terms of explicit physical models.

The approach of spectral density mapping further clarifies the question central to the approach of Lipari and Szabo about the information content of NMR relaxation data. The best result that can be obtained from NMR relaxation measurements is a partial mapping of the spectral density function. The question can then be rephrased as: What is the information content of a spectral density function mapped at a number of frequencies? Since the spectral density function is simply the Fourier transform of the time correlation function, what is the information content of a time correlation function? From various tests (see, for example, Refs. [199–201]) on time correlation functions containing a number ( $> 3$ ) of exponentially decaying terms with random errors of 0.1–0.2%, it follows that at most three exponential terms can be derived, i.e. increasing the number of exponentially decaying terms does not statistically improve the fit [202]. A fit in terms of two exponentially decaying terms can show residuals of significant magnitude, and increasing the number of exponential terms to three leads to a statistically significant better fit, but a further increase in this number does not. Thus, any (multi-)exponentially decaying time correlation function can be described with high accuracy by at most three time constants and two amplitude parameters (assuming that the decay function is normalized to say 1.0). Consequently, any spectral density function can, with high accuracy, be described by the sum of three Lorentzians, i.e. by three time constants and two amplitude parameters. This is exactly the ‘extended’ Lipari and Szabo description, i.e. isotropic overall motion with two decaying functions for internal motions (Eqs. (57), (64) and (75)) or anisotropic overall motion described by two decaying terms and one term for internal motion (Eqs. (79) and (80)). A second aspect of the tests on (multi-)exponential decay curves is that a three-exponential fit gives correct values for the time constant and amplitude of the slowest decaying exponential. The accuracy is better the larger the amplitude. The time constants and amplitudes of the other exponential terms generally lie in between the actual values and cannot be related

to specific processes. Only in the case when the actual time correlation function contains two exponential terms does the shorter time represent the actual time constant. This very nicely compares with the conclusions of Lipari and Szabo mentioned above, namely, that the order parameters can reliably be determined (the accuracy increases with the magnitude of the order parameter) and interpreted; the time constant for overall motion can be determined, but the time constant for internal motion cannot, except for special cases where only one adjustable parameter is present. This leads to the conclusion that NMR relaxation data cannot provide time constants for internal motions except under special circumstances (see below). NMR relaxation data does provide a measure for the time constant of overall motion and can reliably provide a measure for the generalized order parameter, which is in fact the scaling of the contribution of overall motions. The order parameter can be interpreted as a measure for the spatial restriction of internal motions (but see above for the special case of anisotropic overall motion). This also means that any physical model for internal motion which contains as adjustable parameters two time constants and two amplitude parameters, can always be made to fit NMR relaxation data. Only in the special case that the physical model for internal motion contains as adjustable parameters **one time constant and one amplitude parameter** can one establish whether a particular physical model applies or not.

Recently, two papers have appeared that address two problems: firstly, how to derive anisotropic motion from NMR relaxation data, and secondly, how to detect and determine conformational exchange processes. In the first paper by Tjandra et al. [203] the NMR relaxation data of HIV protease is studied. They obtain  $^{15}\text{N}$  relaxation data ( $T_1$ ,  $T_2$  and NOE). NMR relaxation data have a limited information content. Nevertheless, from such data one can determine, for both isotropic as well as anisotropic motion, reasonably accurately the overall or end-over-end tumbling time for each relaxation vector individually. In the case of anisotropic motion additional time constants will show up in the spectral density functions, and reliable determination of, for example, the other time constant related to overall motion, seems at first glance impossible. However, there is an important restraint on these parameters. They are the same

for all residues in a molecule. Consequently, one can use the rather complex spectral density functions which contain the terms resulting from anisotropic tumbling and internal motion. In the analysis of Tjandra et al.  $T_1/T_2$  ratios are considered, since these are the least affected by internal motion; residues with internal motion are excluded from the analysis. In order to identify residues with internal motion two separate cases can be distinguished. Firstly, changes in  $J(\omega_N)$  resulting from internal motions that are faster than overall rotational diffusion but slower than  $\sim 100$  ps can significantly alter the NOE values. Residues with a low NOE value are excluded. Secondly, residues with conformational exchange on a microsecond to millisecond time-scale experience additional line broadening, commonly described by an exchange term  $R_{ex}$ . In this study a residue  $n$  is excluded when the following criterion applies:

$$(\langle T_2 \rangle - T_{2,n})/(\langle T_2 \rangle - (\langle T_1 \rangle - T_{1,n})/\langle T_1 \rangle) > 1.5 \times SD \quad (89)$$

Here, the average is taken over residues that have not been excluded because of a low NOE, and SD represents the standard deviation of these data points. Residues that exhibit slow conformational exchange will have a considerably different  $T_2$  but not  $T_1$ , thus failing to satisfy this equation. After this selection enough residues are left to determine the rotational diffusion tensor. The HIV protease has cylindrical symmetry. Consequently, its diffusion tensor contains two independent components, and two angles are required to attach the diffusion to the molecular frame, i.e. four independent parameters. The results show that these parameters can be determined well.

The next step in the analysis is to seek to determine parameters for internal motion. Previously when an isotropic motional model was used a number of residues were identified as showing conformational exchange. However, after the inclusion of anisotropic motion such residues were no longer identified. Given the values for the rotational diffusion tensor it is straightforward to determine  $S^2$  and  $\tau_e$ . This generally gives a good agreement between the values at 360 MHz and at 600 MHz, except for a small but significant systematic difference in  $S^2$  and  $\tau_e$ . Tjandra et al. [203] now take the ratio  $T_1^{360}/T_1^{600}$  and the ratio  $T_2^{360}/T_2^{600}$ . The systematically lower values of  $S^2$  at

360 MHz compared to those at 600 MHz are reflected in the  $T_1^{360}/T_1^{600}$  and  $T_2^{360}/T_2^{600}$  ratios, which are higher than expected for a rigid rotor. Tjandra et al. [203] propose that the discrepancy results from two separate causes: first, internal motions are not negligible when calculating the diffusion tensor from  $T_1/T_2$  ratios, and second, the chemical shift anisotropy (CSA) of  $^{15}\text{N}$  is, on average, about 10 ppm larger than the commonly used value of 160 ppm.

A second paper, by Phan et al. [204], describes a different approach to the analysis of NMR relaxation data. Again  $^{15}\text{N}$  relaxation data is used ( $T_1$ ,  $T_2$  and NOE), but here measurements are performed at three different field strengths. The analysis shows that the overall dimensions of the protein can be approximated by a cylinder with an axial ratio  $D_{\text{par}}/D_{\text{per}}$  of 1.9. The data can be fitted satisfactorily to a Lipari–Szabo model, taking anisotropy into account. A method is demonstrated for determining the exchange contribution. Instead of the equation used by Tjandra et al. [203] the exchange contribution is determined from  $1/T_1 (= R_1)$  and  $1/T_2 (= R_2)$  using the parameter  $(R_2 - R_1/2)$ :

$$\{R_2 - \frac{R_1}{2}\} = \frac{c^2 + d^2}{3}J(0) + \frac{d^2}{2}J(\omega_H) + R_{ex} \quad (90)$$

where  $c$  and  $d$  represent as before the constants for CSA and dipolar contribution to the relaxation. The term  $(R_2 - R_1/2)$  is subsequently plotted against the field strength. The first term on the right-hand side is field dependent, i.e.  $c \propto B_0$ , as is the term involving  $R_{ex}$ . The term involving  $J(\omega_H)$  also depends on the field but may be neglected due to its smallness. One then obtains (with  $c_1 = c/B_0$  and  $A = R_{ex}/B_0^2$ ):

$$\{R_2 - \frac{R_1}{2}\} \approx (\frac{c_1^2}{3}J(0) + A)B_0^2 + \frac{d^2}{3}J(0) \quad (91)$$

From a plot of  $(R_2 - R_1/2)$  against  $B_0^2$  one can deduce from the intercept  $I_0 = d^2J(0)/3$  the value of  $J(0)$  and from the slope,  $m$ , the exchange constant  $A$ :

$$A = m - \frac{c_1^2 I_0}{d^2} \quad (92)$$

Thus, any spin which has a slope  $m > (c_1^2 I_0/d^2)$  will have an exchange contribution  $AB_0^2$ . Ultimately six residues were identified as having an exchange contribution. However, the condition  $A = 0$  for the other

residues was also not strictly adhered to. For 45 residues (56%) a straight line with a positive slope could be fitted to these data. Practically, the analysis was conducted in detail for 31 residues for which the slopes differed by less than 40%, on going from 50.6 MHz to 60.8 MHz and on going from 50.6 to 76.0 MHz.

It is interesting to note here that Phan et al. [204] do not observe the systematic deviation due to CSA as found by Tjandra et al. [203] in their analysis. In the analysis of Phan et al. [204] this deviation should show up as a slope  $m$  being systematically larger than  $c^2I_0/d^2$  by about 7% (10 ppm/160 ppm). Such a difference is not detectable in the method used by Phan et al. [204] because of the rather large variations in the slopes observed (< 40% for the 31 residues analyzed in detail). This demonstrates the difficulties in the analysis and in determining how errors may propagate.

Relaxation studies on nucleic acids are not new and pioneering work has been done on larger DNA fragments (100 bps) over the last 20 years [205–208]. Studies on short synthetic oligonucleotide sequences have also been performed, but have been limited by the lack of spectral dispersion and the necessity to work at natural isotopic abundance. Only recently have  $^{13}\text{C}$  labeled RNAs become available, opening up the way for more detailed relaxation studies of well-defined sequences. The amount of detailed experimental relaxation data is thus limited (for recent reviews see Refs. [18,195], and Ref. [209], which describes solid state NMR relaxation studies). We will briefly discuss here some specific aspects.

Boxer and co-workers have performed detailed  $^1\text{H}$  and  $^{31}\text{P}$  relaxation studies on a series of oligonucleotides varying in length from 6 to 20 base pairs [210,211]. The overall tumbling time derived from NMR data gave a lower value than derived from dynamic light scattering (DLS). Agreement could be made by assuming an order parameter  $S^2$  of 0.8.

Relaxation studies involving  $^{13}\text{C}$  of short synthetic oligonucleotides are scarce. Williamson and Boxer [210] studied a hairpin molecule with a loop thymidine C6  $^{13}\text{C}$ -enriched to measure the loop mobility. The apparent correlation times for the thymidine C6 in the loop were shorter than those for the cytosine H6–H5 cross-relaxation indicating mobility in the loop. This was confirmed by the rather low order

parameter for the C6–H6 vectors (between 0.6 and 0.7). In a multinuclear NMR study on a DNA octamer it was found that for both the C–H as well as the H–H vectors in the ribose ring the order parameters were smallest at the termini and largest at the center; the bases showed the least evidence of mobility [212]. Borer et al. [213] have reached the same conclusion based on  $^{13}\text{C}$  NMR relaxation studies at three different field strengths. Order parameters calculated from long molecular dynamics simulations also show this same trend [214]. A detailed  $^{13}\text{C}$  relaxation study has recently also been performed by Gaudin et al. [215] on a DNA dodecamer, which has the central three thymines  $^{13}\text{C}$ -labeled at the C1'-position. The relaxation rates  $R(\text{Cz})$ ,  $R(\text{Hz})$ ,  $R(\text{Hz} \rightarrow \text{Cz})$ ,  $R(2\text{HzCz})$ ,  $R(2\text{HzCxy})$  related to the H1'–C1' vector were measured. Instead of deriving spectral density functions and/or order parameters and relaxation time constants from these data, the rates were directly compared with rates calculated for two models of internal motion. A good fit with experimental data was reached for a two-state jump model between  $\chi\text{-syn}$  and  $\chi\text{-anti}$  conformations with  $P(\text{syn})/P(\text{anti}) = 9/91$ , or for a restricted rotation model with  $\Delta\chi = 28^\circ$ , and an internal diffusion coefficient of  $30 \times 10^7 \text{ s}^{-1}$ . King et al. [18] present relaxation data for  $^{13}\text{C}$ -enriched  $\Delta\text{TAR}$  RNA which has a trinucleotide bulge and a six-base loop. Here it is also observed that the ribose nuclei have smaller order parameters than the base nuclei. Furthermore, the order parameters of the C1' and C6/C8 sites vary widely (0.44 to 1.0) with a reasonable correlation between secondary structure and mobility. Interestingly, the loop residues have relatively high order parameters indicating a rather well structured loop.

In the analysis of the relaxation data discussed above both the Lipari and Szabo model-free approach and more complex physical models have been used. It is interesting to note that according to the analysis of Lipari and Szabo the information content of NMR relaxation measurements is such that apart from the overall time constant, two parameters are sufficient to describe the relaxation data. We later argued, using simulation of exponentially decaying time correlation functions, that in the case of highly precise data two extra parameters may be added, i.e. a description in terms of an overall time constant, two smaller time constants and two amplitude parameters may be



possible (this is the extended Lipari and Szabo description). Hence, *any physical model for internal motion superimposed on overall motion that contains as adjustable parameters three time constants and two amplitude parameters can always be made to fit NMR relaxation data. Only in the special case that the physical model contains a smaller number of adjustable parameters, for example, for internal motion contains one time constant and one amplitude parameter, can one establish whether a particular physical model applies or not.* The latter condition is generally physically quite unrealistic for nucleic acids, where often much more complex motions occur, each of which are not well known. One is forced to conclude that for the interpretation of NMR relaxation data one should not go beyond determining and interpreting the order parameter. The generalized order parameter  $S^2$  has for nucleic acids (when anisotropic motion occurs) a more complex interpretation than for proteins (assuming isotropic overall motion); it may not even directly reflect internal motion. For example, the C6–H6 vector lies in a helical region almost perpendicular to the long axis, but in bulge regions this may not be the case. This affects both the effective overall relaxation time and the order parameter. In particular when the vector lies at the ‘magic angle’ with respect to the helix axis, shorter time constants govern the relaxation even in the absence of internal motion. The complexity which results from the presence of anisotropic motion in nucleic acids represents both a challenge and an opportunity. In particular, the dependence of the relaxation data, i.e.  $S^2$ , on the orientation of the dipolar relaxation vector with respect to the overall axis, enables the use of this type of data in structure determination. The important characteristic of such data is that it represents global structural knowledge, which is in contrast to NOE and  $J$ -coupling data, which is strictly local.

## 9. Calculation of structures

The different methods that have been used over the years to determine structures from NMR data, have nowadays more or less converged into one consensus approach, although the area is very active and new methods are regularly developed. This present

consensus on how NMR derived structures should be obtained and how the data should be presented is reflected in a set of IUPAC recommendations [216]. These recommendations form a very useful overview and provide a checklist when performing structure calculations. The consensus approach currently adhered to is to calculate the final set of structures using a simulated annealing (SA) protocol, with X-plor [78] being one of the most popular programs used for this purpose. This is also the method we have mainly used in our calculations. The SA method should, however, not be used as a black box, since it does not allow one to screen with complete certainty the full conformational space consistent with the experimental data. It is thus important to ensure firstly that methods are used that search conformation space as much as possible, and secondly, that the ultimately derived structures are validated. In the following we will discuss a structure calculation protocol and aspects concerning the validation of the derived structure or set of structures.

The consensus structure calculation protocol used by us [12,77,139,217] consists of the following steps. The procedure starts out with generating extended random structures, in order to ensure that conformation space is searched as fully as possible. Two options can be chosen to derive the initial structures from these extended structures: (1) A distance geometry type of approach can be applied; such methods are less computationally demanding and consequently allow the derivation of a set with a large number of initial structures (up to 500) [12,217]. (2) Starting from extended randomized structures, an SA folding protocol is used to calculate the initial structures [77,139]. (3) The best structures, i.e. the converged structures, can subsequently be subjected to rounds of refinement using an SA refinement protocol.

The SA folding protocol we use [77] has similarities to the protocols as described by Santa Lucia and Turner [218] and Varani et al. [36]. It was applied to a two-strand RNA duplex with sequence 5'GGGCUGAAGCCU3', in which a tandem G.A mismatched base pair was formed (see below). The molecule is folded from the extended structures via a number of successive SA runs. In the first and second round the global fold is determined by adding successively more and tighter distance and torsion angle constraints. A high level of convergence was obtained



Fig. 33. Final set of ten best structures of UGAA (see text): (a) structures were fitted on the whole molecule; (b) structures were fitted on the UGAA part of the molecule; (c) structures were fitted on the bottom stem of the molecule.

(75 out of 80). In the final round, i.e. the SA refinement protocol, all constraints are used and the upper and lower bounds set to their final values [77]. Fig. 33 shows an overlay of the final structures. The high number of constraints per residue ( $> 30$ ) leads to a defined structure (local superposition of the central four residues yields and rmsd of  $0.6 \text{ \AA}$ ). The global rmsd gives a value of  $1.25 \text{ \AA}$ . Due to the linear extended nature of helices and the local nature of NMR constraints, the NMR derived structures are well defined locally, but not globally (see Fig. 33). When presenting nucleic acid structures one should therefore show a number of locally superimposed structures (see Fig. 33), instead of one set of structures. We have observed that the most difficult step in the SA protocol is to obtain the global fold. Applying the SA protocol to an intra-molecular triple helix via SA from extended chains leads to very low levels of convergence [139]. Recently, Stein et al. have proposed a method called torsion angle dynamics [219]. This method is less computationally demanding, since the chemical structure is maintained during the dynamics. We have applied this method and observed a considerably improved convergence rate (around 50%), when starting from random extended structures [139].

In the final stage the derived structures have to be analyzed and validated. The usual approach is to present an overview of the structure calculation statistics (see, e.g. Refs. [12,77,139,217]). This overview should contain a description of the constraints, i.e.

NOE or distance constraints, dihedral, hydrogen bond constraints, etc. The NOE or distance constraints should be divided, according to the type of constraint, into intra-residue and inter-residue constraints. In view of the discussion concerning distances and their effect on the accuracy of the derived structure in Section 4, it seems advisable to further divide these two categories, for example, according to the division made in Table 1. For example, constraints involving imino and amino protons indicate base pairing. The usual approach is to base the structure calculation on the experimental data, i.e. the derived structure or set of structures should follow from the experimental data and not be biased by uncertain force field effects. Therefore, the constraint set should contain the experimental constraints and the holonomic constraints, i.e. bond, bond angles, dihedrals, impropers, and hard-sphere van der Waals constraints. The experimental constraints (distance and torsion angle) are usually given as upper and lower bounds, with a quadratic increase in penalty outside the allowed range. The distance between upper and lower bound is the estimated error in the constraint value. To check whether the experimental data is consistent with the derived structure, the overview should therefore give the number of violations of experimental constraints (distance, dihedral angle, etc.), and/or the energies, divided according to the categories bond, bond angles, dihedral, restraint dihedrals, impropers, van der Waals (hard sphere), and restrained distances (plus other categories when applied). The energies provide an

indication of whether the structure is strained (to an unusual extent) by the experimental data. The rmsd of the set of structures that is ultimately derived therefore reflects the uncertainty in the derived structure, as it results from the uncertainty in, or lack of, the experimental data.

We note that the precision of the derived structures, as reflected in the rmsd values of the family of structures, is determined by three main aspects (see also discussion in Sections 4.1 and 4.4): (1) the precision of the derived distance constraints, or for that matter of the torsion angle constraints, is not the major contributor to highly precise structures, but see Section 4.4 for a detailed overview of the type of constraint lists that can be used and the consequences for the precision of the family of structures; (2) the major contributor to highly precise structures is the spread and the number of experimental distance constraints; the spread should be as even as possible throughout the chemical structure, and the number should be at least 10 per residue, but highly precise structures require significantly larger numbers (see Section 4.4 for a detailed description); (3) torsion angle constraints are not the major contributor to highly precise structures; only for already well defined structures will their inclusion lead to significant further improvement in precision, as suggested by the systematic study performed by Allain and Varani [79].

An approach to further test the reliability of the structure calculations is the method of cross validation suggested by Brünger et al. [220]. Cross validation is a statistical method which involves randomly removing about 10% of the constraints from the total list of experimental constraints, and subsequently deriving the structures based on these different subsets of constraint data. The spread in the derived structures then gives an indication of the reliability of the derived structure or set of structures. However, this approach can present a problem for structures derived from NMR data, because important structural features often depend heavily on a few characteristic NOEs. If this is the case, cross validation is not a viable method, since removing these characteristic NOE data would lead to obviously wrong structures, or to a set of structures with too much spread.

Because important structural features often depend on a few characteristic NOE contacts, and the precision depends heavily on the number of NOE-based

constraints, it is very important to make use of all possible NOEs and at the same time to carefully check the constraints against the experimental data. A very valuable approach in this respect is back-calculation of the NOE data. This should be done during each of the various rounds of structure refinement to ensure that all the cross peaks have been correctly interpreted; this includes a check on whether short distances in the calculated structures are indeed absent in the NOE spectra, otherwise lower bounds should be introduced.

A second type of check concerns the geometry of the derived structures. For proteins various programs exist that check the geometry of the derived structures. No such programs are in use for nucleic acids. This is unfortunate, but understandable, since the structure of nucleic acids is determined by a much greater number of torsion angles; the nucleic acid backbone is determined by six variable torsion angles, while for proteins this number is only two. Nevertheless, certain restrictions apply to these torsion angles, i.e. the  $\epsilon$  torsion angle is sterically restricted to approximately 150–280°, and the  $\delta$  torsion angle, related to the sugar ring puckering, lies between about 80° and 145°. The other torsion angles have usual ranges for A- and B-type helices, but may in other cases be found in any of their three possible rotamer states. For a complete overview of the sterically allowed ranges we refer to Ref. [88], and Sections 4 and 5 of this review. Another very important aspect of correct geometry concerns the stereospecificity; the most salient here is the correct placement of H2'' and H2' in DNA, of H2' and OH2'' in RNA, and of H5' and H5'' in both DNA and RNA. We have found during the course of our study on the proton chemical shift calculations (see Section 6 and Ref. [50]), that in quite a number of deposited PDB files the correct stereospecific geometry is violated. The H5' and H5'' protons seem to be placed almost at random, i.e. a high percentage have the wrong geometry, but some H2' and H2'' protons, and sometimes in RNA the H2', were placed in the wrong position. The latter errors are of the utmost relevance, because these protons provide essential experimental distance constraints in the structure calculations. It is therefore important that calculated structures are checked against correct geometry. A program which checks correct stereospecificity is available from the authors

on request [50]. Similar findings were recently published by Schultze and Feigon [221].

Finally, we note that the application of distance constraints in restrained molecular dynamics or in SA protocols is strictly speaking only allowed when the structure under study is rigid. When internal dynamics are present, the SA protocols, as discussed, cannot be applied. One alternative is to use time averaged constraints [222,223]. However, in the molecular dynamic runs only fast internal motions tend to be visible. These are in turn often related to libration motions, i.e. motions within a certain rotamer domain. For this type of motion the distances fluctuate around a middle value, so that one can use one average distance, and averaging would not be required. For situations such as the sugar ring flips between N- and S-states, and torsion angle flips between different rotamer states, distances can be small in one case and large in the other. In order to account for these differences in distances, averaging techniques have to be applied. Torsion angle flips and/or sugar ring flips do not occur often enough in the usual molecular dynamic runs, so that averaging is insufficient and the approach of time averaged constraints is not a good one. One either has to extend the dynamics period or apply ensemble averaging. Both approaches present enormous computational problems. Further developments in this area are awaited.

## 10. Prospects for larger systems

In the field of protein NMR, three-dimensional structure determination can nowadays be successfully performed on proteins with a molecular weight up to 30 kDa. Most NMR studies of nucleic acids, including those using labeled RNA nucleotides, involve systems of limited size, up to 40 nucleotides, which corresponds to a considerably smaller molecular weight of up to 13 kDa.  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling of RNA is still a fairly recent development. It is therefore not surprising that the initial studies have been limited to smaller systems. Although performed on smaller systems, these studies do provide good indications as to what the bottlenecks are for larger systems. It seems that the two main aspects that need to be considered are the  $T_2$  relaxation times (or line widths), and the resonance overlap.

The factor that mainly determines the  $T_2$  relaxation rates is the tumbling time of the molecule studied (see Section 8). We have calculated the tumbling time as a function of the number of base pairs for a number of situations that are likely to be encountered when studying nucleic acids, namely, for a spherical molecule (radius  $r$ ), and for cylindrical molecules, which have either the usual helix radius or double the radius (Fig. 33). Since the helix radius is about 11 Å, short nucleotide sequences (up to 10 base pairs), such as encountered in hairpins studied by NMR, will have a more or less spherical shape. Longer sequences will tend to fold into helices of cylindrical shape. For sequences containing more than 30 to 50 base pairs in a row, a turn may occur, leading to two helices folded side by side (see, for example, the X-ray

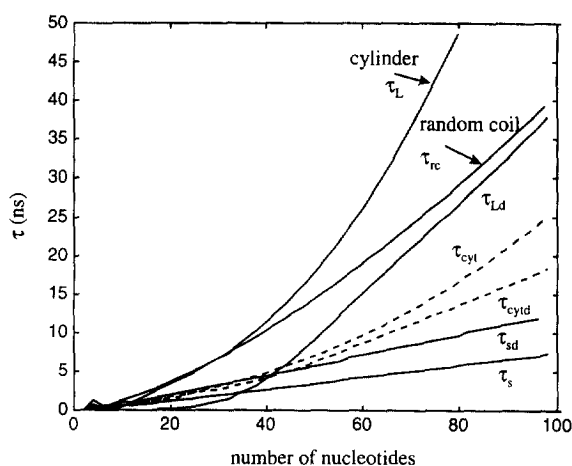


Fig. 34. The rotation times of differently shaped nucleic acid molecules as a function of the number of nucleotides. A cylindrically shaped and a randomly coiled molecule are considered. The rotation time for the end-over-end tumbling of a cylinder,  $\tau_L$ , and for rotation about the long cylinder axis,  $\tau_s$ , are calculated according to Eq. (59) in Section 8 ( $kT = 4.0 \times 10^{-21}$  J;  $\eta = 0.001$  poise). For the calculations it is assumed that the nucleotides form base pairs which fold as an A-type double helix; the length of the cylinder,  $L$ , then equals the number of nucleotides times 1.35. The cylinder radius is taken to be 11 Å. The rotation times,  $\tau_{Ld}$  and  $\tau_{sd}$ , represent end-over-end tumbling and rotation about the long axis, respectively, of a cylinder of length  $L/2$  and radius 22 Å. These rotation times correspond to two helices stacked side by side. For the calculation of a randomly coiled molecule,  $\tau_{rc}$  is taken to be proportional to  $L^{1.5}$  and scaled in such a way that it equals  $\tau_L$  at a length corresponding to 17 base pairs. The rotation times  $\tau_{cyt}$  and  $\tau_{cyld}$  correspond to  $\tau_3$  of Eq. (62) in Section 8, i.e.  $\tau_{cyt} = 3\tau_L\tau_s/(\tau_s + 2\tau_L)$  and  $\tau_{cyld} = 3\tau_{Ld}\tau_{sd}/(\tau_{sd} + 2\tau_{Ld})$ . They represent the shorter rotation time of a relaxation vector perpendicular to the helix axis.

structure of part of the Tetrahymena group I intron [4,5]). The time constant for tumbling of a spherical molecule,  $\tau_{\text{sphere}}$ , increases steeply with the third power of the diameter,  $D$ , with  $D$  taken to be equal to the number of base pairs ( $nb$ ) multiplied by 2.7 Å, so that  $\tau_{\text{sphere}} \propto (nb)^3 \propto D^3$ . Similarly, the time constant for end-over-end tumbling of a cylinder,  $\tau_L$ , increases with the third power of the length,  $L$ , and thus with the number of base pairs, i.e.  $\tau_L \propto (nb)^3 \propto L^3$ . On the other hand, the time constant for tumbling about the long axis is smaller and increases only linearly with length,  $\tau_L \propto (nb) \propto L$ . Included in Fig. 34 are also the tumbling times,  $\tau_{Ld}$  and  $\tau_{sd}$ , for the case where the helix has twice the usual diameter. As expected  $\tau_{Ld}$  is considerably smaller than  $\tau_L$ , and  $\tau_{sd}$  is larger than  $\tau_s$ . We also show  $\tau_{\text{cyl}} = 1/2 (\tau_L + \tau_s)$ , and  $\tau_{\text{cyl}d} = 1/2 (\tau_{Ld} + \tau_{sd})$ . These values should correspond closely to the effective rotation times governing H5 relaxation [195]. We find that for a system of 20 base pairs (40 nucleotides, 12 kDa) the tumbling time,  $\tau_{\text{cyl}}$ , is of the order of 5 ns, while for side by side helices each 20 base pairs in length (80 nucleotides; 24 kDa), the tumbling time,  $\tau_{\text{cyl}d}$ , is about 13 ns. Taking  $\tau_L$  and  $\tau_{Ld}$  as yardsticks we find 12 ns (12 kDa) and 26 ns (24 kDa), respectively. The former values compare quite closely with the tumbling times of proteins, where 3.4 ns is estimated for an 8 kDa protein (eglin c [198,224], and 10.2 ns for a protein of 22 kDa [203]. Thus, the tumbling times for nucleic acids are 1.0 to 2.0 times larger than for proteins of the same molecular weight.

In Fig. 35(A)–(D), estimates are given of transverse relaxation rates as a function of the number of nucleotides for four different situations. In Fig. 35(A) non-exchangeable protons directly bonded to  $^{13}\text{C}$  are considered. The figure shows the transverse rate resulting from dipolar interaction between  $^{13}\text{C}$  and  $^1\text{H}$ . Thus, the curves represent the relaxation rates of in-phase Cx coherence with the  $^{13}\text{C}$  nucleus attached to a proton. In the inset the relaxation rates are shown that ensue for HzCx and Hx coherences. Fig. 35(B) shows transverse relaxation rates of non-exchangeable protons not attached to a  $^{13}\text{C}$  nucleus (or directly bonded to  $^{12}\text{C}$ ), and in Fig. 35(C) the effect on the transverse relaxation rates of deuteration of surrounding protons is examined. Finally, in Fig. 35(D) the transverse relaxation of imino protons, directly bound to  $^{15}\text{N}$ , is shown with and without deuteration

of surrounding protons. The effect of deuteration is simulated by removing short distances from the proton relaxation pathways (see legend). Internal motion, which reduces the relaxation rates by a factor of 0.8, is assumed to be present in some cases.

In Section 7 we presented the transfer functions for a variety of through-bond NMR experiments assuming two different  $T_2$  relaxation times. Setting a somewhat arbitrary limit for breakdown of these experiments at transfer efficiencies ranging from 0.1 to 1%, we arrive at limiting  $T_2$  relaxation rates or line widths of the order of 20–30 Hz, indicated in Fig. 35(A)–(D) by the horizontal broken lines. The way in which through-bond experiments are affected by transverse relaxation depends on the type of coherences present during the longest transfer steps. We consider the experiments in that order. The first set of through-bond experiments encompasses HCN, HCNCH, PCH and HCP. These experiments all contain long transfer steps of the type,  $\text{Cx} \rightarrow \text{CyNz}$  or  $\text{Cx} \rightarrow \text{CyPz}$ , i.e. the transfer steps involve transfer of in-phase transverse coherence of a  $^{13}\text{C}$  nucleus directly attached to a proton. The sensitivity of these experiments will therefore be determined by the relaxation rates of these in-phase transverse  $^{13}\text{C}$  coherences, the rates of which are completely determined by the dipolar interaction of the  $^{13}\text{C}$  nucleus with its directly bonded proton and can be read off from Fig. 35(A). As can be seen from Fig. 35(A), these through-bond experiments will start to fail at rather short helix lengths of around 25 bp's corresponding to sequences of around 50 nucleotides (15 kDa). The range of these experiments can be extended tremendously by means of selective deuteration. The relaxation rate of in-phase transverse coherences of  $^{13}\text{C}$  not bonded to a proton is a factor of over a 100 smaller than for  $^{13}\text{C}$  bonded to a proton. This means that  $^{13}\text{C} \rightarrow ^{15}\text{N}$  or  $^{13}\text{C} \rightarrow ^{31}\text{P}$  transfers can remain effective in specifically deuterated compounds (possibly up to 200 nucleotides). For example, the HCCP–TOCSY experiment on a deuterated RNA, except at the H1' position, will remain a sensitive experiment even for long sequences. The second set of through-bond experiments consists of HCNCH, HPCH, long-range (H,N) HSQC, and long-range (H,C) HSQC. They all contain as the main transfer step either  $^{15}\text{Nz} \rightarrow ^1\text{HxNz}$ ,  $^{31}\text{Pz} \rightarrow ^1\text{HxPz}$ , or  $^{13}\text{Cz} \rightarrow ^1\text{HxCz}$ , where the  $^1\text{H}$  is attached to a  $^{13}\text{C}$  nucleus. Here

one can take the relaxation rates of  $H_x$  as a measure to derive the sensitivity of these transfers. As shown in the inset of Fig. 35(A) these experiments will start to fail for even shorter lengths (20–40 nucleotides). The life span of such experiments can be improved by removing the  $^{13}\text{C}$  nucleus from the  $^1\text{H}$  relaxation pathway. When the  $^{13}\text{C}$  labeling is absent the proton relaxation rates are as shown in Fig. 35(B). As can be seen, the experiments can be performed up to 50 to 70 nucleotides. This range can be extended further, up to 90 to 140 nucleotides, by deuteration of the surrounding protons, as shown in Fig. 35(C). Thus, deuteration and removal of  $^{13}\text{C}$  labeling can extend the range of these experiments to sequences which are over 100 nucleotides long. A third set of through-bond experiments involves those aimed at assignment

of exchanging protons. Here, the main transfer step is  $^{15}\text{N}x \rightarrow ^{13}\text{C}z\text{N}x$ , with the  $^{15}\text{N}$  nucleus directly bonded to a proton and no proton directly attached to the  $^{13}\text{C}$  nucleus. In this case one has to consider the relaxation of in-phase transverse  $^{15}\text{N}$  coherences, shown in the inset of Fig. 35(D). The experiment will only start to fail for much longer sequences (around 100 to 150 nucleotides). In summary, the range of the through-bond experiments can be extended to sequences of over a 100 nucleotides by judiciously applied labeling techniques. In other word, selective deuteration in combination with selective  $^{13}\text{C}$  labeling can make most through-bond experiments, albeit in modified form, sensitive enough even for sequences with more than 100 nucleotides. Finally, we also need to consider, apart from the through-bond experiment,

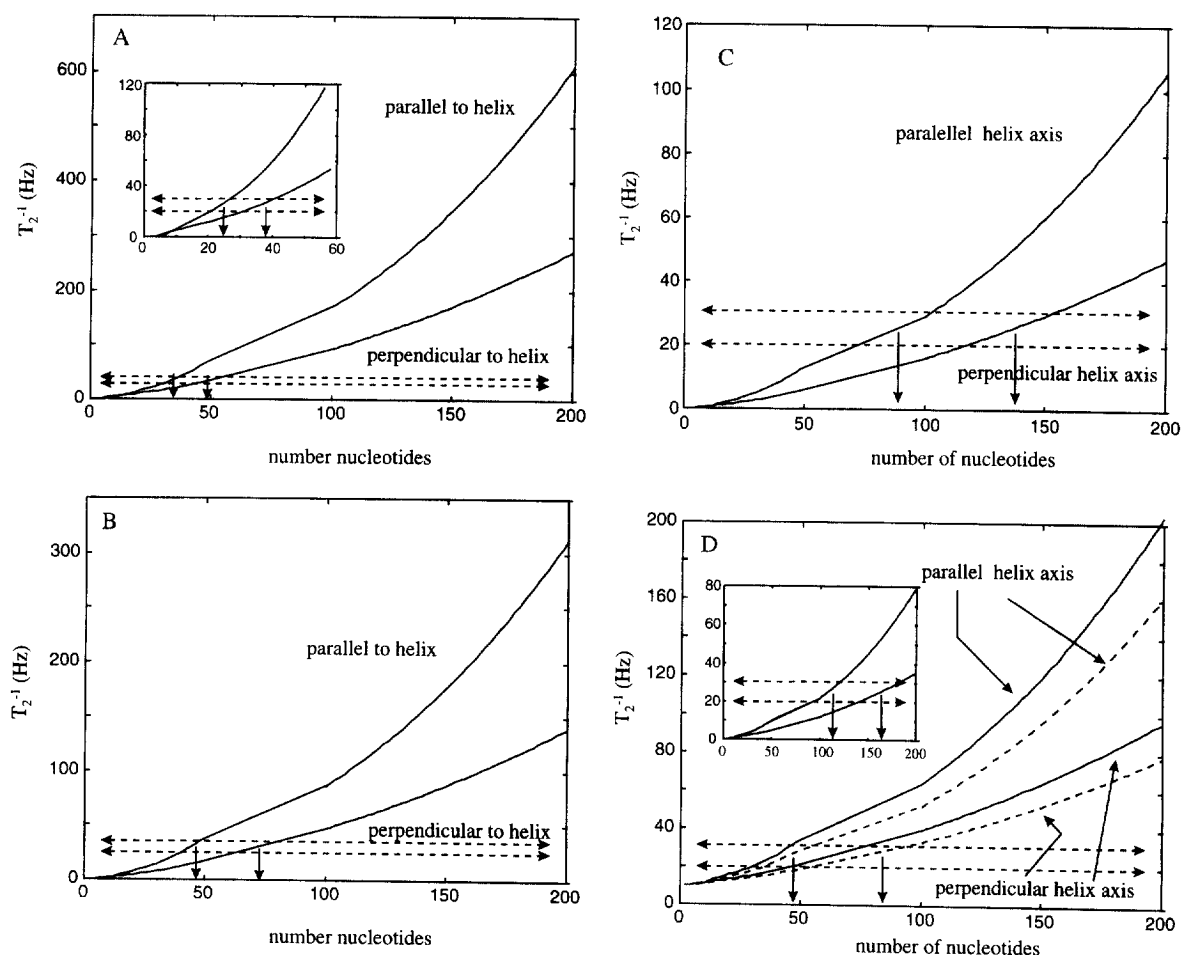


Fig. 35.

the multi-dimensional  $^{13}\text{C}$  and/or  $^{15}\text{N}$  edited NOE experiments, since these form an important part of the assignment and the structural characterization. Multi-dimensional  $^{13}\text{C}$  and/or  $^{15}\text{N}$  edited NOE experiments contain only short transfer steps (3 ms), and are thus expected to start to fail at line widths greater than, say, around 100 Hz, which corresponds to a limit of 100 nucleotides (Fig. 35(A)). NOE information can also be obtained from samples that are partially deuterated instead of  $^{13}\text{C}$  labeled. In this case the upper size limit is difficult to give, but will be above 100 nucleotides, as follows from Fig. 35(A) and (B). These lengths are all considerably beyond those achieved in practice nowadays. It can be concluded from these data that  $T_2$  relaxation times are not the limiting factor for extending the size of RNAs or DNAs studied today.

The other factor that has to be considered is the resonance overlap. As pointed out before, the introduction of  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling does considerably reduce overlap. A prime example is that in RNA the H2' to H5'/5'' resonances overlap completely, while the (C2',C3'), C4' and C5' resonances reside in separate regions. Unfortunately, these  $^{13}\text{C}$  spectral regions are narrow and within these regions rather strong overlap may occur. It seems that in particular the

C2' resonances reside in a region of about 1 to 2 ppm. The C4' resonances do spread out if they are in a non-helical environment, but in a helical environment the spread is also approximately 2 ppm. The H1' and C1' resonances do have a reasonable dispersion. The imino  $^{15}\text{N}$  resonances have limited dispersion, and the  $^{31}\text{P}$  resonances, in a helix, are notorious for their small dispersion. The size limit is thus mainly determined by resonance overlap even in the era of  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling. Today, the size limit lies around 40 nucleotides, but even for this size residue-type-specific labeling is required (see, e.g. Ref. [225]). In the case of uniform labeling, a rough estimate of the size limit is 30 nucleotides. Using this value as a yardstick it is possible to hypothesize on the prospects for larger systems, and on the amount of selective labeling required by the following reasoning. Assume a sequence of length 120, which contains 30 A residues, 30 G residues, 30 C residues and 30 U residues. With 30 nucleotides as the limit for assignment using uniform labeling, it follows that assignment should be possible, when only the 30 A residues are labeled. The same applies in the case of residue-type-specific labeling of the G, C and U residues. Thus, with residue-type-specific labeling and applying through-bond

Fig. 35. Transverse relaxation rates versus the number of nucleotides involving non-exchanging protons (A, B and C) and imino protons (D). The rotation times for end-over-end tumbling and tumbling about the helix axis have been calculated using eqns (59) and (60) (Section 8). The helix is assumed to be of the A-type. When the number of nucleotides is less than or equal to 50, the length of the helix is taken to be equal to half the number of nucleotides times  $2.7 \text{ \AA}$ , and the helix radius is assumed to be  $11 \text{ \AA}$ . When the number of nucleotides is greater than or equal to 100, the length of the helix is taken to be equal to one-quarter the number of nucleotides times  $2.7 \text{ \AA}$ , and the helix radius is assumed to be  $22 \text{ \AA}$ . It is thus assumed that when the number of nucleotides exceeds 100, two side-by-side helices are formed (this is to represent for example the case of a group I intron RNA). Between 50 and 100 nucleotides the values for the relaxation rates are linearly interpolated from the values at 50 and 100 nucleotides. Calculated are the transverse relaxation rates of a proton directly bonded to a  $^{13}\text{C}$  nucleus (A), a  $^{12}\text{C}$  nucleus (B and C), and a  $^{15}\text{N}$  nucleus (D). The relaxation is assumed to result from dipolar interactions alone, either with the directly bonded nucleus (the bond length is assumed to be  $1.095 \text{ \AA}$ ) or with other protons (see eqns (81)–(88), Section 8). Note that the hetero nucleus relaxes by the directly bonded proton with the same rate as the proton is relaxed by the directly bonded hetero nucleus. Two situations are always considered, namely the relaxation vector lies parallel to the helix axis or perpendicular to it. The spectral density function is then calculated according to Eq. (63); in the former case the density function is given by  $J(\omega, \tau_1)$  and in the latter by  $0.75J(\omega, \tau_1) + 0.25J(\omega, \tau_3)$ . The horizontal dotted lines indicate relaxation rates of 20 and 30 Hz, respectively. These values can be considered roughly the level above which triple resonance experiments start to become too insensitive. The two vertical arrows indicate where these levels cross the expected relaxation rates, and should thus indicate roughly the maximum lengths at which triple resonance experiments can be performed. (A) The effect on the proton transverse relaxation rate due to dipolar interaction with a directly bonded  $^{13}\text{C}$  nucleus. The inset shows the sum of the rates due to dipolar interaction with the directly bonded nucleus and due to dipolar interaction with other protons at 2.2, 2.4 and  $3.0 \text{ \AA}$ , respectively; internal motion is taken to be present, scaling the transverse relaxation rates by  $0.8 (= S^2)$ . (B) Transverse proton relaxation rate resulting from dipolar interactions with protons at 2.2, 2.4 and  $3.0 \text{ \AA}$ . (C) The effect of deuteration: transverse proton relaxation rate resulting from dipolar interaction with one proton at  $2.4 \text{ \AA}$ . (D) Transverse imino relaxation rate resulting from dipolar interaction with the directly bonded  $^{15}\text{N}$ , and dipolar interaction with protons at 3.0, 3.0, 2.7 and  $2.7 \text{ \AA}$  (broken lines); in addition, 10 Hz exchange broadening is assumed (drawn lines). The broken lines represent the effect of deuteration of non-exchanging protons: the rate is calculated as above but now with protons at 3.0, 3.0 and  $2.7 \text{ \AA}$ . The inset gives the transverse rate resulting from dipolar interaction with the directly bonded nucleus alone. This represents the relaxation rates of in-phase  $^{15}\text{N}$  relaxation. Internal mobility is taken to be present scaling the rate by a factor of  $0.8 (= S^2)$ .

assignment techniques, one arrives at a size limit of 120 nucleotides. It remains to be seen whether this limit is unrealistically large.

The above considerations show that, for nucleic acids, it is of penultimate importance to establish labeling methods that reduce the overlap in the spectra. The above-mentioned approach of residue-type-specific labeling is a straightforward extension of existing and practiced methods which employ enzymatic means to achieve the labeling. A further extension of the size limit could possibly be achieved when using different site-specific mutant forms of the studied sequence. Finally, site-specific labeling is an ideal approach, but for RNAs is extremely expensive, even more so for larger systems. Obviously, present labeling methods have to be extended to be able to successfully study larger systems ( $> 120$ ) by NMR.

A final aspect that plays a significant role is the development of improved structure calculation methods. As pointed out in Section 6, the first steps have only just been taken towards quantitative utilization of chemical shifts [50,226]. Furthermore, the use of ambiguously assigned NOEs in simulated annealing has been demonstrated for protein NMR by Nilges [227], but has as yet not been tried for nucleic acids. This method is particularly interesting, since Nilges has shown that only a very small number of assigned NOEs is required to derive the correct three-dimensional structure. Furthermore, Smith et al. [228] have recently considered how minimal a dataset is required for NMR studies of large proteins (over 40 kDa), to still be able to derive a three-dimensional structure, albeit of low resolution. Application of these methods in the field of nucleic acids will certainly help to further extend the size limit.

In conclusion, it seems that with the introduction of specific deuteration and partial  $^{13}\text{C}$  labeling techniques, the limit of the systems that can be studied lies well beyond 100 nucleotides. Residue-type-specific labeling can extend the range to about 120 nucleotides. Beyond that, site-specific-labeling techniques are required. It is interesting to consider in this respect the group I domain, for which the structure has recently been determined by means of X-ray diffraction. The total length of this sequence is 160 nucleotides. The structure of the group I intron domain shows two side-by-side helices of approximately 60 base pairs each. This corresponds to a

molecule with a rotational correlation time as described earlier in Fig. 35, for which many through-bond assignment techniques can still be applied, when labeling techniques are judiciously applied. It is also interesting to note that the model of the structure of the complete group I intron does not have a size much larger than this.

## 11. Conclusions

The last five years have seen very exciting innovations in the field of NMR of nucleic acids, spawned by the development of RNA labeling techniques. It is to be expected that this field will further mature in the near future. A number of possible developments, which have been hinted at in this review, are expected to occur in the near future. Rather simple extensions of the present labeling methods for RNA will allow larger systems to be tackled. It is also expected that DNA labeling will become a viable technique, so that parallel developments can occur in this area. Relaxation studies are expected to be performed and to provide answers to questions concerning the conformational exchange processes, which are so important for the functionality of RNAs. Furthermore, the spin relaxation depends on the orientation of its relaxation vector with respect to the long axis of the cylindrical shape of nucleic acid molecules. This dependence may possibly be used as a structural parameter to determine the relative orientation of molecular segments (see Section 8). An exciting development is to make use of the slight alignment via orientation along the magnetic field which shows up in high magnetic fields for biomolecules that have an anisotropic diamagnetic susceptibility tensor [229,230]. This alignment influences the magnitude of the  $^1J_{\text{NH}}$ - and  $^1J_{\text{CH}}$ -couplings in a field-dependent manner. When measurable, this also provides global structural information [229,230]. Furthermore, the first results indicate that  $^1\text{H}$  chemical shifts, which mainly result from magnetic anisotropy effects of the aromatic rings, can be calculated reliably. Since nucleic acids are essentially aromatic in nature it is to be expected that chemical shift based structural constraints will contribute considerably to better, i.e. more detailed, structures. Finally, computational methods based on floating assignments are expected



to be introduced and further improve the NMR structure calculations.

## Acknowledgements

The authors wish to thank Drs J.H. Ippel and J. Schleucher for helpful comments and critical reading of the manuscript. S.W. wishes to thank his former colleagues at the Nijmegen laboratory for the exciting scientific environment they provided; Dr H.A. Heus introduced him to isotope labeled RNA, and made him more aware of the many biological functions of RNA. Furthermore, the authors want to express their appreciation to Professor C.W. Hilbers for his support and advice, and the many scientific discussions concerning NMR and its application to structural studies of nucleic acids.

## References

- [1] R.F. Gesteland, J.F. Atkins, *The RNA World*, Cold Spring Harbor Laboratory Press, 1997.
- [2] H.W. Pley, K.M. Flaherty, D.B. McKay, *Nature* 372 (1994) 68.
- [3] W.G. Scott, J.T. Finch, A. Klug, *Cell* 81 (1995) 991.
- [4] J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, A.A. Szewczak, C.E. Kundrot, T.R. Cech, J.A. Doudna, *Science* 273 (1996) 1696.
- [5] J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, C.E. Kundrot, T.R. Cech, J.A. Doudna, *Science* 273 (1996) 1678.
- [6] C. Oubridge, N. Ito, P.R. Evans, C.H. Teo, K. Nagai, *Nature* 372 (1994) 432.
- [7] K. Vaiegard, J.B. Murray, P.G. Stockley, J. Stonehouse, L. Liljas, *Nature* 371 (1994) 623.
- [8] C. Kang, X. Zhang, R. Ratliff, R. Moyzis, A. Rich, *Nature* 356 (1992) 126.
- [9] S.R. Holbrook, C. Cheong, I. Tinoco Jr., *Nature* 353 (1991) 579.
- [10] R. Chattopadhyaya, K. Grzeskowiak, R.E. Dickerson, *J. Mol. Biol.* 211 (1990) 189.
- [11] C.W. Hilbers, H.A. Heus, M.J.P. Van Dongen, S.S. Wijmenga, *Nucleic Acids and Molecular Biology*, Springer, Berlin, 1994, p. 56.
- [12] M.J.P. Van Dongen, S.S. Wijmenga, G.A. Van der Marel, J.H. Van Boom, C.W. Hilbers, *J. Mol. Biol.* 263 (1996) 715.
- [13] F.W. Smith, J. Feigon, *Nature* 356 (1992) 164.
- [14] A. Kettani, R.A. Kumar, D.J. Patel, *J. Mol. Biol.* 254 (1995) 638.
- [15] J. Feigon, K.M. Koshlap, F.W. Smith, *Methods Enzymol.* 261 (1995) 225.
- [16] K. Gehring, J.L. Leroy, M. Gueron, *Nature* 363 (1993) 561.
- [17] J.R. Williamson, *Nature* 382 (1996) 112.
- [18] G.C. King, J.W. Harper, Z. Xi, *Methods Enzymol.* 261 (1995) 436.
- [19] E.V. Puglisi, J.D. Puglisi, J.R. Williamson, U.L. Raj-Bhandary, *Proc. Natl. Acad. Sci. USA* 91 (1994) 11467.
- [20] F. Aboul-El, J. Karn, G. Varani, *Nucleic Acids Res.* 24 (1996) 3974.
- [21] G. Varani, *Ann. Rev. Biophys. Biomol. Struct.* 24 (1995) 379.
- [22] F.H. Allain, G. Varani, *J. Mol. Biol.* 250 (1995) 333.
- [23] F.H. Allain, G. Varani, *Nucleic Acids Res.* 23 (1995) 341.
- [24] S.R. Mirmira, I. Tinoco Jr., *Biochemistry* 35 (1996) 7664.
- [25] Z.P. Cai, I. Tinoco Jr., *Biochemistry* 35 (1996) 6026.
- [26] H.S. Kang, J.V. Hines, I. Tinoco Jr., *J. Mol. Biol.* 259 (1996) 135.
- [27] K.J. Luebke, I. Tinoco Jr., *Biochemistry* 35 (1996) 11677.
- [28] S.R. Lynch, J.G. Pelton, I. Tinoco Jr., *Magn. Reson. Chem.* 34 (1996) S11.
- [29] S.R. Mirmira, I. Tinoco Jr., *Biochemistry* 35 (1996) 7675.
- [30] J. Nowakowski, I. Tinoco Jr., *Biochemistry* 35 (1996) 2577.
- [31] D. Fourmy, M.I. Recht, S.C. Blanchard, J.D. Puglisi, *Science* 274 (1996) 1367.
- [32] J.L. Battiste, H.Y. Mao, N.S. Rao, R.Y. Tan, D.R. Muhandiram, L.E. Kay, A.D. Frankel, J.R. Williamson, *Science* 273 (1996) 1547.
- [33] A.S. Brodsky, J.R. Williamson, *J. Mol. Biol.* 267 (1997) 624.
- [34] F. Aboul-El, J. Karn, G. Varani, *J. Mol. Biol.* 253 (1995) 313.
- [35] X.Y. Chen, H.S. Kang, L.X. Shen, M. Chamorro, H.E. Varmus, I. Tinoco Jr., *J. Mol. Biol.* 260 (1996) 479.
- [36] G. Varani, F. Aboul-El, F.H. Allain, *Progress in Nuclear Magnetic Resonance Spectroscopy* 29 (1996) 51.
- [37] F.H.T. Allain, C.C. Gubser, P.W.A. Howe, K. Nagai, D. Neuhaus, G. Varani, *Nature* 380 (1996) 646.
- [38] C.C. Gubser, G. Varani, *Biochemistry* 35 (1996) 2253.
- [39] D.P. Zimmer, D.M. Crothers, *Proc. Natl. Acad. Sci. USA* 92 (1995) 3091.
- [40] D.J. Aberhard, personal communication.
- [41] D.M.J. Lilley, *DNA-Protein Structural Interactions*, Oxford University Press, Oxford, 1995.
- [42] C. Altona, J.A. Pikkemaat, F.J.J. Overmars, *Curr. Opin. Struct. Biol.* 6 (1996) 305.
- [43] J.A. Pikkemaat, H. Van den Elst, J.H. Van Boom, C. Altona, *Biochemistry* 33 (1994) 14896.
- [44] F.J.J. Overmars, J.A. Pikkemaat, H. Van den Elst, J.H. Van Boom, C. Altona, *J. Mol. Biol.* 255 (1996) 702.
- [45] R.T. Batey, J.L. Battiste, J.R. Williamson, *Methods Enzymol.* 261 (1995) 300.
- [46] K.B. Hall, *Methods Enzymol.* 261 (1995) 542.
- [47] J.D. Puglisi, J.R. Wyatt, *Methods Enzymol.* 261 (1995) 323.
- [48] A. Pardi, *Methods Enzymol.* 261 (1995) 350.
- [49] J.H. Ippel, S.S. Wijmenga, R. de Jong, H.A. Heus, C.W. Hilbers, E. de Vroom, G.A. Van der Marel, J.H. Van Boom, *Magn. Reson. Chem.* 34 (1996) S156.
- [50] S.S. Wijmenga, M. Kruithof, C.W. Hilbers, *Analysis of <sup>1</sup>H chemical shifts in DNA. Assessment of the reliability of <sup>1</sup>H*

- chemical calculations for use in structure refinement, *J. Biomol. NMR* 10 (1997) 337.
- [51] J.F. Milligan, D.R. Groebe, G.W. Witherell, O.C. Uhlenbeck, *Nucleic Acids Res.* 15 (1987) 8783.
  - [52] J.F. Milligan, O.C. Uhlenbeck, in: J.E. Dahlberg, J. Abelson (Eds.), *Methods in Enzymology*, Academic Press, New York 180 (1989) 51.
  - [53] R.T. Batey, N. Cloutier, H.Y. Mao, J.R. Williamson, *Nucleic Acids Res.* 24 (1996) 4836.
  - [54] T.J. Tolbert, J.R. Williamson, *J. Am. Chem. Soc.* 118 (1996) 7929.
  - [55] R.T. Batey, M. Inada, E. Kujawinski, J.D. Puglisi, J.R. Williamson, *Nucleic Acids Res.* 20 (1992) 4515.
  - [56] S.S. Wijmenga, H.A. Heus, F.J. Van De Ven, C.W. Hilbers, in: C.I. Stassinopoulou (Ed.), *NMR of Biological Macromolecules*, Springer Verlag, Berlin, 1994, p. 304.
  - [57] E.P. Nikonowicz, M. Michnicka, K. Kalurachchi, E. DeJong, *Nucleic Acids Res.* 25 (1997) 1390.
  - [58] A. Ono, T. Makita, S.I. Tate, E. Kawashima, Y. Ishido, M. Kainosho, *Magn. Reson. Chem.* 34 (1996) S40.
  - [59] B.L. Gaffney, P.P. Kung, C. Wang, R.A. Jones, *J. Am. Chem. Soc.* 117 (1995) 12281.
  - [60] M.J. Michnicka, J.W. Harper, G.C. King, *Biochemistry* 32 (1993) 395.
  - [61] IUPAC–IUB Joint Commission Biochemical Nomenclature, *Eur. J. Biochem.* 131 (1983) 9.
  - [62] S.S. Wijmenga, M.M. Mooren, C.W. Hilbers, in: G.C.K. Roberts (Ed.), *NMR of Macromolecules, A Practical Approach*, Oxford University Press, New York, 1993, p. 217.
  - [63] M.J. Blommers, F.J. Van De Ven, G.A. Van der Marel, J.H. Van Boom, C.W. Hilbers, *Eur. J. Biochem.* 201 (1991) 33.
  - [64] M.R. Conte, C.J. Bauer, A.N. Lane, *J. Biomol. NMR* 7 (1996) 190.
  - [65] G. Liparo, A. Szabo, *J. Am. Chem. Soc.* 104 (1982) 4546.
  - [66] G. Liparo, A. Szabo, *J. Am. Chem. Soc.* 104 (1982) 4559.
  - [67] B.A. Borgias, T.L. James, *J. Magn. Reson.* 87 (1990) 475.
  - [68] U. Schmitz, T.L. James, *Methods Enzymol.* 261 (1995) 3.
  - [69] N.B. Ulyanov, T.L. James, *Methods Enzymol.* 261 (1995) 90.
  - [70] C.B. Post, R.P. Meadows, D.G. Gorenstein, *J. Am. Chem. Soc.* 112 (1990) 6796.
  - [71] K. Kaluarachchi, R.P. Meadows, D.G. Gorenstein, *Biochemistry* 30 (1991) 8785.
  - [72] B.A. Luxon, D.G. Gorenstein, *Methods Enzymol.* 261 (1995) 45.
  - [73] R. Boelens, T.M.G. Koning, G.A. Van der Marel, J.H. Van Boom, R. Kaptein, *J. Magn. Reson.* 82 (1989) 290.
  - [74] F.J.M. Van De Ven, M.J.J. Blommers, R.E. Schouten, C.W. Hilbers, *J. Magn. Reson.* 94 (1991) 140.
  - [75] S.S. Wijmenga, H.A. Heus, B. Werten, G.A. Van der Marel, J.H. Van Boom, C.W. Hilbers, *J. Magn. Reson. B.* 103 (1994) 134.
  - [76] C.W. Hilbers, S.S. Wijmenga, H. Hoppe, H.A. Heus, in: O. Jardetzky, J.F. Lefevre (Eds.), *Dynamics and the Problem of Recognition in Biological Macromolecules*, Plenum Press, New York, 1996, p. 193.
  - [77] H.A. Heus, S.S. Wijmenga, H. Hoppe, C.W. Hilbers, The detailed structure of tandem G.A mismatched base pair motifs in RNA duplexes is context dependent, *J. Mol. Biol.* 271 (1997) 147.
  - [78] A.T. Brünger, XPLOR manual [computer program], 1995.
  - [79] F.H.T. Allain, G. Varani, *J. Mol. Biol.* 267 (1997) 338.
  - [80] F.J. Van De Ven, C.W. Hilbers, *Eur. J. Biochem.* 178 (1988) 1.
  - [81] C.W. Hilbers, M.J.J. Blommers, F.J.M. Van De Ven, J.H. Van Boom, G.A. Van der Marel, *Nucleosides and Nucleotides* 10 (1991) 61.
  - [82] D.B. Davies, M. MacCoss, S.S. Danyluk, *J. Chem. Soc. Chem. Comm.* 99 (1984) 536.
  - [83] C.A. Haasnoot, F.A.A.M. de Leeuw, C. Altona, *Tetrahedron* 36 (1980) 2783.
  - [84] J. Van Wijk, B.D. Huckriede, J.H. Ippel, C. Altona, *Methods Enzymol.* 211 (1992) 286.
  - [85] Y.T. van den Hoogen, C.M. Hilgersom, D. Brozda, K. Lesiak, P.F. Torrence, C. Altona, *Eur. J. Biochem.* 182 (1989) 629.
  - [86] J.P. van Duynhoven, J. Goudriaan, C.W. Hilbers, S.S. Wijmenga, *J. Am. Chem. Soc.* 114 (1992) 10055.
  - [87] K. Bock, C. Pederson, *Acta Chim. Scand., Ser. B* 34 (1997) 354.
  - [88] M.M. Mooren, S.S. Wijmenga, G.A. Van der Marel, J.H. Van Boom, C.W. Hilbers, *Nucleic Acids Res.* 22 (1994) 2658.
  - [89] M.M. Mooren, On nucleic acid structure analysis by NMR University of Nijmegen, Netherlands, 1994.
  - [90] S.A. Schroeder, V. Roongta, J.M. Fu, C.R. Jones, D.G. Gorenstein, *Biochemistry* 28 (1989) 8292.
  - [91] D.A. Pearlman, S.H. Kim, *J. Biomol. Struct. Dyn.* 4 (1986) 49.
  - [92] J.V. Hines, S.M. Landry, G. Varani, I. Tinoco Jr., *J. Am. Chem. Soc.* 116 (1994) 5823.
  - [93] A. Bax, G.W. Vuister, S. Grzesiek, F. Delaglio, A.C. Wang, R. Tschudin, G. Zhu, *Methods Enzymol.* 239 (1994) 97.
  - [94] M. Eberstadt, G. Gemmecker, D.F. Mierke, H. Kessler, *Angewandte Chemie – International Edition English* 34 (1995) 1671.
  - [95] Y. Kim, J.H. Prestegard, *J. Magn. Reson.* 84 (1989) 9.
  - [96] H. Widmer, K. Wüthrich, *J. Magn. Reson.* 74 (1987) 316.
  - [97] M.J. Blommers, D. Nanz, O. Zerbe, *J. Biomol. NMR* 4 (1994) 595.
  - [98] J.A. Pikkemaat, C. Altona, *Magn. Reson. Chem.* 34 (1996) S33.
  - [99] R.F. Macaya, P. Schultze, J. Feigon, *J. Am. Chem. Soc.* 114 (1991) 781.
  - [100] M. Leijon, J. Zdunek, H. Fritzsche, H. Sklenar, A. Graslund, *Eur. J. Biochem.* 234 (1995) 832.
  - [101] R. McIntyre, R. Freeman, *J. Magn. Reson.* 96 (1992) 425.
  - [102] J.J. Titman, J. Keeler, *J. Magn. Reson.* 89 (1990) 640.
  - [103] J. Stonehouse, J. Keeler, *J. Magn. Reson. Series A* 112 (1995) 43.
  - [104] T. Szyperski, P. Güntert, G. Otting, K. Wüthrich, *J. Magn. Reson.* 99 (1992) 552.
  - [105] L. Mueller, *J. Magn. Reson.* 72 (1987) 191.
  - [106] S. Seip, J. Balbach, H. Kessler, *J. Magn. Reson. Series B* 104 (1994) 172.
  - [107] P. Schmieder, J.H. Ippel, H. Van den Elst, G.A. Van der Marel, J.H. Van Boom, C. Altona, H. Kessler, *Nucleic Acids Res.* 20 (1992) 4747.

- [108] B.N.M. van Buuren, S.S. Wijmenga, H.A. Heus, S. Mronga, C.W. Hilbers, Determination of all backbone  $J$ -couplings in a uniformly  $^{13}\text{C}$  labeled RNA hairpin, 1997 (unpublished).
- [109] H. Schwalbe, W. Samtag, J.W. Engels, W. Bermel, C. Griesinger, *J. Biomol. NMR* 3 (1993) 479.
- [110] S.S. Wijmenga, K. Hallenga, C.W. Hilbers, *J. Magn. Reson.* 84 (1989) 634.
- [111] S.J. Glaser, H. Schwalbe, J.P. Marino, C. Griesinger, *J. Magn. Reson. Series B* 112 (1996) 160.
- [112] H. Schwalbe, J.P. Marino, G.C. King, R. Wechselberger, W. Bermel, C. Griesinger, *J. Biomol. NMR* 4 (1994) 631.
- [113] D.P. Zimmer, J.P. Marino, C. Griesinger, *Magn. Reson. Chem.* 34 (1996) S177.
- [114] M.M. Mooren, C.W. Hilbers, S.S. Wijmenga, *J. Magn. Reson.* 94 (1991) 101.
- [115] G. Zhu, D. Live, A. Bax, *J. Am. Chem. Soc.* 116 (1994) 8370.
- [116] G.W. Vuister, A. Bax, *J. Magn. Reson.* 98 (1992) 428.
- [117] L. Szilagyi, *Progress in Nuclear Magnetic Resonance Spectroscopy* 27 (1995) 325.
- [118] D.S. Wishart, B.D. Sykes, *Methods Enzymol.* 239 (1994) 363.
- [119] D.A. Case, H.J. Dyson, P.E. Wright, *Methods Enzymol.* 239 (1994) 392.
- [120] M.P. Williamson, T. Asakura, *J. Magn. Reson. Series B* 101 (1993) 63.
- [121] M.P. Williamson, J. Kikuchi, T. Asakura, *J. Mol. Biol.* 247 (1995) 541.
- [122] K. Ōsapaya, D.A. Case, *J. Am. Chem. Soc.* 113 (1991) 9436.
- [123] K. Ōsapaya, Y. Theriault, P.E. Wright, D.A. Case, *J. Mol. Biol.* 244 (1994) 183.
- [124] E. Oldfield, *J. Biomol. NMR* 5 (1995) 217.
- [125] J. Kuszewski, A.M. Gronenborn, G.M. Clore, *J. Magn. Reson. Series B* 107 (1995) 293.
- [126] J. Kuszewski, J. Qin, A.M. Gronenborn, G.M. Clore, *J. Magn. Reson. Series B* 106 (1995) 92.
- [127] C. Giessner-Pretre, B. Pullman, *Q. Rev. Biophys.* 20 (1987) 113.
- [128] C.E. Johnson, F.A. Bovey, *J. Chem. Phys.* 29 (1958) 1012.
- [129] C.W. Haigh, R.B. Mallion, *Progress in Nuclear Magnetic Resonance Spectroscopy* 13 (1980) 303.
- [130] R. Ribas Prado, C. Giessner-Pretre, *J. Molec. Structure* 76 (1981) 81.
- [131] A.D. Buckingham, *Can. J. Chem.* 38 (1960) 300.
- [132] P. Marshall, M. Pople, *Molec. Phys.* 1 (1958) 199.
- [133] F.J. Van De Ven, C.W. Hilbers, *Nucleic Acids Res.* 16 (1988) 5713.
- [134] R. Ghose, J.P. Marino, K.B. Wiberg, J.H. Prestegard, *J. Am. Chem. Soc.* 116 (1994) 8827.
- [135] P. Legault, A. Pardi, *J. Magn. Reson. Series B* 103 (1993) 82.
- [136] H.A. Heus, A. Pardi, *J. Am. Chem. Soc.* 113 (1991) 4360.
- [137] A. Pardi, R. Walker, S. Rapoport, G. Wider, K. Wuthrich, *J. Am. Chem. Soc.* 105 (1983) 1652.
- [138] G.W. Kellogg, B.I. Schweitzer, *J. Biomol. NMR* 3 (1993) 577.
- [139] M.J.P. Van Dongen, J.F. Doreleijers, J.H. Van Boom, G.A. Van der Marel, C.W. Hilbers, S.S. Wijmenga, 1997 (in preparation).
- [140] N.B. Leontis, M.E. Piotto, M.T. Hills, A. Malhotra, I.V. Ouporov, J.M. Nussbaum, D.G. Gorenstein, *Methods Enzymol.* 261 (1995) 183.
- [141] C. Altona, *J. Mol. Biol.* 263 (1996) 568.
- [142] J.A. Pikkemaat, F.J.J. Overmars, C.M. Dreef-Tromp, H. Van den Elst, J.H. Van Boom, C. Altona, *J. Mol. Biol.* 262 (1996) 349.
- [143] M.J.P. Van Dongen, H.A. Heus, S.S. Wijmenga, G.A. Van der Marel, J.H. Van Boom, C.W. Hilbers, *Biochemistry* 35 (1996) 1733.
- [144] M.J.P. Van Dongen, S.S. Wijmenga, R. Eritja, F. Azorín, C.W. Hilbers, *J. Biomol. NMR* 8 (1996) 207.
- [145] G.M. Clore, J.G. Omichinski, K. Sakaguchi, N. Zambrano, H. Sakamoto, E. Appella, A.M. Gronenborn, *Science* 265 (1994) 386.
- [146] G.M. Clore, J.G. Omichinski, K. Sakaguchi, N. Zambrano, H. Sakamoto, E. Appella, A.M. Gronenborn, *Science* 267 (1995) 1515.
- [147] G. Varani, C. Cheong, I. Tinoco Jr., *Biochemistry* 30 (1991) 3280.
- [148] E.P. Nikonowicz, A. Pardi, *J. Mol. Biol.* 232 (1993) 1141.
- [149] A. Pardi, E.P. Nikonowicz, *J. Am. Chem. Soc.* 114 (1992) 9202.
- [150] J.P. Simorre, G.R. Zimmermann, A. Pardi, B.T. Farmer II, L. Mueller, *J. Biomol. NMR* 6 (1995) 427.
- [151] J.P. Simorre, G.R. Zimmermann, L. Mueller, A. Pardi, *J. Biomol. NMR* 7 (1996) 153.
- [152] V. Sklenar, T. Dieckmann, S.E. Butcher, J. Feigon, *J. Biomol. NMR* 7 (1996) 83.
- [153] R. Fiala, F. Jiang, D.J. Patel, *J. Am. Chem. Soc.* 118 (1996) 689.
- [154] J.P. Marino, J.H. Prestegard, D.M. Crothers, *J. Am. Chem. Soc.* 116 (1994) 2205.
- [155] P. Legault, B.T. Farmer II, L. Mueller, A. Pardi, *J. Am. Chem. Soc.* 116 (1994) 2203.
- [156] E.R. Zuiderweg, *J. Magn. Reson.* 89 (1990) 533.
- [157] S.W. Fesik, H.L. Eaton, E.T. Olejniczak, E.R. Zuiderweg, L.P. McIntosh, F.W. Dahlquist, *J. Am. Chem. Soc.* 112 (1990) 886.
- [158] A. Bax, G.M. Clore, A. Gronenborn, *J. Magn. Reson.* 88 (1990) 425.
- [159] L.E. Kay, P. Keifer, T. Saarinen, *J. Am. Chem. Soc.* 114 (1992) 10663.
- [160] J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletsky, S.J. Glaser, O.W. Sorensen, C. Griesinger, *J. Biomol. NMR* 4 (1994) 301.
- [161] M. Sattler, M. Schwendinger, J. Schleucher, C. Griesinger, *J. Biomol. NMR* 6 (1995) 11.
- [162] M. Sattler, P. Schmidt, J. Schleucher, O. Schedletsky, S.J. Glaser, C. Griesinger, *Journal of Magnetic Resonance B* 108 (1995) 235.
- [163] S.S. Wijmenga, E. Steensma, C.P.M. van Mierlo, *J. Magn. Reson.* 124 (1997) 459.
- [164] A. Majumdar, H. Wang, R.C. Morshauer, E.R. Zuiderweg, *J. Biomol. NMR* 3 (1993) 387.
- [165] H. Wang, E.R. Zuiderweg, *J. Biomol. NMR* 5 (1995) 207.
- [166] F.J.M. Van De Ven, M.E.P. Philippens, *J. Magn. Reson.* 97 (1992) 202.

- [167] V. Sklenar, R.D. Peterson, M.R. Rejante, J. Feigon, J. Biomol. NMR 3 (1993) 721.
- [168] B.T. Farmer II, L. Mueller, E.P. Nikonowicz, A. Pardi, J. Biomol. NMR 4 (1994) 129.
- [169] V. Sklenar, R.D. Peterson, M.R. Rejante, E. Wang, J. Feigon, J. Am. Chem. Soc. 115 (1993) 12181.
- [170] B.T. Farmer, L. Mueller, E.P. Nikonowicz, A. Pardi, J. Am. Chem. Soc. 115 (1993) 11040.
- [171] S. Tate, A. Ono, M. Kainosho, J. Am. Chem. Soc. 116 (1994) 5977.
- [172] V. Sklenar, R.D. Peterson, M.R. Rejante, J. Feigon, J. Biomol. NMR 4 (1994) 117.
- [173] L.E. Kay, M. Ikura, A. Bax, J. Magn. Reson. 91 (1991) 84.
- [174] L.E. Kay, M. Ikura, R. Tschudin, A. Bax, J. Magn. Reson. 89 (1990) 496.
- [175] H.A. Heus, S.S. Wijmenga, F.J. Van De Ven, C.W. Hilbers, J. Am. Chem. Soc. 116 (1994) 4983.
- [176] J.P. Marino, J.H. Prestegard, D.M. Crothers, J. Am. Chem. Soc. 116 (1994) 2205.
- [177] S.S. Wijmenga, H.A. Heus, H.A. Leeuw, H. Hoppe, M. van der Graaf, C.W. Hilbers, J. Biomol. NMR 5 (1995) 82.
- [178] G. Varani, F. Aboul-Ela, F. Allain, C.C. Gubser, J. Biomol. NMR 5 (1995) 315.
- [179] R. Ramachandran, C. Sich, M. Grune, V. Soskic, L.R. Brown, J. Biomol. NMR 7 (1996) 251.
- [180] G. Otting, H. Senn, G. Wagner, K. Wuthrich, J. Magn. Reson. 70 (1986) 500.
- [181] R. Kaptein, R. Boelens, V.P. Chuprina, J.A.C. Rullmann, M. Slijper, Methods Enzymol. 261 (1995) 513.
- [182] J.L. Battiste, R. Tan, A.D. Frankel, J.R. Williamson, J. Biomol. NMR 6 (1995) 375.
- [183] J.R. Williamson, J.L. Battiste, H. Mao, A.D. Frankel, Nucleic Acids Symp. Ser. (1995) 46.
- [184] J.D. Puglisi, L. Chen, A.D. Frankel, J.R. Williamson, Proc. Natl. Acad. Sci. USA 90 (1993) 3680.
- [185] P.J. Folkers, M. Nilges, R.H. Folmer, R.N. Konings, C.W. Hilbers, J. Mol. Biol. 236 (1994) 229.
- [186] J.J. Prompers, R.H. Folmer, M. Nilges, P.J. Folkers, R.N. Konings, C.W. Hilbers, Eur. J. Biochem. 232 (1995) 506.
- [187] R.H. Folmer, M. Nilges, R.N. Konings, C.W. Hilbers, EMBO J. 14 (1995) 4132.
- [188] F.J. Van de Ven, Multidimensional NMR in Liquids; Basic Principles and Experimental Methods, VCH Publishers, New York, 1995.
- [189] M.M. Tirado, G.J. de la Torre, J. Chem. Phys. 73 (1980) 1986.
- [190] S. Broersma, J. Chem. Phys. 63 (1959) 423.
- [191] F. Perrin, J. Phys. Radium 5 (1934) 497.
- [192] D.E. Woessner, J. Chem. Phys. 37 (1962) 647.
- [193] D.E. Woessner, J. Chem. Phys. 42 (1964) 1855.
- [194] D.E. Woessner, B.S. Snowden, G.H. Meyer, J. Chem. Phys. 50 (1968) 710.
- [195] A.N. Lane, Methods Enzymol. 261 (1995) 413.
- [196] G. Wagner, Curr. Opin. Struct. Biol. 3 (1993) 748.
- [197] J. Peng, G. Wagner, J. Magn. Reson. 98 (1992) 308.
- [198] J. Peng, G. Wagner, Biochemistry 32 (1992) 8571.
- [199] C. Houssier, C.T.O. Konski, in: S. Krause (Ed.), Molecular Optics, Plenum Press, New York, 1980, p. 309.
- [200] S.S. Wijmenga, Electric birefringe experiments on aqueous polyelectrolyte salt solutions. Influence of concentration and flexibility, Thesis, University of Leiden, 1984.
- [201] S.S. Wijmenga, F. van der Touw, M. Mandel, in: B. Sedlacek (Ed.), Physical Optics of Dynamic Phenomena and Processes in Macromolecular Systems, Walter de Gruyter and Co, Berlin, 1985, p. 87.
- [202] P.R. Bevington, Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill, New York, 1969.
- [203] N. Tjandra, P. Wingfield, S. Stahl, A. Bax, J. Biomol. NMR 8 (1996) 273.
- [204] I. Phan, J. Boyd, I.D. Campbell, J. Biomol. NMR 8 (1996) 369.
- [205] R.A. Komoroski, A. Allerhand, Proc. Natl. Acad. Sci. USA 69 (1972) 1804.
- [206] R.A. Komoroski, A. Allerhand, Biochemistry 13 (1974) 369.
- [207] M.E. Hogan, O. Jardetzky, Biochemistry 19 (1980) 3460.
- [208] J. Keepers, T. James, J. Am. Chem. Soc. 104 (1982) 929.
- [209] B.H. Robinson, G.P. Drobny, Methods Enzymol. 261 (1995) 451.
- [210] J.R. Williamson, S.G. Boxer, Biochemistry 28 (1989) 2819.
- [211] W. Eimer, J.R. Williamson, S.G. Boxer, R. Pecora, Biochemistry 29 (1990) 799.
- [212] A.N. Lane, Carbohydr. Res. 221 (1991) 123.
- [213] P.N. Borer, S.R. LaPlante, A. Kumar, N. Zanatta, A. Martin, A. Hakkinen, G.C. Levy, Biochemistry 33 (1994) 2441.
- [214] M. Slijper, R. Boelens, A.L. Davis, R.N.H. Konings, G.A. Van der Marel, J.H. Van Boom, R. Kaptein, Biochemistry 36 (1997) 249.
- [215] F. Gaudin, F. Paquet, L. Chanteloup, J.M. Beau, T.T. Nguyen, G. Lancelot, J. Biomol. NMR 5 (1995) 49.
- [216] J.L. Markley, A. Bax, Y. Arata, C.W. Hilbers, R. Kaptein, B.D. Sykes, K. Wuthrich, IUPAC recommendations, draft, 1997.
- [217] M.J.P. Van Dongen, M.M.W. Mooren, E.F.A. Willems, G.A. Van der Marel, J.H. Van Boom, S.S. Wijmenga, C.W. Hilbers, Nucleic Acids Res. 25 (1997) 1537.
- [218] J. Santa Lucia Jr., D.H. Turner, Biochemistry 32 (1993) 12612.
- [219] E.G. Stein, L.M. Rice, A.T. Brunger, J. Magn. Reson. 124 (1997) 154.
- [220] A.T. Brunger, G.M. Clore, A.M. Gronenborn, R. Saffrich, M. Nilges, Science 261 (1993) 328.
- [221] P. Schultze, J. Feigon, Nature 387 (1997) 668.
- [222] A.E. Torda, R. Scheck, W.F. van Gunsteren, J. Mol. Biol. 214 (1990) 223.
- [223] A.E. Torda, R.M. Brunne, T. Huber, H. Kessler, W.F. van Gunsteren, J. Biomol. NMR 3 (1993) 55.
- [224] G. Wagner, J. Biomol. NMR 3 (1993) 375.
- [225] T. Dieckmann, J. Feigon, J. Biomol. NMR 9 (1997) 259.
- [226] J. Nowakowski, J.L. Miller, P.A. Kollman, I. Tinoco Jr., J. Am. Chem. Soc. 118 (1996) 12812.
- [227] M. Nilges, J. Mol. Biol. 245 (1995) 645.
- [228] B.O. Smith, Y. Ito, A. Raine, S. Teichmann, L. Ben-Tovim, D. Nietlispach, R.W. Broadhurst, T. Terada, M. Kelly, H. Oschkinat, T. Shibata, S. Yokoyama, E.D. Laue, J. Biomol. NMR 8 (1996) 360.

- [229] N. Tjandra, A. Bax, *J. Magn. Reson.* 124 (1997) 512.
- [230] N. Tjandra, S. Grzesiek, A. Bax, *J. Am. Chem. Soc.* 118 (1996) 6264.
- [231] P. Büchner, W. Maurer, H. Rüterjans, *J. Magn. Reson.* 29 (1978) 45.
- [232] C.D. Poulter, C.L. Livinnston, *Tetrahedron Lett.* 13 (1979) 755.
- [233] R.L. Lipnick, J.P. Fissekis, *J. Org. Chem.* 44 (1979) 1627.
- [234] T.M. Logan, E.T. Olejniczak, R.X. Xu, S.W. Fesik, *J. Biomol. NMR* 3 (1993) 225.
- [235] R.D. Bertrand, W.B. Moniz, A.N. Garroway, G.C. Chingas, *J. Am. Chem. Soc.* 100 (1978) 5227.
- [236] L. Mueller, R.R. Ernst, *Molec. Phys.* 38 (1979) 963.
- [237] A.J. Shaka, C.J. Lee, A. Pines, *J. Magn. Reson.* 77 (1988) 274.
- [238] A. Mohebbi, A.J. Shaka, *J. Chem. Phys. Lett.* 178 (1991) 374.
- [239] L. Braunschweiler, R.R. Ernst, *J. Magn. Reson.* 53 (1983) 521.
- [240] A. Majumdar, E.R. Zuiderweg, *J. Magn. Reson. Series A* 113 (1995) 19.
- [241] H. Geen, R. Freeman, *J. Magn. Reson.* 93 (1991) 93.
- [242] S. Grzesiek, A. Bax, *J. Am. Chem. Soc.* 114 (1992) 6291.
- [243] M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR* 2 (1992) 661.
- [244] S. Grzesiek, A. Bax, *J. Am. Chem. Soc.* 115 (1993) 12593.
- [245] D. Marion, M. Ikura, R. Tschudin, A. Bax, *J. Magn. Reson.* 85 (1989) 393.
- [246] S.L. Patt, *J. Magn. Reson.* 96 (1992) 94.
- [247] A.J. Shaka, P. Barker, R.J. Freeman, *J. Magn. Reson.* 64 (1985) 547.
- [248] M.H. Levitt, *Progress in Nuclear Magnetic Resonance Spectroscopy* 18 (1986) 61.
- [249] V. Sklenar, *J. Magn. Reson. Series A* 114 (1995) 132.
- [250] A.G. Palmer, J. Cavanagh, P.E. Wright, M. Rance, *J. Magn. Reson.* 93 (1991) 151.
- [251] B.L. Tomlinson, H.D.W. Hill, *J. Chem. Phys.* 59 (1973) 1775.
- [252] R. Konrat, I. Burghardt, G. Bodenhausen, *J. Am. Chem. Soc.* 113 (1991) 9135.
- [253] U. Eggenberger, P. Schmidt, S.J. Sattler, S.J. Glaser, C. Griesinger, *J. Magn. Reson.* 100 (1992) 604.
- [254] L. Emsley, G. Bodenhausen, *J. Chem. Phys. Lett.* 165 (1990) 469.
- [255] J. Boyd, N.J. Soffe, *J. Magn. Reson.* 85 (1989) 406.
- [256] A.J. Shaka, J. Keeler, T.A. Frenkiel, R.J. Freeman, *J. Magn. Reson.* 52 (1983) 335.
- [257] D. Marion, K. Wuthrich, *Biochem. Biophys. Res. Commun.* 113 (1983) 967.
- [258] G. Kontaxis, J. Stonehouse, E.D. Laue, J. Keeler, *J. Magn. Reson. Series A* 111 (1994) 70.
- [259] D.B. Davies, P. Rajana, H. Sadikot, *J. Chem. Soc., Perkin Trans. 2* (1985) 279.
- [260] M.H. Kolk, M. van der Graaf, S.S. Wijmenga, C.W.A. Pleij, H.A. Heus, C.W. Hilbers, *Science* 280 (1998) 434.