Pure & Appl. Chem., Vol. 70, No. 1, pp. 117–142, 1998.
Printed in Great Britain.

INTERNATIONAL UNION OF PURE
AND APPLIED CHEMISTRY

IUPAC–IUBMB–IUPAB INTER-UNION TASK GROUP ON THE
STANDARDIZATION OF DATA BASES OF PROTEIN AND NUCLEIC ACID
STRUCTURES DETERMINED BY NMR SPECTROSCOPY

RECOMMENDATIONS FOR THE PRESENTATION OF NMR STRUCTURES OF PROTEINS AND NUCLEIC ACIDS

(IUPAC Recommendations 1998)

Prepared for publication by

JOHN L. MARKLEY[1], AD BAX[2], YOJI ARATA[3], C. W. HILBERS[4], ROBERT KAPTEIN[5], BRIAN D. SYKES[6], PETER E. WRIGHT[7] AND KURT WÜTHRICH[8]

[1]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA; [2]Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, MD, USA; [3]Water Research Institute, Tsukuba, Japan; [4]Laboratory of Biophysical Chemistry, University of Nijmegen, The Netherlands; [5]Department of Chemistry, University of Utrecht, The Netherlands; [6]Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada; [7]Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA; [8]Institut für Molekularbiologie und Biophysik, ETH-Hönggerberg, Zürich, Switzerland (convenor of the task group to whom correspondence should be addressed).

Recommendations for the presentation of NMR structures of proteins and nucleic acids (Recommendations 1998)

Abstract: The recommendations presented here are designed to support easier communication of NMR data and NMR structures of proteins and nucleic acids through unified nomenclature and reporting standards. Much of this document pertains to the reporting of data in journal articles; however, in the interest of the future development of structural biology, it is desirable that the bulk of the reported information be stored in computer-accessible form and be freely accessible to the scientific community in standardized formats for data exchange. These recommendations stem from an IUPAC-IUBMB-IUPAB inter-union venture with the direct involvement of ICSU and CODATA. The Task Group has reviewed previous formal recommendations and has extended them in the light of more recent developments in the field of biomolecular NMR spectroscopy. Drafts of the recommendations presented here have been examined critically by more than 50 specialists in the field and have gone through two rounds of extensive modification to incorporate suggestions and criticisms.

TABLE OF CONTENTS

INTRODUCTION

Solution-state NMR spectroscopy is used widely to determine the structures of peptides, proteins, and protein-ligand complexes, as well as those of nucleic acids and their complexes with proteins, drugs, and other molecules. As the field has developed, a certain consensus has evolved on the presentation of NMR solution structures. This has been helped indirectly by guidelines established for depositing primary experimental data and resulting structures in data banks such as the Protein Data Bank (ref. 1), BioMagResBank (ref. 2), Nucleic Acid Database (ref. 3), and by conventions used for inclusion in abstracting services, for example, *Macromolecular Structures* (ref. 4). In consideration of the accumulated experience over the past few years in presenting NMR structures, the time appeared to be appropriate for a formal examination of

reporting conventions used in the past and for the development of a set of generally accepted guidelines for the future. With these goals in mind, the present Task Group was convened as an IUPAC/IUBMB/IUPAB Inter-Union venture with financial support from ICSU and CODATA. The present recommendations build upon earlier recommendations for biochemical nomenclature (ref. 5), the presentation of proton (ref. 6) and non-proton (ref. 7) NMR data for publication, and parameters and symbols for use in NMR (ref. 8).

1. Definition of the system studied

1.1. Names of the molecules used

It is helpful to use a notation that specifies the molecule type, natural species of origin, and fragment (if applicable): for example, "fragment consisting of residues 10—218 of murine Blk, a B-cell specific protein tyrosine kinase (PTC)." Note that recommendations for the naming of a variety of biological macromolecules have been published (ref. 9—14). Authors are encouraged to include identifiers used by publicly accessible databases (ref. 15), as these may be helpful for cross-referencing purposes. Such identifiers include the accession codes used by the Protein Identification Resource (ref. 16), SWISS-PROT (ref. 17), the Enzyme Commission of IUBMB (ref. 18), and the Chemical Abstracts Service (ref. 19).

1.2. Source

- Genus, species, strain or variant of gene (cloned or synthetic).
- Expression vector and host.
- If chemically synthesized, a description of the methods used.

1.3. Evidence for homogeneity and chemical identity

- Confirmation of chemical identity (e.g., from mass spectrometry or chemical sequencing).
- Analytical method used to establish chemical homogeneity.

1.4. Sequence or chemical structure

- Full sequence of the molecule (or reference to it).
- Unambiguous definition of the sequence numeration of the molecule studied.
- Description of additional covalent linkages and their locations: *e.g.*, disulfide bridges, covalently attached cofactors, or metal ions.
- Description of any differences between the system studied and the naturally-occurring biomolecule (*e.g.*, fragment with additional or fewer residues at either end of the biopolymer, post-transcriptional modifications in nucleic acids or post-translational modifications in proteins, or lack thereof).
- Specification of noncovalent cofactors or prosthetic groups.
- Quaternary structure (number and kind of subunits; symmetry, if known).

1.5. Solution conditions

- Specification of the solvent constituents and their isotopic compositions.
- Concentration of each solute component, including buffers, salts, antibacterial agents, *etc.*
- Temperature and pressure along with methods used for their measurement.
- Value of the pH, or pH* for uncorrected pH meter readings in ^2H_2O. Note that p^{2}H (or "pD") applies only to measurements made with electrodes filled with ^2H_2O.
- Comment on special measures taken to minimize self-aggregation (if applicable).
- Type of sample cell used.

1.6. Isotope-labeled proteins and nucleic acids

Essential information would include descriptions of the methods used to prepare the labeled molecules and to determine the positions and levels of isotopic labeling in the system used. Percent incorporation should be indicated when known, as illustrated in the following examples.

- Uniform (random) labeling with ^{15}N (extent of incorporation unknown): [*U*-^{15}N]-ribonuclease.
- Uniform (random) labeling with ^{13}C, ^{15}N, and ^{2}H (labeling levels known): [*U*-98% ^{13}C; *U*-90% ^{15}N; *U*-65% ^{2}H]-d(GCGCAATTGCGC).
- Residue-selective labeling: [98% ^{15}N]-Cys rubredoxin.
- Site specific labeling: [95% $^{13}C^{\alpha}$]-Trp28 lysozyme.
- Insertion of residues with natural isotopic composition into otherwise uniformly labeled proteins or nucleic acids are designated by *NA* for natural abundance: [*U*-98% ^{13}C; *NA*-F,Y,W]-lysozyme.

2. Atom identifiers for reporting chemical shift assignments

The 1983 IUPAC-IUB recommendations for peptides and proteins (ref. 21) were never widely adopted by the protein science community. Thus the present recommendations for peptides and proteins follow the earlier IUPAC-IUB "tentative rules" of 1969 (ref. 20) currently employed by biomolecular databases, with extensions and clarifications as required for NMR data. These rules designate all atoms by Greek letters (or Roman counterparts) and employ the main chain precedence rule for numbering prochiral sites. The present recommendations for nucleic acids follow the IUPAC-IUB nomenclature (ref. 22), with three exceptions. First, notation for the hydrogens at the C5′ position in ribose and deoxyribose follows the widespread use of H5′ (for *pro-S*) and H5″ (for *pro-R*), rather than H5′1 and H5′2 (the latter notation (ref. 22) is inconsistent with the numbering convention used with amino acids, in which such atoms would be designated as 2 and 3). Second, H2′ (for *pro-S*) and H2″ (for *pro-R*) are used to designate the hydrogens at the C2′-position in deoxyribose. Third, the notation proposed for hydroxyl groups of ribose and deoxyribose rings in reference 22 is adapted from that used with amino acids, so that the hydrogen atoms are denoted by $H^{O2'}$, $H^{O3'}$, and $H^{O5'}$, as appropriate, whereas the hydroxyl oxygens are denoted by $O^{5'}$, etc.

In cases not covered by the present recommendations (*e.g.*, unusual or modified amino acid residues or nucleic acid bases), authors are encouraged to refer to notation in use by the databases for such groups. If new atom notation needs to be introduced, it will be helpful if authors define this by providing stereospecific diagrams as appropriate.

2.1. Proteins and peptides

2.1.1. Standard nomenclature

Figure 1 defines the atom-naming conventions recommended here for proteins. Coordinate files should contain all atoms, including hydrogen atoms. The 1969 recommendations (ref. 20) did not deal explicitly with nomenclature for hydrogen atoms, and the rules are subject to different interpretations. The interpretation recommended here (Fig. 1) is that at any position of the side chain, the attached atom leading to the main chain always has the highest priority. For cyclic amino acids such as proline, the priority leads out from the main chain C^{α} atom rather than from the N atom (i.e., the priority is given by $C^{\alpha} > C^{\beta} > C^{\gamma} > \ldots > N$). In the peptide backbone, the nitrogen is denoted by N, its attached hydrogen by H, and the carbonyl carbon and oxygen by C and O, respectively (ref. 20). In cases where the single letter designations are ambiguous, they may be primed (e.g., N′, C′, O′) (ref. 20); C′ is used quite widely to denote the peptide backbone carbonyl carbon. Although H alone is used by the biomolecular data banks (PDB and BMRB), it is recommended here that the symbol H^{N} (in widespread use in the NMR community) be used as the unambiguous designator for the backbone amide hydrogen. The hydrogens of an N-terminal amine are designated as H^{1}, H^{2}, and H^{3} (when protonated) or H^{1} and H^{2} (when unprotonated). The oxygens of a C-terminal carboxyl or carboxylate are O′ and O″, and the hydrogen of the carboxyl is H″. IUPAC rules (ref. 18) label the hydrogens on methyl groups and protonated amines by "1", "2", and "3", in the conventional way.

Thus, for example, the atom designations in amino acid side chains would be: $H^{\beta 1}$, $H^{\beta 2}$, and $H^{\beta 3}$ for the three methyl hydrogens of alanine; $H^{\zeta 1}$, $H^{\zeta 2}$, and $H^{\zeta 3}$ for the three amine hydrogens of a protonated lysine; and $H^{\zeta 1}$ and $H^{\zeta 2}$ for the two amine hydrogens of a neutral lysine.

In situations where Greek letters are unavailable (for example, on some computer displays), they may be replaced by upper case Roman letters (α = A, β = B, γ = G, δ = D, ε = E, ζ = Z, η = H). When superscripts are not feasible, the full descriptor can be placed on a single line. When a full descriptor is used in a subscript, its components are displayed on a single line: *e.g.*, $^3J_{H'H\alpha}$ for a coupling constant or $d_{H\alpha H\beta}$ (commonly abbreviated $d_{\alpha\beta}$) for a distance detected by a nuclear Overhauser effect (NOE). The computer representation for ′ is a single quote and for ″ is two single quote symbols.

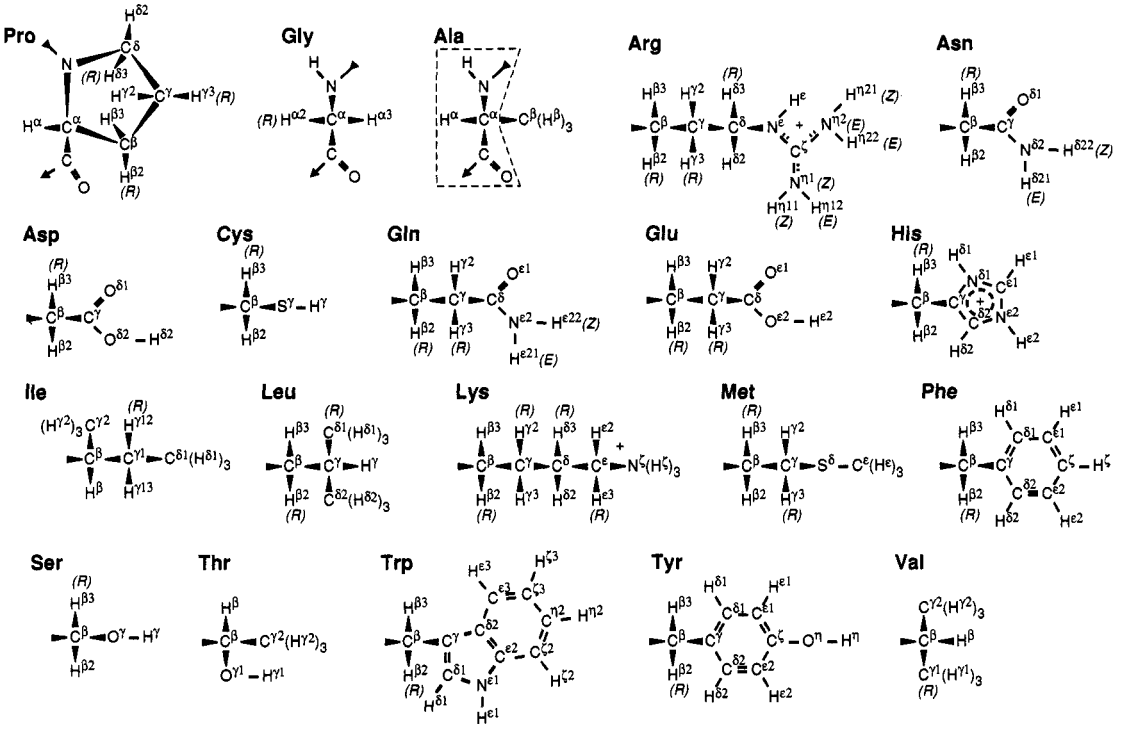

Fig. 1. *Recommended atom identifiers for the twenty common amino acids follow the 1969 IUPAC-IUB guidelines* (ref 20). Backbone atoms are shown for Pro, Gly, and Ala but not for the other L-amino acids (where they correspond to those bounded by the dashed line in the Ala structure). Greek letters are used as atom identifiers. The C^{α} or the substituent closer to C^{α} (in the order C^{α}, C^{β}, C^{γ}, . . .) takes precedence over atoms in branches in defining stereochemical relationships. For example, if tetrahedral carbon *C* has four substituents *X*, *Y*, *Z*, and *Z′* (with priority $X > Y > Z = Z'$; i.e., *Z* and *Z′* are diastereotopic substituents designated provisionally as unprimed and primed), their numbering is derived as follows: if one sights down the X—C axis (with the X atom toward the viewer), the equivalent atoms, Z and Z′, are designated Z_2 and Z_3, such that Y, Z_2, and Z_3 follow a clockwise orientation. The side-chain -NH_2 nitrogens of Arg are designated as $N^{\eta 1}$ and $N^{\eta 2}$ by their relationship (*cis* or *trans*, respectively) to C^{δ}. The hydrogen atoms of the side-chain -NH_2 groups of Asn, Gln, and Arg are distinguished by numbers (1 or 2) on the basis of their relationship (*cis* or *trans*, respectively) to the heavy atom three bonds closer to the main chain (C^{β} for Asn, C^{γ} for Gln, N^{ε} for Arg). Thus, each -NH_2 hydrogen of Arg is distinguished by two numbers, the first indicating the nitrogen to which it is attached and the second indicating the stereochemistry of the hydrogen itself. Numbering of Phe and Tyr rings gives higher priority to the atom with the smaller absolute value of the χ^2 torsion angle (ref. 20). For example, the ring carbons of Phe and Tyr lying in the plane with the smaller χ^2 torsion angle are designated as $C^{\delta 1}$ and $C^{\varepsilon 1}$. Indicated for reference in parentheses are the *pro-R*/*pro-S* designations for prochiral tetrahedral groups (with only the *pro-R* indicated as "*R*") (ref. 25, 26) and the *E*/*Z* designations for planar groups (ref. 27, 28).

2.1.2. Notation for assignment ambiguities

It is important to distinguish clearly between resonances that have complete stereochemical assignments and those that do not. Resonances with complete stereospecific assignments are denoted by the symbols shown in Fig. 1. Resonances having ambiguous assignments are denoted by special symbols, as defined below. Knowledge about particular kinds of assignment ambiguity (stereospecific and non-stereospecific) can be important in structure determinations. Such information defines, for example, what kind of pseudoatoms (see **3.1.** below) need to be invoked in a structure determination.

2.1.2.1. Non-stereochemical ambiguity

A slash (/) should be used to indicate ambiguity in assignments. The slash is used to join the ambiguous atom designators and chemical shifts. For example, four chemical shifts (6.84, 6.90, 7.35, and 7.42 ppm) associated ambiguously with four hydrogens of a given tryptophan ring ($H^{\varepsilon 3}$, $H^{\zeta 2}$, $H^{\zeta 3}$, $H^{\eta 2}$) are represented as: $H^{\varepsilon 3/\zeta 2/\zeta 3/\eta 2}$ = 6.84/6.90/7.35/7.42 ppm. A single resonance at 4.30 ppm that is known to arise from H^{α} of either Thr 10 or Thr 28 would be indicated as T10 H^{α}/T28 H^{α} = 4.30 ppm. Because signals from individual hydrogens on methyl groups and amines are ordinarily not resolved by NMR spectroscopy, the signals could be represented by the slash rule (for example, $H^{\beta 1}/H^{\beta 2}/H^{\beta 3}$, for the hydrogens of an alanine methyl group), but it is simpler when designating spectral features just to omit the numeration (H^{β}, for the alanine methyl).

2.1.2.2. Stereochemical ambiguity involving prochiral centers

The nomenclature described in **2.1.2.1.** also applies to situations in which two diastereotopic substituents have not been assigned stereospecifically. For example, resonances at 2.44 ppm and 3.12 ppm that were ambiguously assigned to the $H^{\beta 2}$ and $H^{\beta 3}$ atoms of a given residue are reported as $H^{\beta 2/\beta 3}$ = 3.12/2.44 ppm. A single peak at 2.44 ppm that could be assigned to $H^{\beta 2}$, $H^{\beta 3}$, or both is reported as $H^{\beta 2/\beta 3}$ = 2.44 ppm. If it has been established that both have the same chemical shift, they are reported as: $H^{\beta 2}$ = 2.44 ppm, $H^{\beta 3}$ = 2.44 ppm.

2.1.2.3. Other stereochemical ambiguity

When flips of the symmetrical ring of a phenylalanine or tyrosine are slow on the chemical shift time scale, signals from the δ- and ε-atoms can have non-equivalent shifts. In such cases, information about assignments relative to the χ^2 torsion angle is needed to define the specific atom designators 1 and 2. This information often is not known in advance of a full structure determination, so it is useful to specify this level of ambiguity again with the slash rule. For example, if proton signals at 6.60, 6.95, 7.12, and 7.22 ppm have been identified with a particular, slowly-rotating, tyrosine ring at position 12, with no further information, they are reported as Y12 $H^{\delta 1/\varepsilon 1/\delta 2/\varepsilon 2}$ = 7.12/6.95/7.22/6.60 ppm. If, in addition to the above, it is known that the signals at 7.12 and 7.22 ppm arise from δ-hydrogens, then the signals are reported as Y12 $H^{\delta 1/\delta 2}$ = 7.12/7.22 ppm and Y12 $H^{\varepsilon 1/\varepsilon 2}$ = 6.60/6.95 ppm. If it has been further determined that the signals at 6.95 and 7.12 ppm correspond to hydrogens on the same side of the ring, they are reported as $H^{\delta x}$ = 7.22 ppm, $H^{\delta y}$ = 7.12 ppm, $H^{\varepsilon x}$ = 6.60 ppm, $H^{\varepsilon y}$ = 6.95 ppm.

In cases where the signals from the side-chain amide hydrogens of Asn or Gln, or the guanidino NH_2 nitrogens or hydrogens of Arg are known to be coincident, the same chemical shift is assigned to both atom designators. In cases where information for unambiguous, individual assignments is lacking, the ambiguity is represented by the "slash rule". For example, ambiguous assignments would be indicated by Gln $H^{\varepsilon 21/\varepsilon 22}$, Arg $H^{\eta 11/\eta 12}$, or Arg $H^{\eta 11/\eta 12/\eta 21/\eta 22}$.

2.2. Nucleic acids

2.2.1. Standard nomenclature

Figure 2 shows the recommended designators for the atoms of ribose, deoxyribose, and the common bases (ref. 22). The IUPAC-IUB notation for the methyl group of T is $C7(H7)_3$. The phosphorus-bound oxygens are named O3′, O5′, OP1, and OP2. In the CIP rules (Cahn-Ingold-Prelog, ref. 25, 26), OP1 corresponds to *pro-R* and OP2 to *pro-S*.

2.2.2. Notation for assignment ambiguities

The slash rule is used to designate diastereotopic substituents of prochiral centers that have not been assigned stereospecifically. For example, if H5′ and H5″ of a given ribose are identified at 4.12 ppm and 4.53 ppm but have not been assigned individually, they are designated as H5′/H5″ = 4.53/4.12 ppm.

3. Additional atom identifiers for reporting conformational constraints and structures

Structural constraints involving atoms for which the NMR signals have not been assigned individually are frequently employed in the determination of macromolecular structures from NMR data. In such cases, special symbols are used to identify the structural constraints. These may take the form of pseudoatoms that represent groups of atoms, or of symbols that link specific constraints with NMR chemical shifts.

3.1. Pseudoatom nomenclature

NOE constraints to groups of atoms for which the resonances have not been assigned individually are commonly specified by pseudo-structures in which certain atoms have been joined together into pseudoatoms. Relevant pseudoatoms for the standard amino acid residues have been defined (ref. 30) and are in common use in the NMR literature. An NOE to one or all protons in the group is measured relative to a location central to the group (ref. 31). All pseudoatoms needed to describe proteins and nucleic acids can be specified in terms of two letters, Q and M, in conjunction with the usual atom identifiers (Roman letters are used instead of Greek letters for atom positions within amino acids, and R is used to denote a ring). M describes the location of methyl groups, and Q is used in all other situations. *Table 1 lists all pseudoatoms in the* common amino acids and nucleotides.

3.2. NOE constraints that can be identified by chemical shifts

NOEs used for structural constraints that are attributed to pairs of diastereotopic or

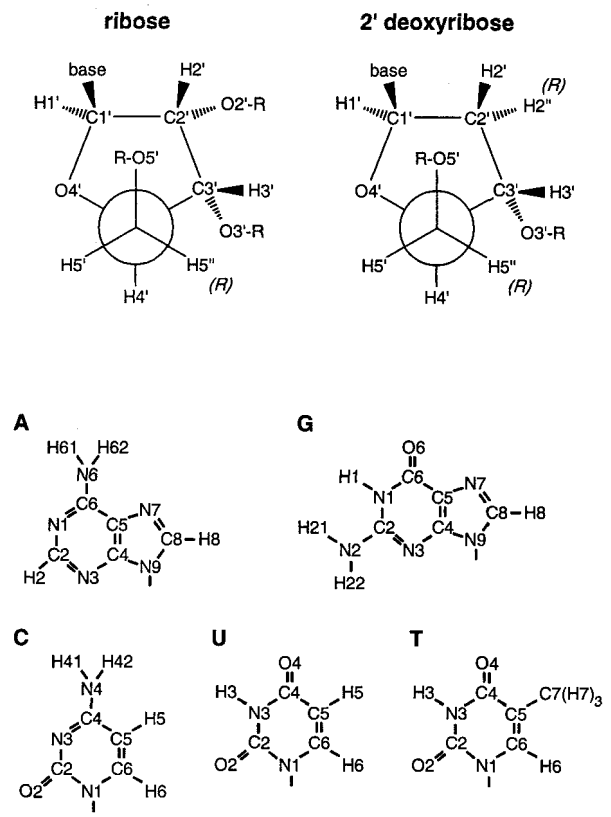

Fig. 2. Nomenclature, structures, and atom numbering for the sugars and the bases contained in common nucleotides (ref. 22). The identifiers shown here for the hydrogen atoms largely follow previous recommendations as discussed in section **2.** of the text. "*R*" indicates *pro-R*.

TABLE 1. Pseudoatoms for the Common Amino Acids and Nucleotides Used in the Determination of Structures of Proteins and Nucleic Acids from NMR Data.[a]

Residue	Pseudoatom	^{1}H atoms represented[b]
Gly	QA	α-methylene
Ala	MB	β-methyl
Val	MG1, MG2	γ1-, γ2-methyl
	QG	all six γ-methyl
Ile	MG, MD	γ2-, δ1-methyl
	QG	γ1-methylene
Leu	MD1, MD2	δ1-, δ2-methyl
	QB	β-methylene
	QD	all six δ-methyl
Pro	QB, QG, QD	β-, γ-, δ-methylene
Ser, Asp, Cys, His, Trp	QB	β-methylene
Thr	MG	γ2-methyl
Asn	QB	β-methylene
	QD	δ2-amido
Glu	QB, QG	β-, γ-methylene
Gln	QB, QG	β-, γ-methylene
	QE	ε2-amido
Lys	QB, QG, QD, QE	β-, γ-, δ-, ε-methylene
	QZ	ζ-amino
Arg	QB, QG, QD	β-, γ-, δ-methylene
	QH1, QH2	η11 and η12, η21 and η22
	QH	all four η-guanidino
Met	QB, QG	β-, γ-methylene
	ME	ε-methyl
Phe, Tyr	QB	β-methylene
	QD, QE	δ1- and δ2-ring, ε1- and ε2-ring
	QR	all ring
β-D-ribose	Q5′	5′-methylene
2′-β-D-deoxyribose	Q2′	2′-methylene
	Q5′	5′-methylene
A	Q6	6-amino
C	Q4	4-amino
G	Q2	2-amino
T	M7	7-methyl

[a] Ref. (28). [b] See Figures 1 and 2.

symmetry-related protons which have non-degenerate chemical shifts and have not been assigned individually, can be distinguished unambiguously on the basis of the chemical shifts. For example, an NOE observed at 3.12 ppm to one of a pair of methylene proton resonances at 2.44 ppm and 3.12 ppm, which were assigned ambiguously to the $H^{\beta 2/\beta 3}$ atoms of Tyr57 is reported as an NOE to Tyr57 $H^{\beta 2/\beta 3}$ (3.12 ppm). Alternatively, the symbols "*L*" and "*H*" can be used to represent the resonances at lower frequency (historically referred to as "downfield") and higher frequency (historically "upfield"), respectively. In the above example, the NOE would then be to Tyr57 $H^{\beta L}$.

4. Nomenclature for specifying conformation

Torsion angles are defined in terms of the substituent with highest preference on each of the two atoms that span the rotatable bond (ref. 20, 35). The IUPAC-IUB convention (ref. 20, 22) specifies that the main chain atoms of a peptide or nucleic acid take precedence over others. For example, in a peptide, the backbone

carbon and nitrogen atoms take precedence over other heavier atoms such as the carbonyl oxygen or cysteine C^{β} (ref. 20). Otherwise atom preferences follow standard conventions (ref. 25, 26). A given torsion angle θ about the B—C bond of a molecule A—B—C—D (where A is the atom with highest preference attached to B, and D is the atom with highest preference attached to C) is the angle between the planes containing A—B—C and B—C—D, or alternatively the angle between the projections of B—A and C—D onto a plane normal to B—C. The torsion angle is written in full as θ (A,B,C,D). When the projections of the two bonds B—A and C—D coincide (eclipsed or *cis* conformation), $\theta = 0°$; when the projections of the two bonds B—A and C—D are opposed (*trans* conformation), $\theta = 180°$. If for a given torsion angle θ that is neither 0° nor 180°, when looking (in either direction) along the central bond B—C, the minimal rotation of the *front* bond that would be required to achieve the eclipsed conformation is clockwise, θ is considered to be positive; if the required minimal rotation is counter-clockwise, θ is considered to be negative. As defined below, various abbreviations are used to denote specific torsion angles in proteins and nucleic acids; in other situations (or to avoid any ambiguity) the four atoms that determine the torsion angle should be specified.

4.1. Polypeptides

4.1.1. Backbone torsion angles: ϕ_i, ψ_i, *and* ω_i

The ϕ_i torsion angle describes rotations about the N_i—C^{α}_i bond (relevant four atoms C'_{i-1}—N_i—C^{α}_i—C'_i), and the ψ_i torsion angle describes rotations about the C^{α}_i—C'_i bond (relevant atoms N_i—C^{α}_i—C'_i—N_{i+1}). The ω_i torsion angle describes rotations about the peptide bond, C'_{i-1}—N_i (relevant atoms C^{α}_{i-1}—C'_{i-1}—N_i—C^{α}_i).

4.1.2. Side-chain torsion angles: χ^j_i

Side-chain torsion angles for residue i are designated by χ^j_i, where j represents the rotatable bond: $j = 1$ for the C^{α}—C^{β} bond (where the relevant atoms are N—C^{α}—C^{β}—X); $j = 2$ for the C^{β}—C^{γ} bond (where the relevant atoms are C^{α}—C^{β}—C^{γ}—X), *etc.* Figure 3 shows how the χ^1 angle is related to the angle between the projections of the bonds C^{α}—H^{α} and C^{β}—H^{β} (or C^{β}—$H^{\beta 2}$) for the common amino acids. In amino acids with branched side chains, two superscripts are used, the first to indicate the position of the bond and the second to indicate the branch (*e.g,* the C^{α}—C^{β}—$C^{\gamma 1}$—$C^{\delta 1}$ torsion angle of isoleucine at residue position i is denoted by $\chi^{2,1}_i$) (ref. 20).

4.1.3. Specification of turns, disulfide bonds, and proline rings

Turns can be defined unambiguously by the torsion angles involved, i.e., ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2}, and ψ_{i+2} (ref. 36). Disulfide bridges between residues i and j can be described by the torsion angles χ^1_i, χ^2_i, χ^3_i, χ^2_j, and χ^1_j; in addition, it is recommended that the handedness be indicated by *R* or *S* (ref. 37). Proline ring puckers can be defined by the torsion angles χ^1_i, χ^2_i, χ^3_i and χ^4_i (see caption to Fig. 4); alternatively, use can be made of the fact that the value of χ^1_i alone is indicative of the two major pucker forms (ref. 38) (Fig. 4).

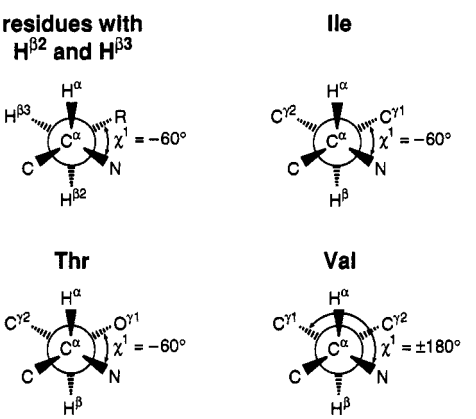

Fig. 3. Definition of the χ^1 angle for the common amino acids. Note that although the H^{α}—C^{α}—C^{β}—$H^{\beta(\beta 2)}$ torsion angles are equivalent in all the examples shown, the value of χ^1 for Val differs from those of the other amino acids as a consequence of the definition of this torsion angle.

4.2. Nucleic acids

Notation in current use in the nucleic acid field is presented here. In line with current usage, multiple notations are provided in certain instances. Older conventions for defining torsion angles can be found in reference 22.

4.2.1. Backbone and glycosidic torsion angles

For residue i the backbone (α, β, γ, δ, ε, ζ) and glycosidic (χ) torsion angles for purine (Pu) and pyrimidine (Py) bases are defined as follows (ref. 22) (see also Fig. 5):

α	$O3'_{i-1}$— P_i— $O5'_i$— $C5'_i$
β	P_i— $O5'_i$— $C5'_i$— $C4'_i$
γ	$O5'_i$— $C5'_i$— $C4'_i$— $C3'_i$
δ	$C5'_i$— $C4'_i$— $C3'_i$— $O3'_i$
ε	$C4'_i$— $C3'_i$— $O3'_i$— P_{i+1}
ζ	$C3'_i$— $O3'_i$— P_{i+1}— $O5'_{i+1}$
χ(Py)	$O4'_i$— $C1'_i$— $N1_i$— $C2_i$
χ(Pu)	$O4'_i$— $C1'_i$— $N9_i$— $C4_i$

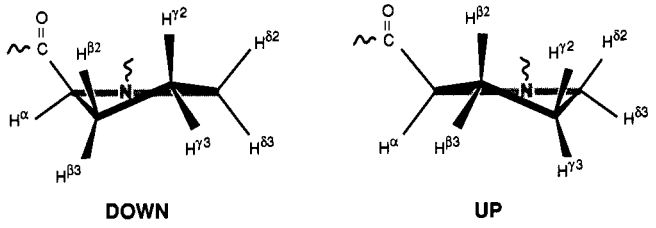

Fig. 4. Ring pucker in the proline ring (ref. 38). In the DOWN conformation, the torsion angles χ^1 and χ^3 are positive and χ^2 and χ^4 are negative; in the UP conformation, the torsion angles χ^1 and χ^3 are negative and χ^2 and χ^4 are positive. Typical values for the torsion angles also depend on whether the Xxx—Pro peptide bond is *cis* or *trans*. Average values (in degrees) found from a number of X-ray structures of proteins for χ^1, χ^2, χ^3, χ^4 were (ref. 38): *trans* DOWN, 18, –29, 17, –4; *cis* DOWN, 26, –38, 18, –7; *trans* UP, –27, 35, –34, 20; *cis* UP, –21, 36, –36, 17.

4.2.2. Sugar pucker: torsion angles and angle of pseudorotation

Each of the sugar torsion angles ν_0, ν_1, ν_2, ν_3 and ν_4 is defined by four atoms as follows (ref. 22) (see also Fig. 5):

ν_0	C4′— O4′— C1′— C2′
ν_1	O4′— C1′— C2′— C3′
ν_2	C1′— C2′— C3′— C4′
ν_3	C2′— C3′— C4′— O4′
ν_4	C3′— C4′— O4′— C1′

Note that ν_3 and δ, which characterize torsion angles about the same bond (C3′— C4′), will have different values for any given conformation.

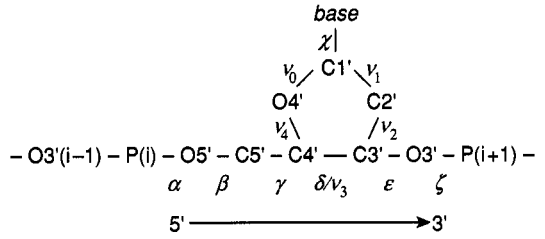

Fig. 5. Designation of the torsion angles in the sugar-phosphate backbone (α, β, γ, δ, ε, ζ), the glycosidic bond (χ), and the endocyclic torsion angles in the sugar ring (ν_0 — ν_4) (ref. 22).

The sugar is generally non-planar, and it is recommended that the sugar ring conformation (pucker) be described by specifying two parameters: the phase angle of pseudorotation (P) and the puckering amplitude ψ_m. In practice, the pseudorotation parameters can be determined from measurements of the three-bond J-couplings between the protons attached to C1′, C2′, C3′, and C4′. The following equations characterize relationships among the torsion angles defined by these protons and the pseudorotation parameters (ref. 39, 40):

$$\phi_{1'2'} = 121.4 + 1.03\ \psi_m \cos(P - 144°)$$
$$\phi_{1'2''} = 0.9 + 1.02\ \psi_m \cos(P - 144°)$$
$$\phi_{2'3'} = 2.4 + 1.06\ \psi_m \cos(P)$$
$$\phi_{2''3'} = 122.9 + 1.06\ \psi_m \cos(P)$$
$$\phi_{3'4'} = -124.0 + 1.09\ \psi_m \cos(P + 144°)$$

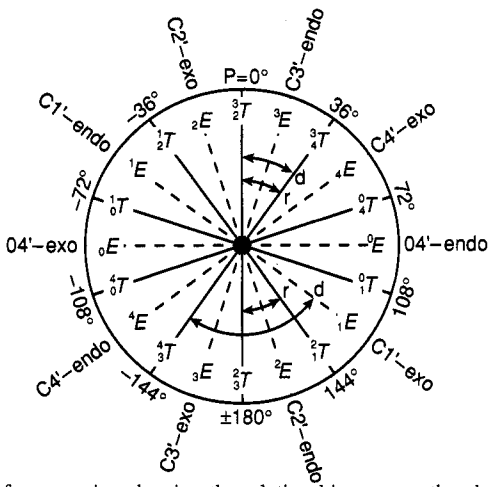

Fig. 6. Pseudorotation cycle of the furanose ring showing the relationships among the phase angle of pseudorotation (*P*), the envelope (*E*) and twist (*T*) notation, and the *endo* and *exo* notation. Conformations corresponding to the northern half are designated as *N*-type, and those corresponding to the southern half are designated as *S*-type. Ranges of the *P* values usually observed experimentally for the *N* and *S* conformations are represented by '*r*' for ribo- and '*d*' for 2′-deoxyribofuranose rings of β -D-nucleosides and nucleotides (ref. 22). Note that the symmetrical twist (*T*) conformations are found at even multiples of 18° and that the symmetrical envelope conformations are found at odd multiples of 18°.

In these equations the non-equilateral character of the sugar ring is accounted for. The ϕ's can be derived from the *J*-couplings by means of appropriate Karplus equations.

Relationships among the phase angle of pseudorotation and the alternative *E/T* and *endo/exo* notations are illustrated by the pseudorotation wheel in Fig. 6. In the *E/T* nomenclature, the puckered forms are designated by *E* (envelope form) and *T* (twist form) (ref. 22). For example, $P = +18°$ is equivalent to C3′-*endo* or ${}^{3}E$, and $P = -162°$ is equivalent to C3′-*exo* or ${}_{3}E$. Symmetrical twist conformations, with two atoms at equal distance with respect to the plane defined by the three remaining ring atoms, can also be represented; for example, $P = 0°$ is equivalent to C3′-*endo*/ C2′-*exo*, or to ${}^{3}_{2}T$.

4.2.3. Value ranges for torsion angles specifying commonly observed conformations

In practice, torsion angles are often not known exactly, but can be assigned to particular conformational regions. The notation used most frequently by spectroscopists and X-ray crystallographers for ranges of torsion angles is *cis* , *trans*, – *gauche,* and *+gauche* as shown in Fig. 7. The IUPAC-IUB Commission on Nucleotide Nomenclature recommended the Klyne-Prelog notation (ref. 22), with ± *synperiplanar* (± *sp*), ± *synclinal* (± *sc*), ± *anticlinal* (± *ac*), and ± *antiperiplanar* (± *ap*) as defined in Fig. 7. The terms *syn* (0° ± 90°) and *anti* (180° ± 90°) have special meanings in nucleotide chemistry in that they are used to define the orientation of the base with respect to the sugar. Table 2 lists the predominant conformations of nucleic acids and the associated ranges of torsion angles.

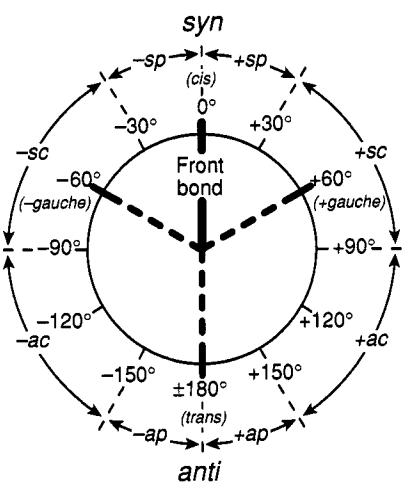

Fig. 7. Relationships among different terminologies used for the approximate description of a torsion angle in terms of conformational regions. Torsion angles are defined by holding the front bond at zero position, as indicated in the figure, and by measuring the angle between the front bond and the back bond while looking along the central bond. The Klyne-Prelog convention (whose ranges are abbreviated as: ± *sp*, ± *synperiplanar*; ± *sc*, ± *synclinal*; ± *ap*, ± *antiperiplanar*; ± *ac*, ± *anticlinal*) has been recommended by the IUPAC-IUB Commission on Nucleotide Nomenclature (ref. 22). However, spectroscopists commonly use a different nomenclature which specifies four 60° regions: *cis* (0° ± 30°), *trans* (180° ± 30°), *+gauche* (60° ± 30°) and *–gauche* (–60° ± 30). The terms *syn* (0° ± 90°) and *anti* (180° ± 90°) are discussed in the text.

TABLE 2. Notation Commonly used to Describe the Major Conformational Forms of Polynucleotides in Terms of Torsion Angle Ranges Obtained from Fiber Diffraction Measurements.[a]

	α	β	γ	δ	ε	ζ	χ
A-RNA	g^- - *sc*	*t* + *ap*	g^+ + *sc*	g^+ + *sc*	*t* - *ap*	g^- − *sc*	*anti* − *ap*
A-DNA	g^- − *sc*	*t* +*ap*	g^+ +*sc*	g^+ +*sc*	[*t*] − *ac*	g^- − *sc*	*anti* − *ap*
B-DNA	g^- − *sc*	[*t*] -*ac*	g^+ +*sc*	*t* +*ap*	*t* +*ap*	[g^-] − *ac*	*anti* − *ac*
Z-DNA G[b]	g^+ +*sc*	*t* − *ap*	*t* +*ap*	g^+ +*sc*	g^- − *sc*	[[g^+]] +*ac*	*syn* +*sc*
Z-DNA C[c]	[[g^-]] − *ac*	*t* − *ap*	g^+ +*sc*	[*t*] +*ac*	[[g^-]] − *ac*	g^- − *ac*	*anti* − *ap*

[a] For reference, the alternative Klyne-Prelog notation (ref. 22) is given below each recommended symbol. The *cis* (*c*), *trans* (*t*), *+gauche* (g^+), *−gauche* (g^-) notation does not provide a definition for the torsion angle ranges 90° to 150° and − 90° to − 150° (Fig. 7). Several torsion angles in standard polynucleotide conformations fall outside the ranges defined by these symbols. This information is also given in the table in that angle values outside the range by less than 10° are enclosed by brackets and those outside the range by more than 10° are enclosed by double brackets. In these cases, the Klyne-Prelog notation provides a more accurate representation of the torsion angle range. [b]Guanine residue in Z-DNA. [c]Cytosine residue in Z-DNA.

5. NMR data acquisition, processing, and referencing

This section gives recommendations on how to present information needed to provide a comprehensive description of the experiments used in a structure determination.

5.1. Data acquisition

The following information describes the data collection process. In particular, a complete account is needed for experiments employed in the collection of conformational constraints.

- NMR hardware used. Customarily the magnetic field is specified by the resonance frequency of a particular nucleus, usually ^{1}H; other minimal information includes the type of console, type of probe head and the size and nature of the sample tube/container. Additional details might be desirable for the description of novel experimental protocols or home-built systems.

- Protocols for the acquisition of raw data. This normally includes a diagram of the pulse sequence and/or an ASCII file with the code for the pulse program, or a reference if previously published routines were used. It is important to include information on details of solvent suppression, timing parameters, r.f. power levels (or equivalent 90° pulse duration), phase cycling, pulse shapes, any magnetic field gradients (magnitudes and durations), the modes of acquisition for each dimension (including method for quadrature detection), the recycle delay, and the total time required for data acquisition.
- Description of time domain data. This usually includes acquisition times, spectral widths, possible use of non-linear sampling, and the number of real or complex data points in each time domain.

5.2. Data processing

A comprehensive description of data processing normally includes the following information:

- Software packages used.
- Enhancement of the time-domain data. This may include digital filtering, solvent filtering, convolution with a window function, zero filling, or linear prediction.
- Time-domain to frequency-domain conversion. The Fourier transform is the most common method, but alternatives may be employed, such as maximum entropy, maximum likelihood, or Bayesian analysis. The final data size and digital resolution in each dimension are part of a complete documentation.
- Enhancement of frequency-domain data. Common methods include symmetrization, baseline flattening, and ridge or solvent suppression.
- Methods used to extract spectral parameters such as resonance frequencies, peak volumes, and coupling constants.

5.3. Referencing of chemical shifts

According to general convention, NMR spectra are displayed with positive frequencies to the left and negative frequencies to the right of the zero-frequency reference (ref. 6, 7). The IUPAC Commission on Molecular Structure and Spectroscopy (ref. 6) has recommended that the ^{1}H signal of tetramethylsilane (TMS) be used as the primary reference for the resonance frequencies (and hence chemical shifts) of protons, and they propose to use the same signal as an indirect reference for other nuclei (ref. 41). The traditional primary chemical shift standard, however, for aqueous solutions has been the methyl signal of a water-soluble derivative of TMS. Following recent literature (ref. 42), it is recommended that the primary chemical shift standard for all nuclei in aqueous biological investigations be the methyl signal of internal 2,2-dimethylsilapentane-5-sulfonic acid (DSS) at low concentration. When it is inappropriate to use DSS as an internal standard, *e.g.*, in cases where DSS binds to the system under investigation, it is still desirable to use the position of DSS as the 0 ppm point. In such cases, reference 42 lists alternative reference compounds and corresponding referencing protocols. In such cases, precise details should be given for the reference used and the conversion factors employed. In the interest of clearly differentiating such chemical shifts from the default IUPAC recommendation for TMS as the reference (ref. 41), it is recommended that the notation δ_{DSS} be used. The chemical shifts of the methyl peaks in the two references are quite similar (ref. 43).

5.3.1. ^{1}H chemical shifts

The methyl resonance of internal DSS at low concentration directly provides the preferred reference signal for 0 ppm: notation, δ_{DSS} or δ(DSS).

5.3.2. Chemical shifts of other nuclei such as ^{2}H, ^{13}C, ^{15}N, and ^{31}P.

Whereas a number of direct and indirect approaches have been described for the referencing of chemical shifts of nuclei other than ^{1}H, the indirect referencing method (ref. 44) has become the preferred approach. Non-proton chemical shifts are referenced indirectly to the ^{1}H standard using conversion factors derived from ratios of NMR frequencies. The relative frequencies are designated by the symbol Ξ, with ^{1}H conventionally at exactly 100 MHz. Ξ values for the nuclei most commonly utilized in studies of proteins and nucleic acids are listed in Table 3. Thus, for example, the zero frequency for

TABLE 3. Relative Frequencies (Ξ) of Nuclei of Biomolecular Interest to be Used for Indirect Referencing.[a]

Nucleus	Secondary reference sample	Ξ /MHz	Reference
^{1}H		100.000000	by definition
^{2}H	DSS (internal)	15.3506088	ref. 45
^{13}C	DSS (internal)	25.1449530	ref. 42
^{15}N	liquid NH_3 (external)	10.1329118	ref. 42
^{31}P	$(CH_3O)_3PO$ (internal)	40.4808636	ref. 46

[a] These Ξ values were derived originally (as indicated in the table) from the ratio of the signal frequency of a secondary reference to that of internal DSS in D_2O as the primary reference.

^{13}C chemical shifts is obtained from the experimentally determined ^{1}H frequency for DSS by multiplying the latter frequency by the $^{13}C/^{1}H$ Ξ ratio, 25.1449530 /100.000000 = 0.251449530. In order to tie in with the chemical shift scales used in much of the previous literature, these conversion factors were derived from the $^{1}H/^{2}H$ frequency ratio of water, ^{13}C frequency of the methyl groups of internal DSS, the ^{15}N frequency of external liquid ammonia, and the ^{31}P frequency of internal trimethylphosphate, respectively (see reference 42 for details). It is recommended here that the relative frequencies (Table 3) be used as fixed constants not subject to further change and that they be used for data collected at all temperatures (with no applied temperature correction) and with the primary reference as either internal TMS (δ) or DSS (δ_{DSS}). It is anticipated that standardized Ξ values for additional nuclei of interest to biomolecular spectroscopists will be developed and published as recommendations (ref. 41).

6. Resonance assignments

Resonance assignments constitute a prerequisite for the commonly used structure determination procedures. Because the final quality of the structure depends to a large extent on the completeness of the underlying resonance assignments, their full description is important. Deposition of chemical shift assignments is imperative (see **10.** below). Quantitative information on the completeness of assignments obtained, including a listing of the missing assignments, is of prime interest. If, as part of automated assignment procedures, parameters are available that provide a measure for the reliability of a particular assignment, it is desirable that such parameters be presented along with the assignments themselves.

6.1. Sequence-specific assignment of polypeptide chains

The assignment process for polypeptides consists of sequence-specific assignments of backbone and side chain resonances.

6.1.1. Sequential backbone assignment

In most cases, a brief description should suffice, since present assignment strategies employ one or both of the two following approaches.

Assignments derived from sequential NOEs and ^{1}H- ^{1}H J-correlation. If samples with solution conditions (*e.g.*, pH, temperature, ionic strength) different from those reported in the assignment table were used in determining or validating assignments, this information needs to be noted. It is useful to report other information relevant to the assignment process, such as improvement of the spectral resolution by isotopic labeling, including $^{1}H/^{2}H$ exchange of amide groups. It is recommended that this include description of the types of *J*-correlation experiments used in the sequential assignment process.

Sequential assignments derived from homonuclear and heteronuclear J-correlations across peptide bonds. Although uniform double labeling with ^{15}N and ^{13}C is most commonly employed with this approach, other labeling patterns may be used. If non-uniform labeling was employed, important information includes a description of the labeling strategy and a list of the labeled proteins used. In cases where NOEs were used in parallel to establish or validate assignments, this information is of interest.

6.1.2. Amino acid side-chain assignment

Side chain assignments, although conceptually simple, are often complicated by spectral overlap, strong coupling, and rapid transverse relaxation. Various isotope labeling strategies and NMR techniques have been devised to overcome these problems. Unless novel methods have been employed, it should suffice to reference the approaches used. It is recommended that the report contain a list of the experiments used, the fraction of the side chain resonances that have been correlated through *J*-couplings to the backbone, and the extent to which NOEs have been used in the assignment or validation process.

6.2. Sequence-specific assignment of nucleic acids

The assignment process for nucleic acids consists of two phases. In one part, the sugar-phosphate backbone resonances are assigned, together with the purine H8 and the pyrimidine H5 and H6 resonances. This assignment usually is based on combined use of NOEs, ^{1}H-^{1}H *J*-correlation, and ^{1}H- ^{31}P *J*-correlation, and, in cases where isotopic labeling is used, on ^{1}H- ^{13}C, ^{13}C- ^{13}C, ^{13}C- ^{15}N, and ^{13}C- ^{31}P *J*-correlations. In the second part, the remaining base protons (adenine H2 and all exchange labile protons) are assigned, usually by NOE methods, but possibly also by isotopic labeling and *J*-correlation. For both parts, it is important to report how the resonance assignments were established and what assumptions regarding the conformation of the oligonucleotide were used in making assignments that involve NOEs. Further pertinent information includes temperature and solvent conditions and, if applicable, the level and pattern of isotopic enrichment.

6.3. Stereospecific assignment of diastereotopic substituents

It is useful to indicate in the assignment table which method was used to make each stereospecific assignment. Unless a novel approach was used, the description may consist of a brief reference to one of the methods in common use.

Analysis of J-couplings and short-range NOEs. It is useful to report the *J*-couplings and the experiments used for their measurement along with any NOEs utilized in making the stereospecific assignments. If computational methods, such as systematic grid searches, were used, they should be referenced or described briefly.

Stereoassignment by isotopic labeling. It is customary to report how the labeling was achieved (*e.g.*, isotopic composition of cell growth medium) and what the observed levels of enrichment were for the pertinent sites in the biomolecule.

Stereospecific assignments by reference to preliminary structures. Indicate the software package and any statistical criteria (confidence level) employed.

6.4. Conformational equilibria

The criteria used to deduce the existence of multiple states (as may result, for example, from partial folding, oligomeric heterogeneity, cofactor or ligand binding heterogeneity, peptide bond *cis/trans* isomerization, or hairpin-duplex equilibrium) are of prime interest. A separate designator is supplied for each individual state and associated with its NMR parameters. Fractional occupancies and interconversion rates for the states (or bounds on these) may be included if known.

7. Conformational features derived from diagnostic NMR parameters

7.1. Polypeptide secondary structure

It is a special feature of protein structure determination by NMR that the secondary polypeptide structure, including the connections between individual segments of regular secondary structure, can be determined early on in connection with the resonance assignments, before the complete structure calculation is even started. Reports on such identification of regular secondary structures and tight turns will be of interest also in the foreseeable future, for example, as a preliminary structural characterization of a novel protein or in connection with studies on protein folding. The information used for secondary structure identification is concisely documented in a survey diagram of the type shown in Fig. 8. The data in Fig. 8 include spin-spin coupling constants $^{3}J_{\mathrm{HNH}\alpha}$, sequential NOEs $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$, medium-range NOEs $d_{\alpha N}(i,i+3)$, $d_{\alpha\beta}(i,i+3)$, $d_{\alpha N}(i,i+4)$, $d_{NN}(i,i+2)$, and $d_{\alpha N}(i,i+2)$ (ref. 47), and conformation-dependent chemical shifts for the α-carbons, $\Delta\delta$ ($^{13}C^{\alpha}$). The diagram can be expanded readily for inclusion of additional parameters of diagnostic value for secondary structure determination. Figure 8 gives complete data for identification of helical structure (ref. 33) and N-caps in helices (ref. 49). Although segments of extended chain also can be

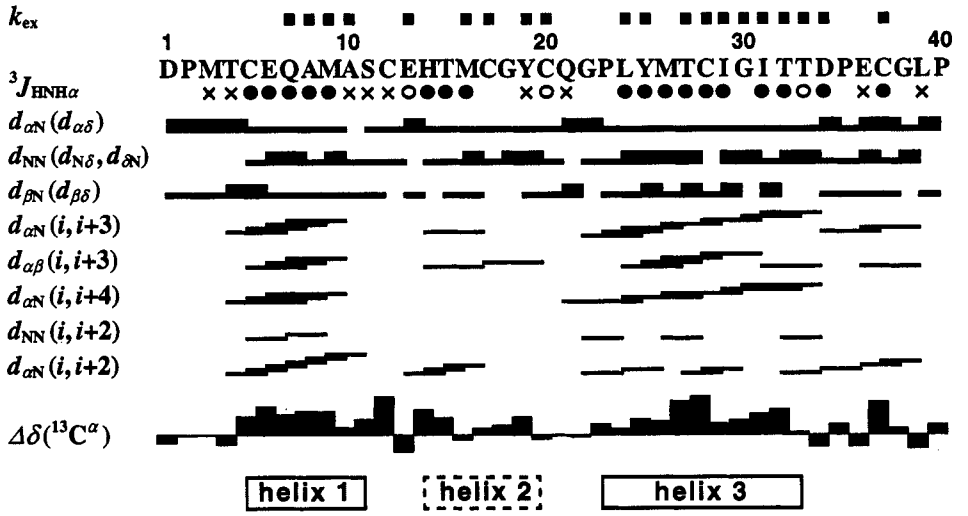

Fig. 8. Example showing methods recommended for presenting NMR data supporting the identification of regular secondary structure in proteins. The 40-residue protein, pheromone Er-2, is used as an illustration (ref. 48). Above the amino acid sequence, black squares identify residues with observably slow hydrogen exchange rates, k_{ex}, at the backbone amide (the conditions of the exchange experiment should be specified). Below the amino acid sequence, filled circles identify residues with $^3J_{HNH\alpha} < 6.0$ Hz, indicative of local α-type conformation; open circles correspond to $^3J_{HNH\alpha} > 8.0$ Hz, indicative of residues in extended chain conformation; crosses identify residues with $^3J_{HNH\alpha}$ values 6.0—8.0 Hz. For the sequential proton-proton NOE connectivities, $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$ ($d_{\alpha\delta}$, $d_{N\delta}$ and $d_{\beta\delta}$ for Xxx—Pro dipeptides, $d_{\alpha N}$, $d_{\delta N}$, and $d_{\beta N}$ for Pro—Xxx dipeptides), thick and thin bars indicate strong and weak NOE intensities, respectively. The observed medium-range NOEs $d_{\alpha N}(i,i+3)$, $d_{\alpha\beta}(i,i+3)$, $d_{\alpha N}(i,i+4)$, $d_{NN}(i,i+2)$, and $d_{\alpha N}(i,i+2)$ are indicated by lines connecting the two residues that are related by the NOE. $^{13}C^{\alpha}$ chemical shifts relative to the random coil values, $\Delta\delta(^{13}C^{\alpha})$, are plotted at the bottom of the figure, where positive values are shifts to lower field. The sequence locations of three helices are indicated at the bottom; broken lines are used to indicate that the identification of helix 2 from these data is uncertain.

recognized from presentations of this type (Fig. 8), additional information on long-range NOEs is needed for the identification of β-sheets. A two-dimensional drawing of the β-sheets enables a concise presentation of these data, which may even include some details on the experiments used (Fig. 9).

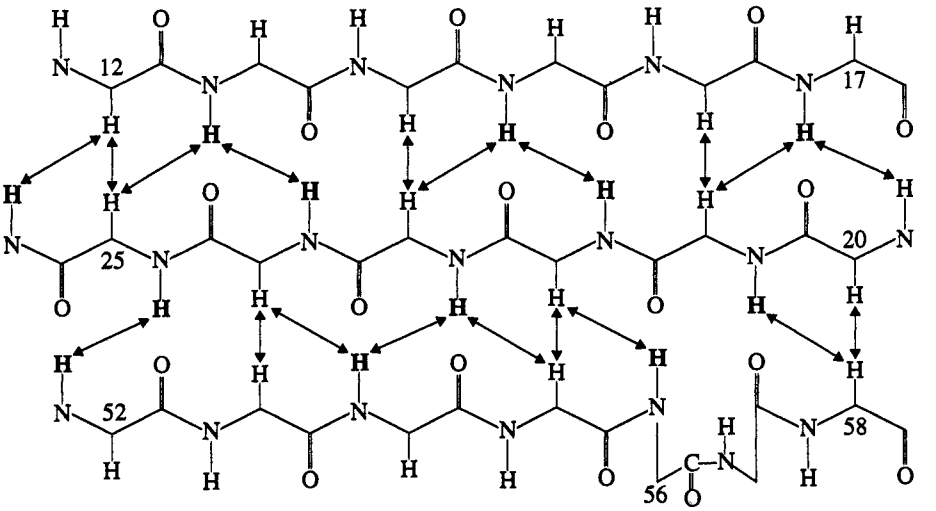

Fig. 9. Presentation of NMR evidence for the presence of β-sheet in a protein (the example shown is Tendamistat (ref. 33). Slowly exchanging amide protons are shown in bold. Interstrand, long-range ^{1}H—^{1}H NOEs $d_{\alpha\alpha}(i,j)$, $d_{\alpha N}(i,j)$, and $d_{NN}(i,j)$ are indicated by double-headed arrows. Similar representations can be drawn for parallel β-sheets.

7.2. Polynucleotides

7.2.1. Mononucleosides

The conformation of a mononucleoside within a nucleic acid structure is determined by the sugar conformation and the glycosidic torsion angle. Diagnostic characterization of these structural features often is possible from simple inspection of the NMR spectra.

- In a ribose ring, $^3J_{H1'H2'} \approx 1$ Hz is diagnostic for the *N*-puckered conformation ($P = 0°$ to 18°), and $^3J_{H1'H2'} \approx 7.9$ Hz is diagnostic for the *S*-puckered conformation ($P = 144°$ to 162°).
- In a deoxyribose ring, $^3J_{H1'H2'} \approx 1.8$ Hz is diagnostic for the *N*-puckered conformation ($P = 0°$ to 18°), and $^3J_{H1'H2'} \approx 10$ Hz is diagnostic for the *S*-puckered conformation ($P = 144°$ to 162°).
- The distance between H6 (in pyrimidines) or H8 (in purines) and the sugar ring proton H1′ varies from ~ 2.5 Å (0.25 nm) in the *syn* ($\chi \approx 60°$) orientation to ~ 3.7 Å (0.37 nm) in the *anti* ($\chi \approx 240°$) orientation. The *anti* orientation, which is found in A- and B-type helices, therefore, is characterized by cross-relaxation cross peaks of low intensity. The *syn* conformation, which is found, for example, in the G-residues in Z-DNA, is characterized by cross-relaxation cross peaks of high intensity. Pyrimidine residues are rarely found in the *syn* orientation.
- The distance between H6 (in pyrimidines) or H8 (in purines) and the sugar ring proton H2′ varies from ~ 2 Å (0.2 nm) in the *anti* orientation to ~ 4 Å (0.4 nm) in the *syn* orientation. Because these distances are nearly independent of the sugar pucker, the magnitudes of their cross-relaxation cross peaks are particularly useful for estimating the magnitude of the glycosidic angle χ.
- Until a specific format for the presentation of these diagnostic data has been agreed upon, description in the text is recommended.

7.2.2. Backbone angles

In favorable situations, heteronuclear ^1H - ^{31}P coupling constants can be used to recognize preferred backbone conformations. The preferred orientation of the backbone angle β (P—O5′—C5′—C4′) in helix-type structures is in the *trans* range, which is characterized by relatively small values for $^3J_{H5'P}$ and $^3J_{H5''P}$ (about 2 to 5 Hz). Unusual folding patterns (e.g., in loop structures) can lead to β values in the g^+ range, characterized by $^3J_{H5''P} \approx 24$ Hz, or the g^- range, characterized by $^3J_{H5'P} \approx 24$ Hz. In B-type DNA helices, the torsion angles β and γ are in the preferred β^t and γ^+ conformations, and the fragment P—O5′—C5′—C4′—H4′ forms an all-*trans*, planar, *W*-type geometry. This results in a long-range $^4J_{H4'P}$ coupling constant of 2-3 Hz, which is large enough to lead to discernible ^1H - ^{31}P hetero-COSY cross peaks. In A-type RNA helices, β and γ adopt similar preferred conformations, but the resulting geometry is not of the 'ideal', all-*trans*, planar, *W*-type, and thus no ^1H - ^{31}P cross peaks are visible for the H4′ - P combination. The backbone torsion angle ε, which has the preferred orientation ε^t, is characterized by a similar effect: when ε switches to the ε^- orientation (ε^+ is forbidden), the fragment H2′—C2′—C3′—O3′—P can form a planar, *W*-type geometry, which is characterized by the presence of $^1H2'$—^{31}P hetero-COSY cross peaks. Until a specific format for the presentation of these data has been proposed, it is recommended that they be presented either in a table or as part of the text.

7.2.3. Secondary structure

The observation of imino proton resonances between 11.5 and 14.5 ppm is indicative of the presence of base-paired helical regions in a nucleic acid. The resonance intensity in this spectral range provides a lower limit to the number of helical base pairs. The occurrence of imino proton resonances at $\delta_{DSS} > 14.5$ ppm is indicative of the presence of protonated cytosine involved in base-pair hydrogen bonding.

8. Collection of input constraints and structure calculation

The collection of constraints used in calculating the structure of a protein or nucleic acid usually is an iterative process, where an initial fold, determined on the basis of a small number of constraints supports the identification of additional constraints, which supplement the original constraints in the next round of refinement. Thus, the collection of input constraints can be considered an integral part of the structure calculation process, and both aspects are dealt with in this section.

8.1. Conformational constraints

8.1.1. Distance constraints from NOE data

It is customary to describe the methods used to assign NOE cross peaks and those employed to convert NOE cross peak intensities into interproton distance constraints. It is essential that the procedure followed in determining the constraints be described in sufficient detail that a worker could repeat their derivation given equivalent NOE data. It is important to specify whether all types of protons were treated equally, or whether special corrections were made, for example, to account for motional averaging or interactions involving methyl groups. For oligomeric proteins, it is important to describe if and how intermolecular NOEs were distinguished from intramolecular interactions.

8.1.2. Torsion angle constraints from *J*-coupling data

It is usual to specify the methods used to relate *J*-couplings to structural constraints. This includes the specific Karplus-type relationships employed, their representation by an energetic penalty function during the structure calculation (see for example, ref. 50), and possibly additional criteria used to resolve the up to four-fold degeneracy of the Karplus relation between torsion angle and *J*-coupling.

8.1.3. Torsion angle constraints from chemical shifts

If chemical-shift-derived torsion angle constraints are used in the structure refinement procedure, it is important to describe the energetic penalty function used, the method by which the predicted shift was calculated, and the refinement stages in which chemical shift information was included.

8.1.4. Constraints representing disulfide bonds, hydrogen bonding, and metal coordination

Standard upper and lower distance constraints are available to represent disulfide bonds that have been identified either by chemical methods or from NMR results (ref. 51). Similarly, standard upper and lower distance constraints are available to represent hydrogen bonds inferred from diagnostic NMR data (see **7.**) (ref. 51). Hydrogen bonding distances for the Watson-Crick and Hoogsteen base pairs are usually taken from Saenger (ref. 52) with error estimates of ± 0.2 Å (0.02 nm). Justification is needed, in all cases, for the selection of the ligating atoms and, in the case of metal coordination, for the coordination geometry used in the structure calculation. In all cases, documentation of how these constraints were incorporated in the energetic penalty function, and at what stage of the analysis, is of interest.

8.1.5. Supplementary constraints for structure refinement

Structural information may be derived from a variety of additional sources, including evidence of hydration, effects of paramagnetic shift and relaxation reagents, isotope shifts, nuclear quadrupole couplings, photochemical chemically induced dynamic nuclear polarization (photo CIDNP) experiments, field dependence of *J*-couplings, anisotropy of relaxation times, etc. It is important to describe in some detail the way in which these data were collected and the means by which they were applied to the structure calculation. If an energetic penalty function was used, its functional form should be specified. It is important to document all assumptions and additional sources of information used in calculating the structure from the experimental data.

8.1.6. Input for the final structure calculation

Data bank deposition of the input for the final structure calculation is considered to be as important as the deposition of the final atomic coordinates (**10.**). Standard formats for the deposition of these constraints (**8.1.1.** — **8.1.5.**) are available from the data banks: BMRB (ref. 2) and PDB (ref. 1).

In publications of structures of proteins and nucleic acids, it is customary to provide the following as a concise overview of the types of input constraints determined. (Prior to this count, NOEs that cannot result in a constraint due to the distance limits imposed by the covalent geometry, for example, upper limit distance constraints > 3.3 Å (0.33 nm) for vicinal protons, should be eliminated. Upper and lower limits belonging to the same NOE should be counted as one constraint.):

- The number and type of NOE (or rotating-frame Overhauser effect (ROE)) constraints, classified as, intraresidue NOEs ($i - j = 0$), sequential NOEs ($|i - j| = 1$), medium-range NOEs ($|i - j| < 5$), long-range NOEs (all other intramolecular), and intermolecular constraints (if applicable).
- The number of torsion angle or *J*-coupling constraints used, classified by the type of angle (ϕ, ψ, χ^1, etc., in polypeptides; β, γ, δ, ε, χ, the sugar pucker constraints, and, if used, the staggered conformations of the α and ζ angles in nucleic acids).
- The number of hydrogen bond constraints (with a distinction made between the number of constrained hydrogen bonds and the number of constraints used to describe these bonds).
- If applicable, other types of constraints.

8.2. Methods used for structure calculation and refinement

The computational methods used for structure determination and refinement (including names and versions of software packages used) need to be specified in adequate detail, and it is important to provide references to parameter sets, potential functions and values of force constants used, or to include these values explicitly in the report if they have not been published elsewhere.

8.2.1. Structure calculation

It is desirable that the process for deriving the three-dimensional structure from the NMR data be described clearly, including the starting conformation(s), numbers and types of constraints used, and the criteria used to reject unfavorable folds. It is important to describe the protocol used in minimizing the energetic penalty function during the various stages of refinement (for example, the initial and final temperatures together with the number of steps, the step size, and the force constants used in simulated annealing protocols).

8.2.2. Preliminary structures to assign additional constraints

Currently, most structure determination protocols remove ambiguities from NOE cross-peak assignments by reference to an initial fold calculated with an incomplete set of (ideally unambiguous) constraints. It is important to document any ambiguity in the initial set of constraints along with the methods used for assigning additional constraints, for example, by reference to the software used and by reporting adequate details on the procedures (number of iterations, criteria and cut-off values) employed for the rejection of NOEs at various stages in the refinement.

8.2.3. Structure refinement

Structure refinement procedures which rely on computationally intensive programs for relaxation matrix refinement or chemical shift calculations are frequently only incorporated during the final stages of the structure determination. It is important to provide sufficient details regarding the use of these procedures, including their effects on the average structure used as input for the refinement, and their effect on the root-mean-square deviation of the ensemble of structures.

9. Reporting three-dimensional structures

9.1. Presentation of structures

In reporting three-dimensional NMR structures it is common practice to display the entire ensemble of conformers used for statistical analysis. In addition, a representative conformer, suitable for detailed structural interpretation, often is presented. The representative conformer has been identified variously as the structure that best satisfies the NMR constraints, as the energy-minimized structure derived from the averaged coordinates of the ensemble, or as the structure that is closest to the averaged coordinates of the ensemble (ref. 53). Subtleties in the averaging of NMR properties (ref. 54), which differ from averaging in X-ray crystallography, need to be considered. It is important to specify the total number of conformers calculated and the criteria for selection of the subset of conformers, including the representative conformer used for display and analysis.

9.2. Agreement of structures with constraints

- It is important to report complete statistics as to how well the structures satisfy the constraints. The maximum violations of distance bounds, torsion angle constraints, coupling constants, chemical shifts, or other experimental constraints used in the structure calculation, are normally reported, together with the average violation per constraint (± standard deviation).
- When computational methods are used that provide a measure of the agreement of the structures with the constraints or with direct spectroscopic data such as NOE intensities, the final value of the target function or figure of merit (R-factor) normally is reported. In the particular case of relaxation matrix calculation of NOE intensities, it is appropriate to specify the functional form of the R-factor (*e.g.*, direct, r^6 weighted, etc.).
- It is usual to specify the deviations from idealized covalent geometry for bond lengths, bond angles, and impropers and to indicate which idealized geometry was used. Since covalent bond lengths and angles are restrained to "ideal" values in all methods currently used to compute NMR structures, deviations from ideality do not provide an independent assessment of the overall quality of the structures but can identify problem regions where constraints might be too tight or where constraints may be in conflict as a result of conformational averaging.

9.3. Precision of structures

An ensemble of conformers calculated from the same set of input data is widely used to represent an NMR-derived structure. The precision of the structure determination commonly is expressed in terms of a statistical analysis of the variation of the atomic coordinates and torsion angles among these conformers. The precision may then be reported either as average pairwise root mean square (rms.) deviation or the rms. deviation relative to the mean coordinates. A meaningful description of the precision of local structure is provided by the circular variance (ref. 55) or the angular order parameter (ref. 56) for torsion angles. It is useful to specify the method of superposition used in obtaining the mean coordinates. For proteins, the average rms. deviation for Cartesian coordinates usually is reported for a set of backbone heavy atoms, *i.e.*, sets of (C^{α}), (N, C^{α}, C), or (N, C^{α}, C, O) atoms, all side-chain heavy atoms, and all heavy atoms (backbone plus side chain). It is important to indicate whether the reported rms. deviations apply to all amino acid residues or only to a selected subset. For nucleic acids, the atoms used for calculations of rms. deviations should be specified.

Useful supplementary information includes evaluation of the rms deviations along the sequence in light of the density of constraints per residue, in order to better evaluate the significance of the apparent precision.

9.4. Validation of structures

In common with other methods for three-dimensional structure analysis, for example X-ray diffraction in single crystals, there is no direct method for absolute validation of the result of an NMR structure determination. Nonetheless, a number of criteria are available for investigating whether the result of an NMR

structure determination is "reasonable." Such approaches have been developed only recently and are being used to evaluate the results of both NMR and X-ray structure determinations. Additional techniques are currently in development in different laboratories. These commonly are based on the database of currently-available three-dimensional structures (ref. 55—61). Criteria used for structure validation include the following:

- Conformational energy (either Lennard-Jones or total energy) with associated force field used.
- For proteins, a Ramachandran (ϕ, ψ) plot for the backbone torsion angles in the family of conformers.
- For more detailed assessment of the stereochemical quality of a protein structure, several computer programs are available (reviewed in reference 60). These may identify regions of the structure in which potential problems require further evaluation. Most of these programs have been written primarily for checking X-ray coordinates, rather than an ensemble of conformers obtained by NMR, but a recently published suite of programs (ref. 61) has been designed for the validation of NMR structures.

10. Data bank deposition of NMR structures and supporting data

Deposition in public data banks of the quantitative and semi-quantitative data and references to specific procedures used to manipulate the data, as described in these recommendations, is strongly encouraged, even in cases where this duplicates information tabulated in journal articles. The key elements in a data deposition for an NMR structure include:

- Representation of the covalent structure(s) of the molecule(s) in the system reported (generally a sequence plus indications of cross links and special modifications) plus information on bound cofactors or other ligands and oligomeric structure.
- Tabulation of assigned chemical shifts.
- Tabulation of assigned coupling constants.
- Tabulation of constraints used in the structure calculation: distance constraints from NOEs (or the spectroscopic data from which the constraints were derived), torsion angle constraints from *J*-couplings and chemical shifts, H-bond constraints, disulfide bridge constraints, and supplementary constraints: it is important to indicate how the constraints were calibrated and the nature of any pseudoatom or other corrections (e.g., to account for spin diffusion or spin multiplicity) included in the constraint values and to specify what type of averaging protocol (*e.g.*, center averaging, r^{-6} averaging, etc.) has been used (ref. 54, 62).
- Cartesian coordinates for the family of conformers that represent the result of the structure determination as well as for a single representative conformer: it is important to include all available atoms in the deposition, including the protons.
- Concise description of the solution conditions used for the structure determination (temperature, pH, ionic composition, etc.).
- Literature citations for the studies that originated the deposited data.

Compilers of databases are encouraged to accommodate the deposition of supplementary information of the kind described in the preceding sections either as free text or in a more organized format. This may include a brief description of the computational techniques (including specification of the software) used to process the input data, derive and refine the structures, and validate the results.

Additional information usefully available from a database includes tables of NMR relaxation rates, hydrogen exchange rates or protection factors, and thermodynamic parameters characterizing conformational equilibria and ligand binding. Some authors may wish to deposit electronic files containing the free induction decays from a representative cross-relaxation experiment and/or lists of peaks and intensities of the primary data sets.

In analogy with the reporting of structures determined by X-ray crystallography, it is desirable that journal editors require, as a condition for acceptance of a publication, database deposition of at least (1) the atomic

coordinates, (2) the assigned chemical shifts, (3) the assigned *J*-couplings, and (4) a compilation of all input constraints used for the structure determination.

Conclusion

The recommendations presented here are designed to support easier communication of NMR data and NMR structures for proteins and nucleic acids through unified nomenclature and reporting standards. Much of this document pertains to the reporting of data in journal articles. However, in the best interest of the future development of structural biology, it is desirable that the bulk of the reported information be stored in computer-accessible form and be freely accessible to the scientific community. For such purpose, the macromolecular crystallographic community is advocating use of the mmCIF format (ref. 63, 64), which is compatible with the STAR/CIF format (ref. 65, 66); mmCIF has been implemented in the Nucleic Acid Database. In recognition of the desirability of developing full compatibility between machine-readable crystal diffraction and NMR data, it is recommended that a compatible format be used for NMR data from proteins and nucleic acids. It may prove desirable, in addition, for the databases to develop full compatibility with the ASN.1 data exchange format (ref. 67—69), which has been adopted by the U.S. National Center for Biotechnology Information and is being used in the chemical exchange format developed by the Chemical Abstracts Service (ref. 70). With these goals in mind, international committees, in association with the Protein Data Bank and BioMagResBank, are in the process of using the present document to develop data dictionaries and author-input protocols for the deposition of macromolecular NMR data (ref. 71).

References and Notes

1. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, & T., Tasumi, M. (1977) The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures, *J. Mol. Biol.* **112**, 535—542. The Protein Data Bank can be accessed on the Internet at http://www.pdb.bnl.gov/.
2. Seavey, B. R., Farr, E. A., Westler, W. M., & Markley, J. L. (1991) A Relational Database for Sequence-Specific Protein NMR Data, *J. Biomol. NMR* **1**, 217—230; Markley, J. L., Farr, E. A., & Seavey, B. R. (1991) Appendix: Check List for Protein NMR Data to be Submitted for Publication or for Database Inclusion, *J. Biomol. NMR* **1**, 231—236. The Biomolecular NMR Data Bank (BioMagResBank or BMRB) can be reached on the Internet at http://www.bmrb.wisc.edu/.
3. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R., & Schneider, B. (1992) The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids, *Biophys. J.* **63**, 751—759. The Nucleic Acid Data Bank can be reached on the Internet at http//www.ndbserver.rutgers.edu:80/.
4. Hendrickson, W. A., & Wüthrich, K. (1991—) *Macromolecular Structures*, Current Biology, London, UK.
5. Nomenclature Committee of IUBMB (NC-IUBMB) and IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) (1992) Newsletter 1992, *Biochem. Int.* **26**, 567—575; and references therein.
6. Commission on Molecular Structure and Spectroscopy (1972) Recommendations for Presentation of NMR Data for Publication in Chemical Journals, *Pure Appl. Chem.* **29**, 627—628.
7. Commission on Molecular Structure and Spectroscopy (1976) Presentation of NMR Data for Publication in Chemical Journals--B. Conventions Relating to Spectra from Nuclei other than Protons (Recommendations 1975), *Pure Appl. Chem.* **45**, 219.
8. Harris, R. K., Kowalewski, J., and Cabral de Menezes, S. (1997) Parameters and Symbols for use in Nuclear Magnetic Resonance, *Pure Appl. Chem.* 1997, **69**, 2489–2495.
9. Liébecq, C. (ed.) (1992) Biochemical Nomenclature and Related Documents - A Compendium, 2nd ed., Portland Press on behalf of IUBMB. London.
10. Bonnett, R. (1978) in The Porphyrins (D. Dolphin, Ed.), Vol. 1, Part A., Academic Press, New York, pp. 1—27.
11. IUPAC-IUB Commission on Biochemical Nomenclature (CBN) (1975) Nomenclature of Iron-Sulfur Proteins. Recommendations 1973, *Eur. J. Biochem.* **35**, 1—2.
12. IUPAC-IUB Commission on Biochemical Nomenclature (1975) The Nomenclature of Peptide Hormones. Recommendations 1974, *Biochemistry* **14**, 2559—2560; reprinted in the Compendium (ref. 9).
13. IUPAC-IUB Commission on Biochemical Nomenclature (1972) Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino Acids). Revised Recommendations 1971, *Biochemistry* **11**, 942—944; reprinted in the Compendium (ref. 9).

14. Nomenclature Committee of the International Union of Biochemistry (NC-IUB) Nomenclature for Incompletely Specified Bases in Nucleic Acids (1985), *Eur. J. Biochem.* **150**, 1—5; reprinted in the Compendium (ref. 9).
15. Access to various databases (*e.g.*, GenBank, DDBJ, EBI (EMBL), EC Enzyme Data Bank, PIR, SWISS-PROT, PRF, PDB, NDB, BMRB) is provided over the Internet. Useful compilations of Internet addresses are found, for example, on the home pages of the Protein Data Bank (http://www.pdb.bnl.gov/), BioMagResBank (http://www.bmrb.wisc.edu), the NCSA Biology Workbench (http://biology.ncsa.uiuc.edu), and Pedro's Biomolecular Research Tools (http://www.public.iastate.edu/~pedro/research_tools.html).
16. Protein Identification Resource, Georgetown University, Washington, D.C., USA. The Protein Identification Resource can be searched with tools located at: http://www.gdb.org/Dan/proteins/pir.html.
17. SWISS-PROT Annotated Protein Sequence Database (http://www.expasy.hcuge.ch/sprot/sprot-top.html).
18. International Union of Biochemistry and Molecular Biology (1992) Enzyme Nomenclature. Recommendations 1992, Academic Press, Orlando, Florida.
19. Chemical Abstracts Service, Columbus, OH, USA (http://info.cas.org/).
20. IUPAC-IUB Commission on Biochemical Nomenclature (CBN) (1970) Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. Tentative Rules 1969, *Biochemistry* **9**, 3471—3479; reprinted in the Compendium (ref. 9).
21. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) (1984) Nomenclature and Symbolism for Amino Acids and Peptides. Recommendations 1983, *Eur. J. Biochem.* **138**, 9—37; reprinted in the Compendium (ref. 9).
22. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) (1983) Abbreviations and Symbols for the Description of Conformations of Polynucleotide Chains. Recommendations 1982, *Eur. J. Biochem.* **131**, 9—15; reprinted in the Compendium (ref. 9).
23. Hanson, K. R. (1966) Applications of the Sequence Rule. I. Naming the Paired Ligands *g,g* at a Tetrahedral Atom X_{ggij}. II. Naming the Two Faces of a Trigonal Atom Y_{ghi}, *J. Am. Chem. Soc.* **88**, 2731—2742.
24. The 1983 recommendations (ref. 21) specified the use of numbers to designate atoms and the use of the R/S rules of Cahn, Ingold, & Prelog (ref. 25 and 26, see also ref. 27—29) to define stereochemistry. According to the *R/S* rules, a tetrahedral carbon *X* with prochiral substituents $A > B > C = C'$ (where the priority is according to decreasing atomic mass and *C* and *C'* have equivalent mass and are labeled arbitrarily) is labeled as follows: one sights down the *A*—*X* axis and assumes that *C'* takes precedence over *C* (as it would if it were replaced, for example, with a heavier isotope). Then, if *B*, *C'*, and *C* (projecting away from the viewer) are clockwise, *C'* is labeled as *pro-R* and *C* is labeled as *pro-S*; if they are counterclockwise, *C'* is labeled as *pro-S* and *C* is labeled as *pro-R*.
25. Cahn, R. S., Ingold, C. K., & Prelog, V. (1966) Specification of Molecular Chirality, *Angew. Chem. Int. Ed. Engl.* **5**, 385—415; 511.
26. Prelog, V., & Helmchen, G. (1982) Basic Principles of the CIP-System and Proposals for a Revision, *Angew. Chem. Int. Ed. Engl.* **21**, 567—583.
27. Blackwood, J. E., Gladys, C. L., Loening, K. L., Petrarca, A. E., & Rush, J. E. (1968) Unambiguous Specification of Stereoisomerism about a Double Bond, *J. Am. Chem. Soc.* **90**, 509—510.
28. Blackwood, J. E., Gladys, C. L., Petrarca, A. E., Powell, W. H., & Rush, J. E. (1968) Unique and Unambiguous Specification of Stereoisomerism about a Double Bond in Nomenclature and Other Notation Systems, *J. Chem. Doc.* **8**, 30-32.
29. IUPAC Commission on Nomenclature of Organic Chemistry (CNOC) (1976) Nomenclature of Organic Chemistry, Section E: Stereochemistry. Recommendations 1974, *Pure Appl. Chem.* **45**, 11—30 (1976); reprinted in the Compendium (ref. 9).
30. Wüthrich, K., Billeter, M., & Braun, W. (1983) Pseudo-Structures for the 20 Common Amino Acids for Use in Studies of Protein Conformations by Measurements of Intramolecular Proton—Proton Distance Constraints with Nuclear Magnetic Resonance, *J. Mol. Biol.* **169**, 949—961.
31. Since the experimental NOEs involve distances to the individual protons rather than to the reference point of the pseudoatom, NOE distance constraints involving pseudoatoms must contain a correction factor. A conservative and straightforward approach is to set these corrections equal to the distance between the reference point of the pseudoatom and the protons that it replaces. Tables listing these corrections for the common amino acid residues have been published (ref. 30, 33) as well as more refined corrections (for example, see ref. 32, 34).
32. Koning, T. M. G., Boelens, R., & Kaptein, R. (1990) Calculation of the Nuclear Overhauser Effect and Determination of Proton—Proton Distances in the Presence of Internal Motions, *J. Magn. Reson.* **90**, 111-123.
33. Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley Interscience, New York.
34. Güntert, P., Braun, W., & Wüthrich, K. (1991) Efficient Computation of Three-Dimensional Protein Structures in Solution from Nuclear Magnetic Resonance Data Using the Program DIANA and the Supporting Programs CALIBA, HABAS and GLOMSA, *J. Mol. Biol.* **217**, 517—530.
35. Historical changes in the definition of these angles have contributed to confusion. Their use in numerous textbooks does not conform to the current IUPAC-IUBMB nomenclature: Clauwaert, J., & Xia, J.-Z. (1993) Straightening out the Dihedral Angles, *Trends Biochem. Sci.* **18**, 317—318; 464.

36. Wilmott, C. M., & Thornton, J. M. (1990) β-turns and their Distortions: A Proposed New Nomenclature, *Protein Eng.* **3**, 479—493.
37. Richardson, J. S. (1981) The Anatomy and Taxonomy of Protein Structure, *Adv. Protein Chem.* **34**, 167—339.
38. Milner-White, E. J., Bell, L. H., & Maccallum, P. H. (1992) Pyrrolidine Ring Puckering in *cis*- and *trans*-Proline Residues in Proteins and Polypeptides. Different Puckers are Favoured in Certain Situations, *J. Mol. Biol.* **228**, 725—734.
39. de Leeuw, H. P. M., Haasnoot, C. A. G., & Altona, C. (1980) Empirical Correlations between Conformational Parameters in β-D-Furanoside Fragments Derived from a Statistical Survey of Crystal Structures of Nucleic acid Constituents, *Isr. J. Chem.* **20**, 108—126.
40. Haasnoot, C. A. G., de Leeuw, F. A. A. M., de Leeuw, H. P. M., & Altona, C. (1981) The Relationship between Proton-Proton NMR Coupling Constants and Substituent Electronegativities *Org. Magn. Reson.* **15**, 43—52.
41. Harris, R. K., Becker, E. D., Bremser, W., Cabral de Menezes, S., Goodfellow, R., & Granger, P. IUPAC (Commission on Molecular Structure and Spectroscopy, I.5) Recommendations for NMR Nomenclature. A. Nuclear Spin Properties and Conventions for Chemical Shifts, (in preparation).
42. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., & Sykes, B. D. (1995) ^{1}H, ^{13}C and ^{15}N Chemical Shift Referencing in Biomolecular NMR, *J. Biomol. NMR.* **6**, 135—140.
43. In a dilute aqueous solution at 298 K (0.3 mM DSS and 0.03 mM TMS in D_2O at pH* 7), the signal of TMS is at δ_{DSS} = -0.0173 ppm (i.e., is at low frequency, historically referred to as "upfield", from that of DSS); in a dilute organic solution at 298 K (0.03 mM TMS and 0.01 mM DSS in perdeuterated dimethyl sulfoxide), the signal of TMS is at δ_{DSS} = 0.0246 ppm (i.e., to high frequency from that of DSS) from that of DSS.
44. Live, D. H., Davis, D. G., Agosta, W. C., & Cowburn, D. (1984) Long Range Hydrogen Bond Mediated Effects in Peptides: ^{15}N NMR Study of Gramicidin S in Water and Organic Solvents, *J. Am. Chem. Soc.* **106**, 1939—1943.
45. The Ξ value recommended here for ^{2}H was determined from the ratios of ^{1}H and ^{2}H signals of a water sample at 298 K as measured by Dr. Frits Abildgaard and David M. LeMaster at the National Magnetic Resonance Facility at Madison. This value is a surrogate for the isotopomers of DSS, and its derivation assumes that there is no primary isotope effect on the ratio.
46. The Ξ value recommended here for ^{31}P is the average of values determined independently, one at the National Magnetic Resonance Facility at Madison by F. Abildgaard (40.4808651), and the other at the National Institutes of Health by A. Bax (40.480862). Measurements were made as 298 K in D_2O solution containing 10% trimethylphosphate as the indirect reference and DSS as the direct reference.
47. NOE distances may be represented rigorously by notation analogous to that used for *J*-couplings. However, since protons are always implied, a shorthand notation is widely used: for example, $d_{\alpha N}$ for $d_{H\alpha HN}$, d_{NN} for d_{HNHN}, $d_{\beta N}$ for $d_{H\beta HN}$, and $d_{\alpha\beta}$ for $d_{H\alpha H\beta}$.
48. Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) Solution Conformation of Proteinase Inhibitor IIA from Bull Seminal Plasma by ^{1}H NMR Spectroscopy and Distance Geometry, *J. Mol. Biol.* **182**, 295—315.
49. Gronenborn, A. M., & Clore, G. M. (1994) Identification of N-Terminal Helix Capping Boxes by Means of ^{13}C Chemical Shifts, *J. Biomol. NMR* **4**, 455—458.
50. Eberstadt, M., Gemmecker, G., Mierke, D. F., Kessler, H. (1995) Scalar Coupling Constants—Their Analysis and their Application for the Elucidation of Structures, *Angew. Chem. Int. Ed. Engl.* **34**, 1671—1695.
51. Ottiger, M., Szyperski, T., Luginbühl, L., Ortenzi, C., Luporini, P., Bradshaw, R. A., & Wüthrich, K. (1994) The NMR Solution Structure of the Pheromone Er-2 from the Ciliated Protozoan *Euplotes raikovi*, *Protein Science* **3**, 1515—1526.
52. Saenger, W., (1983) Principles of Nucleic Acid Structure, Springer Verlag, New York, pp. 123—124.
53. Sutcliffe, M. J. (1993) Representing an Ensemble of NMR-Derived Protein Structures by a Single Structure, *Protein Sci.* **2**, 936—944.
54. Fletcher, C. M., Jones, D. N. M., Diamond, R., Neuhaus, D. (1996) Treatment of NOE Constraints Involving Equivalent or Nonstereoassigned Protons in Calculations of Biomacromolecular Structures, *J. Biomol. NMR* **8**, 292—310.
55. MacArthur, M. W., & Thornton, J. M. (1993) Conformational Analysis of Protein Structures Derived from NMR Data, *Proteins* **17**, 232—251.
56. Hyberts, S. G., Goldberg, M. S., Havel, T. F., Wagner, G. (1992) The Solution Structure of Eglin c Based on Measurements of Many NOEs and Coupling Constants and its Comparison with X-ray Structures, *Protein Sci.* **1**, 736—751.
57. Colovos, C., & Yeates, T. O. (1993) Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions, *Protein Sci.* **2**, 1511—1519.
58. Sippl, M. J. (1993) Recognition of Errors in 3-Dimensional Structures of Proteins, *Proteins* **17**, 355—362.
59. Vriend, G., & Sander, C. (1993) Quality Control of Protein Models: Directional Atomic Contact Analysis, *J. Appl. Cryst.* **26**, 47—60.

60. MacArthur, M. W., Laskowski, R. A, Thornton, J. M. (1994) Knowledge-Based Validation of Protein-Structure Coordinates Derived by X-ray Crystallography and NMR Spectroscopy, *Current Opinion Struct. Biol.* **4**, 731—737.
61. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996) AQUA and PROCHECK—NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR* **8**, 477—486.
62. Jardetzky, O., Roberts, G. C. K. (1981) NMR in Molecular Biology, Academic Press, New York, pp. 126—142.
63. Watenpaugh, K. D., Berman, H. M., Bourne, P. E., & Fitzgerald, P. M. D. (1993) The Macromolecular CIF Dictionary, *Acta Crystallogr. Ser. A* **49** (Supplement), IVI International Congress (IUCr), Abstract OCM—18.02.07.
64. Fitzgerald, P. M. D., Berman, H., Bourne, P., & Watenpaugh, K (1993) The Macromolecular CIF Dictionary, Abstract No. D008, American Crystallographic Association Annual Meeting, Albuquerque, NM, May 23—28 (available from: http://ndbserver.rutgers.edu:80/mmcif/).
65. Hall, S. R., & Spadaccini, N. (1994) The STAR File: Detailed Specifications, *J. Chem. Inf. Comput. Sci.* **34**, 505—508.
66. Hall, S. R., Allen, F. H., & Brown, I. D., [International Union of Crystallography, Commission on Crystallographic Data, Commission on Journals, Working Party on Crystallographic Information] (1991) The Crystallographic Information File (CIF): a New Standard Archive File for Crystallography, *Acta Crystallogr. Ser. A* **47**, 655—685.
67. International Organization for Standardization (1987) Information Processing Systems—Open Systems Interconnection—Specification of Abstract Syntax Notation One (ASN.1), Technical Report ISO8824.
68. International Organization for Standardization (1987) Information Processing Systems—Open Systems Interconnection—Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN.1), Technical Report ISO8825.
69. Ostell, J. (1990) GenInfo ASN.1 Syntax: Sequences, Technical Report 1 & 2, National Center for Biotechnology Information, National Library of Medicine.
70. Mockus, J., & Steckert, T. D. (1994) CXF Chemical eXchange Format Version 1.0, Chemical Abstracts Service: Columbus, OH, USA.
71. Ulrich, E. L., & Markley, J. L. (1994) A Data Dictionary and Information File for the Exchange and Archiving of NMR Experimental Results, 35th Experimental Nuclear Magnetic Resonance Conference, April 10—15, Asilomar, California, Abs. WP 112, p. 264. The BioMagResBank web pages (http://www.bmrb.wisc.edu/) contain complete information about this effort including user entry protocols and software.

Corrigendum:

IUPAC Recommendations on Nomenclature and Symbols and Technical Reports from Commissions

IUPAC–IUBMB–IUPAB Inter-union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy (J.L. Markley, A. Bax, Y. Arata, C.W. Hilbers, R. Kaptein, B.D. Sykes, P.E. Wright and K. Wüthrich). Recommendations for the presentation of NMR structures of proteins and nucleic acids (IUPAC Recommendations 1998). *Pure & Applied Chemistry* 1998, **70**, 117–142.

The authors have noted some further corrections which they wish to make to the printed version. These are as follows:

p. 122, final line: '(ref. 25, 26)' should be '(ref. 23, 24)'.
p. 125, footnote a to TABLE 1: 'Ref. (28)' should be 'Ref. (30)'.
p. 125, sixth line from the bottom: '(historically referred to as "downfield")' should be '(historically referred to as "upfield")'; '(historically "upfield")' should be '(historically "downfield")'.
p. 133, caption to Fig. 8: '(ref. 48)' should be '(ref. 51)'.
p. 135, third line: 'constraints supports' should be 'constraints, supports'