NMR Determination of Oligonucleotide Structure

NMR, in conjunction with appropriate computational searching algorithms, has become the method of choice for determining solution structures. Methodology is now available to determine an accurate, high-precision (i.e., high-resolution) structure of nearly any DNA or RNA double helix of up to 15 to 20 base pairs (bp) via NMR if sufficient care and effort are expended (Schmitz and James, 1995). Care is imperative, as it is also possible to obtain worthless structures; easy-to-use restrained molecular dynamics (rMD) programs currently available may yield a structure even if the user has provided them with poor experimental data or has used an inappropriate conformational searching protocol. It is also feasible to determine the low-resolution structure of even larger oligonucleotides (possibly up to the size of tRNA). Although one can model a chemically reasonable, even useful, structure from small amounts of data, this does not constitute structure determination.

Before resorting to the relatively time-consuming task of "determining a high-resolution structure," however, one should consider the knowledge to be gained as a result. In the case of proteins or single-stranded nucleic acids that fold into a tertiary structure, some useful insights into function can be derived from low-resolution structures or even models. The fairly subtle structural variations in a DNA or RNA·DNA helix that are sequence dependent, and consequently guide protein, mutagen, or drug recognition, demand a detailed high-resolution structure to be very useful. The discussion that follows will emphasize the steps necessary for determination of a high-resolution structure, with some mention of the variations for determining structures of nucleic acids possessing significant tertiary structure. The discussion will also be very strongly biased in describing approaches that have been successful in our lab.

One book has appeared that focuses solely on NMR of nucleic acids (James, 1995). The reader can find there details about many topics discussed below, much additional information about RNA and DNA structure determination, and references.

OVERVIEW OF OLIGONUCLEOTIDE STRUCTURE DETERMINATION BY NMR

Before embarking upon nucleic acid structure determination via NMR, it is advisable to assess the likelihood of success. The size limitation of the method has already been mentioned. Another question to consider is whether the nucleic acid possesses features suggesting that it may have a structure. As a first consideration, solubility in water implies that success is likely. A single-stranded DNA or RNA that does not have a sequence capable of forming at least four or five Watson-Crick base pairs in a helix is highly unlikely to possess a sufficiently stable structure in aqueous solution—although the i-motif for oligo-C sequences at low pH is an obvious exception to this rule (Leroy et al., 1993). Nucleic acids, fortunately, are highly soluble in water. They may, however, aggregate via nonspecific interactions, especially in the presence of multivalent counterions or ligands. Aggregation will result in broadening of NMR resonances, with consequent loss of signal intensity and ability to resolve individual signals.

A flow chart describing the process by which oligonucleotide structure is determined is presented in Figure 7.2.1. As with proteins, assignment of resonances is often the bottleneck in structure determination for structures exhibiting tertiary structure, such as many RNA elements. Making proton resonance assignments in a nucleic acid duplex, however, is fairly quick and straightforward (van de Ven and Hilbers, 1988). The pattern of cross-peaks in a proton two-dimensional nuclear Overhauser effect (2D NOE) spectrum makes distinction between left- and right-handed helices obvious. In the case of a right-handed helix, through-space connectivities between protons in sequential residues are evident from the appearance of cross-peaks corresponding to the pertinent protons in the 2D NOE spectrum. Figure 7.2.2 demonstrates the major sequential assignment strategy. With data in hand, it is often possible to "walk through" up to 200 proton assignments in a day or so. If the sequence is bound by a ligand or contains loops, bulges, or other deviations from a simple right-handed helix, however, assignment of resonances may require many months of work and
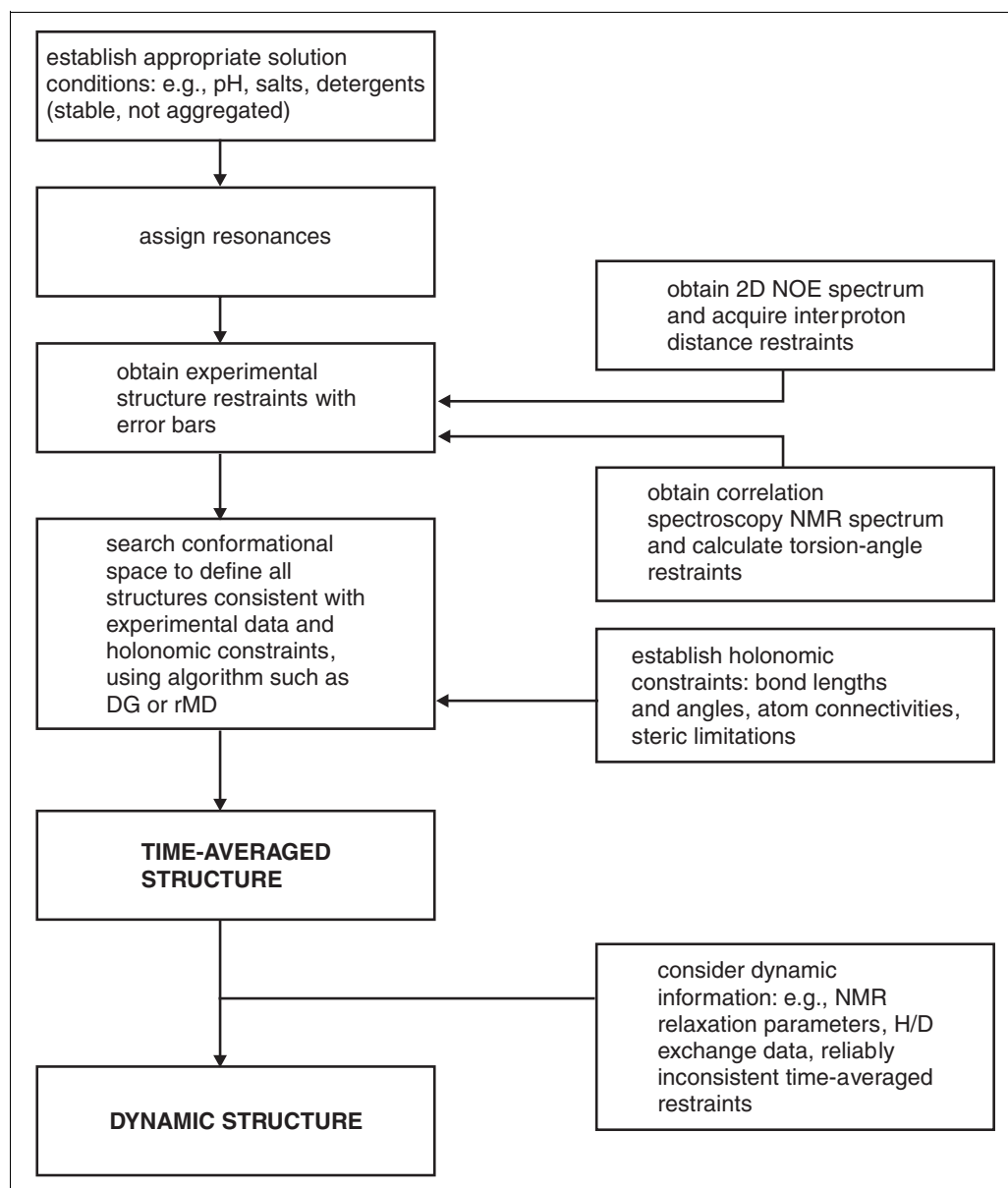
Biophysical Analysis of Nucleic Acids

Figure 7.2.1 Scheme for deriving oligonucleotide structure using experimental NMR data in conjunction with conformational searching procedures such as distance geometry (DG) and restrained molecular dynamics (rMD; sometimes called simulated annealing). 2D NOE, two-dimensional nuclear Overhauser effect.

many different types of 2D NMR experiments, including at least total correlation spectroscopy (TOCSY). If the structure is sufficiently large or complicated, it may be necessary to use ^{13}C or ^{15}N labeling of the sample and multidimensional (e.g., 3D) heteronuclear NMR.

The structure of any molecule can be determined given a sufficient number of experimental structural restraints—e.g., internuclear distances and bond torsion angles, together with the holonomic constraints of bond lengths, bond angles, and steric limitations. A key to the determination of good-quality structures is to use as many structural restraints as possible. For high-resolution structures, the restraints should also possess the best accuracy possible. A means to determine the bounds or error bars on these structural restraints is also quite valuable, as at least an estimate will be needed for structure refinement.

These experimental structural restraints are used with algorithms, such as distance geometry (DG) and rMD, that search conformational space to define all structures consistent with the experimental restraints. If these resulting structures are closely related and there is confidence
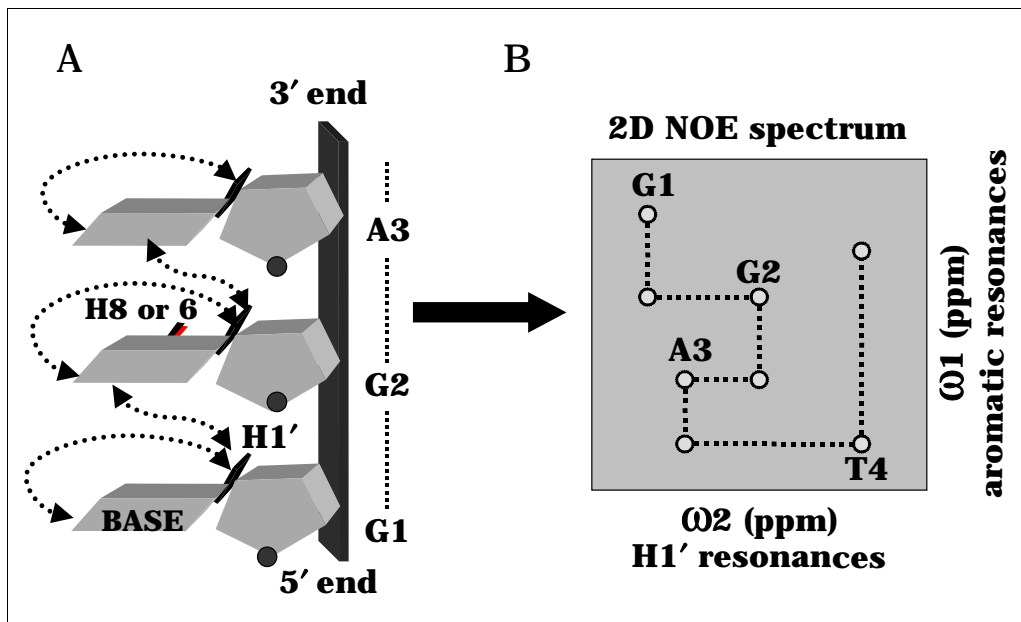
Figure 7.2.2 Schematic representations of (**A**) the sequential connectivity between protons in the sugar ring (shown as H1′) and protons in the base (H6 for pyrimidines and H8 for purines) and (**B**) the "walk" between the congruous cross-peaks in the proton homonuclear two-dimensional nuclear Overhauser effect (2D NOE) spectrum. Note, for example, that a particular H1′ proton has a 2D NOE cross-peak at a frequency ω_1 corresponding to the resonance frequency of an aromatic proton (H6 or H8) from its own residue as well as at the resonance frequency of the H6 or H8 from its 3′-neighboring residue. Likewise, each aromatic proton "sees" its own and its neighbor residue's H1′. While there is some overlap of regions, the different proton types, e.g., H6, H1′, H2′, etc., have their own characteristic range of resonance frequencies. In a real spectrum, there are many additional cross-peaks; i.e., a cross-peak should be found between a given proton and any other proton within ~5 Å. In addition to the H1′-aromatic proton walk, there is an analogous H2′- (and, for DNA, H2″) -to-aromatic proton walk. Of course, intrasugar ring protons yield cross-peaks. There is also a walk between sequential imino protons. The ability to assign these hundreds of resonances devolves from the entire pattern of interrelated connectivities; all cross-peaks must be accounted for. The only serious difficulty in making resonance assignments is that as the number of residues increases and as the structure becomes more monotonous, there is more overlap of cross-peaks obscuring discernment of connectivities. The overlap problem has driven the use of isotopic labeling and use of 3D or 4D NMR in order to spread out the overlapping cross-peaks.

that pertinent parts of conformational space have not been neglected, one can conclude that the structure has been determined. It is preferable that the array of structures generated by the search algorithms be large enough to map out the conformational space that will accommodate all available experimental data.

With the generation of a number of structures satisfying the experimental data, an assessment of the structures is in order. Typically, one compares the atomic coordinates of all of the satisfactory structures to obtain an assessment of the apparent precision or "resolution" of structure determination. In addition to visually inspecting structural features, it is often useful to calculate torsion angles and helical

parameters to compare with canonical or other nucleic acid structures.

BASIC NMR SPECTRAL PARAMETERS

Before considering in detail the determination of oligonucleotide structures from NMR data, it is worthwhile to step back and briefly consider the nature of the NMR data that will be examined. A rigorous, mathematically oriented coverage of the principles can be found in texts by Abragam (1961) and Ernst et al. (1987), while a good introductory text is that by Derome (1987). With an eye towards protein structure determination, a good description of background and applications is given by Cavanagh et al. (1996).

Biophysical Analysis of Nucleic Acids

7.2.3

The nuclear magnetic resonance phenomenon can be described succinctly as follows. If a sample is placed in a magnetic field and is subjected to radiofrequency (RF) energy at the appropriate frequency, nuclei in the sample can absorb the energy. The frequency of the radiation necessary for absorption of energy—i.e., the resonance frequency—depends on the type of nucleus (e.g., ^1H and ^{31}P have a different value for the gyromagnetic ratio γ) and also on the chemical environment of the nucleus. For example, the methyl, methylene, and hydroxyl protons of ethanol absorb at three different frequencies, and H2 protons of two different adenine residues in a DNA duplex absorb at different frequencies, since they are in different chemical environments by virtue of sequence-dependent structural differences. After absorption of energy by the nuclei, the length of time and the manner in which the nuclei dissipate that energy can also be used to reveal information regarding a variety of dynamic processes.

The magnetic field $\mathbf{B}_{local}$ perceived by the nuclei in a molecule will be very slightly altered from the field $\mathbf{B}_o$ in which the sample is placed, as a result of currents in the molecule's electrons. The frequency ν at which a nucleus absorbs RF energy depends linearly on the local magnetic field:

$$\nu = \gamma \mathbf{B}_{local}/2\pi$$

Equation 7.2.1

Equation 7.2.1 is often referred to as the Larmor equation. As the precise chemical environment will determine the local magnetic field, different nuclei of the same type will absorb energy at slightly different frequencies. Because the local magnetic field variations are small compared with that of the magnet in which the sample is placed, the frequency shifts will be relatively small. So if the absorption frequency ν is several hundred megahertz, differences in resonance frequencies for two different hydrogen nuclei will be on the order of several hertz. Although one cannot easily determine absolute RFs to an accuracy of ± 1 Hz, it is possible to determine the relative positions of two signals in the NMR spectrum with even greater accuracy than this. Consequently, a reference signal is chosen, and the difference between the position of the signal of interest and that of the reference, termed the chemical shift, is assessed. The chemical shift is usually expressed in terms of parts per million (ppm), actually a dimensionless number, by

$$\delta = \frac{\nu_{ref} - \nu_{sample}}{\nu_{ref}} \times 10^6$$

Equation 7.2.2

where the difference between the resonance frequency of the reference and that of the sample nucleus ($\nu_{ref} - \nu_{sample}$) measured in hertz (e.g., 90 Hz) divided by the spectrometer's operating frequency (e.g., 600 MHz) gives the chemical shift (e.g., 0.15 ppm). Typical ranges in chemical shifts for nuclei in biochemically important samples are 15 ppm for ^1H, 250 ppm for ^{13}C, 400 ppm for ^{15}N, and 35 ppm for ^{31}P.

A nucleus with a magnetic moment may interact with other nuclear spins resulting in mutual splitting of the NMR signal from each nucleus into multiplets. This is termed scalar coupling. For ^1H, ^{13}C, ^{15}N, and ^{31}P, the number of components into which a signal is split is $n + 1$, where n is the number of other nuclei interacting with the nucleus. For example, a nucleus (e.g., ^{13}C or ^1H) interacting with three methyl protons will give a signal split into a quartet. To a first approximation, the relative intensities of the multiplets are given by binomial coefficients: 1:1 for a doublet, 1:2:1 for a triplet, and 1:3:3:1 for a quartet. The difference between any two adjacent components of a multiplet is the same and yields the value of the spin-spin coupling constant J (in hertz).

One important feature of spin-spin splitting is that it is independent of magnetic field strength. Therefore, increasing the magnetic field strength will increase the chemical shift difference between two peaks in hertz (not parts per million) but will not alter the coupling constant. To simplify a spectrum, especially with ^{13}C and ^{15}N NMR, it is common to employ decoupling. Strong irradiation of the protons at their resonance frequency will cause a collapse of the multiplet in the ^{13}C or ^{15}N resonance into a singlet. In oligonucleotide samples, the signal width may be broader than the splitting, so the latter may not always be apparent. However, measurement of J can provide valuable structural information (as discussed below; see section on Acquisition of Torsion-Angle Restraints).

While it is possible to obtain chemical-shift and coupling-constant information from a one-dimensional (1D) NMR spectrum for small organic molecules, it is usually necessary to employ 2D (or sometimes 3D and 4D) NMR for olignonucleotides. Just as a biochemist may employ two-dimensional gel electrophoresis to spread out peaks in two dimensions, it is possible to resolve the many additional NMR sig-

nals coming from larger biopolymers by utilizing additional dimensions. While the plethora of different possible ways to achieve this can be complicated, basically, by using different combinations of pulses of RF energy, it is possible not only to spread the signals out over two or three dimensions, but also to establish relationships between the nuclei giving rise to the different signals. Nuclei that are scalar-coupled to one another can be identified by the appearance of cross-peaks in the 2D NMR spectrum. Nuclei that are in close proximity in the molecule can be identified via the nuclear Overhauser effect manifest in signal intensities.

ACQUISITION OF INTERPROTON DISTANCE RESTRAINTS

All structural determinations to date have utilized interproton distance restraints from NOE data. Some have also utilized torsion-angle restraints from the oligonucleotide backbone or sugar ring.

2D NOE Spectral Requirements

The intensities of the cross-peaks in a 2D NOE spectrum are related to the distances between protons that are in close spatial proximity (5 to 6 Å) in a structure, and can be used to estimate those distances. One might anticipate that accurate cross-peak intensities would be desired in order to obtain accurate distances. Acquiring intensities accurately reflecting the interproton distance generally requires that (1) the pulse repetition time (T_R) of the experiment be long compared with the longest proton T_1 (spin-lattice relaxation time) value, i.e., such that relaxation be nearly complete; (2) contributions to the cross-peaks from multiple quantum coherences be minimal; (3) good flat baseplanes in the well-digitized spectrum be produced to enable good integration; and (4) peak overlap be minimal (or very reliable deconvolution software be available for spectral analysis). Note that these requirements are not very stringent. For example, if (to take an arbitrary figure) an error of 5% in the measured interproton distance r were acceptable, an error in cross-peak intensity of ~35% could be tolerated. (This assumes only the first-order relationship that cross-peak intensity varies as r^{-6}, as discussed below; see section on Extracting Interproton Distances from 2D NOE Cross-Peak Intensities). The adenine H2 proton in a typical DNA duplex has a T_1 value of ~5 sec, and T_1 values are even longer in RNA. Using a T_R of 2 sec, as commonly seen in the literature, would a render cross-peak involving an adenine

H2 proton only 14% of its fully relaxed intensity. As other peaks will also have attenuated intensities, resulting distance errors will not be as large as this might imply, since relative peak intensities are used in making distance estimates (also discussed below). However, our lab has typically used T_R values of 8 to 12 sec to minimize intensity distortions due to incomplete relaxation. More recently, it has also become possible to correct for intensity distortions caused by using a short T_R (Liu et al., 1996).

Due to the dynamic range problem, acquiring 2D NOE spectra for oligonucleotides in H_2O solution, which is required for observation of imino and amino proton resonances, is more trouble than obtaining spectra for oligonucleotides in a D_2O solution, which is sufficient for nonexchangeable proton data. Most structural studies of oligonucleotides have not utilized distances entailing exchangeable protons. That has generally been wise, since imino and amino protons of nucleic acids exchange with bulk water and the intensity of 2D NOE cross-peaks involving exchangeable protons can be strongly altered by that exchange. If such exchange is not taken into account, calculation of distances to exchangeable protons is prone to significant error (Landy and Rao, 1989). However, the additional structural information that can be obtained by using such distances is extremely valuable. For example, cross-strand distances define the spatial relationship between the two strands in duplex structures, but there are very few such distances entailing solely nonexchangeable protons. The bases in DNA and RNA have a paucity of protons, so structural data provided by those imino and amino protons is far more valuable than structural data emanating from the proton-rich sugars. We have shown that we can incorporate effects of exchange with bulk water in the case of exchangeable protons and hence extract these important distance restraints (Liu et al., 1993). There is a simple, useful method for measuring the exchange rate that we have found useful (Adams and Lerner, 1992); I strongly recommend making the effort, since the information gained is so valuable.

Extracting Interproton Distances from 2D NOE Cross-Peak Intensities

The effect of cross-relaxation between two neighboring protons during the mixing time period τ_m of the 2D NOE experiment is to transfer magnetization between them (Macura and Ernst, 1980). The efficiency of this transfer

depends on (1) the length of τ_m, (2) the distance between the two protons, (3) surrounding protons, and (4) the rate of molecular motion, generally characterized by a correlation time τ_c (Keepers and James, 1984). With transfer of magnetization, the cross-peak intensities in the spectrum will be modified. Consequently, cross-peak intensities have structural information—i.e., distances—embedded in them. In interesting molecules, the two protons giving rise to a particular cross-peak are not the only protons in the molecule. Rather, they belong to an array of all protons in the molecule that, in principle, experience dipole-dipole interactions with all the others. So cross-relaxation between the two protons is part of a coupled relaxation network. There are different methods of analyzing 2D NOE spectra to obtain interproton distances. The commonly employed two-spin or isolated-spin-pair approximation (ISPA) can be used:

$$r_{ij} = r_{ref}(a_{ref}/a_{ij})^{\frac{1}{6}}$$

Equation 7.2.3

where r_{ij} is the interproton distance to be estimated and a_{ij} is the corresponding 2D NOE cross-peak intensity, while r_{ref} and a_{ref} are, respectively, a known interproton distance (e.g., cytosine H5-to-H6 distance = 2.46Å) and its cross-peak intensity. This equation results from truncation after the linear term of the Taylor-series expansion of the complete relaxation rate expression accounting for all proton dipole-dipole interactions:

$$\mathbf{a}(\tau_m) = e^{-\mathbf{R}\tau_m} = 1 - \mathbf{R}\tau_m + \left(\frac{1}{2}\right)\mathbf{R}^2\tau_m^2 - \cdots$$

Equation 7.2.4

where **a** is the matrix of 2D NOE intensities and **R** is the matrix describing the complete dipole-dipole relaxation network (i.e., all interproton interactions). The off-diagonal element R_{ij} (row i, column j) of the relaxation matrix **R** corresponds to the cross-relaxation rate between protons i and j due to the fluctuating (from Brownian motion) dipolar interaction between them, and hence depends on the internuclear distance as $(r_{ij})^{-6}$ and on the rate of reorientation of the ij vector relative to the magnetic field in which the nucleic acid sample has been placed.

The truncation leading to Equation 7.2.3 is valid in the limit of short mixing time, which is equivalent to assuming each cross-peak intensity depends only on the cross-relaxation rate between the two pertaining protons, i.e.,

ISPA. ISPA can lead to sizable systematic, as well as random, distance errors due to multispin effects commonly called "spin diffusion" (Keepers and James, 1984; Borgias and James, 1989; Post et al., 1990). The general method of assessing the impact of spin diffusion has been to obtain 2D NOE buildup curves, i.e., cross-peak intensities as a function of mixing time. Spin diffusion is usually considered negligible if the buildup is initially linear. Constructing an NOE buildup curve requires that several NOE spectra be analyzed, and still cannot properly account for multispin effects (Keepers and James, 1984; Post et al., 1990; Thomas et al., 1991). For example, for mixing times generally accepted as being sufficiently short (i.e., 50 to 100 msec) and not including internal motions, ISPA can result in systematic errors of 45% to 80% in actual distances >3.5 Å, a range that is quite important in defining molecular structure. Limiting analysis to 2D NOE spectra with short mixing times also yields cross-peaks with lower signal-to-noise ratio, limiting the number of distances that can be extracted.

Improving the accuracy of distances derived from NOE cross-peak intensities requires consideration of all structure-dependent relaxation pathways of the entire proton system, which can be done with complete relaxation matrix methods (Keepers and James, 1984; Borgias and James, 1989; Post et al., 1990). For example, theoretical NOE spectra including spin-diffusion effects can be calculated for a proposed structure. Conversely, this approach can be used to compute accurate distances from NOE intensities acquired at any mixing time. Rearrangement of Equation 7.2.4 leads to

$$\mathbf{R} = \frac{-\ln\left[\dfrac{\mathbf{a}(\tau_m)}{\mathbf{a}(0)}\right]}{\tau_m}$$

Equation 7.2.5

Complete relaxation matrix methods utilize linear algebra, so Equation 7.2.5 can be solved by matrix diagonalization techniques to yield the relaxation matrix **R** (Borgias and James, 1989, 1990). Some cross-peak intensities from NOE spectra of nucleic acids cannot be measured due to peak overlap and spectral signal-to-noise limitations, so it has been found expedient to utilize a hybrid matrix approach, where all experimentally unobserved NOEs are taken from a model structure that is presumed to be similar to the actual molecular structure (Boelens et al., 1989; Borgias and James, 1989, 1990; Post et al., 1990). The equation using the

hybrid matrix composed of experimental and model intensities is solved iteratively. It has been found that the model structure utilized is not very significant (Borgias and James, 1989, 1990). Exact details of methods differ, but the elements of the relaxation matrix are varied until a consistent fit to the experimentally observed NOEs is obtained. The programs IRMA (Boelens et al., 1989) and MORASS (Post et al., 1990) integrate this process directly into the conformational search for the final structure via rMD, but the program MARDIGRAS (Borgias and James, 1989, 1990; Liu et al., 1995) simply varies cross-relaxation rates until the best solution is found. Interproton distances are readily obtained from the converged relaxation matrix elements.

This latter approach offers the advantage that the conformational search methodology, force field, or penalty function does not influence the structural restraints, so the subsequent conformational search uses independent restraints that have not been biased by the search methodology itself. An additional advantage is that it is not necessary that all distances be satisfied by a single structure. A set of MARDIGRAS-derived distances can exhibit some mutually inconsistent restraints resulting from conformational flexibility and leading to dynamically averaged, and therefore possibly inconsistent, distances (Ulyanov et al., 1995). Consequently, examination of these distance inconsistencies may aid in recognition of and insights about conformational flexibility (Ulyanov et al., 1995; Tonelli and James, 1998). More information and access to MARDIGRAS can be found at the web page *http://picasso.ucsf.edu/software.html.*

An important aspect of the MARDIGRAS algorithm is that all intensities are used to obtain all distances simultaneously—i.e., the entire cross-relaxation network is employed; this is in distinct contrast to the old paradigm of a one-to-one mapping of intensities and distances. This means that a cross-peak that one would ordinarily not observe at low mixing times can be reliably examined at longer mixing times; even though the cross-peak intensity itself may be dominated by so-called spin diffusion, the cross-relaxation rate and consequently distance may be reliably inferred from the overall pattern of intensities, since the existence of that pair of protons will influence the intensities of other cross-peaks (Borgias and James, 1990; Liu et al., 1995).

Any method used for calculating interproton distances from NOE data requires an estimate of the correlation time. A detailed assessment of methods for evaluating correlation times has been presented elsewhere (Lane, 1995). In general, an isotropic correlation time is assumed for the overall tumbling of a nucleic acid. This assumption is reasonable up to a length of ~15 bp in a duplex, which might be considered as a cylinder. The assumption that one can treat a biopolymer as a sphere works quite well, as long as the molecule is not so distended that the ratio of the long molecular axis to the short molecular axis exceeds ~2.5.

Whenever possible, it is best to estimate the correlation time using ^{13}C relaxation-time measurements, since proton relaxation times are influenced by the multispin effects noted above. Natural-abundance ^{13}C relaxation-time measurements in nucleic acids can be time consuming, however. While subject to error, obtaining an estimate of τ_c from the ratio of T_1 and T_2 relaxation times for individual resolved protons is perhaps the easiest approach. A recent comparison of correlation times obtained using these two methods yielded comparable values (Tonelli et al., 1998). Although this rule of thumb has not been published anywhere, one can obtain a decent estimate of the correlation time for tumbling of a nucleic acid or protein in aqueous solution at approximately room temperature by taking the value of the molecular mass (daltons) in picoseconds and dividing by 2; for example, a structured 14,000-Da biopolymer has $\tau_c \approx 7$ nsec. Fortunately, τ_c accuracy is not paramount: an error by a factor of 2 translates to distance errors of no more than 12%, usually much less. Nevertheless, it is advisable to use both upper and lower bound estimates for τ_c in calculating distance restraints.

Estimation of Interproton Distance Bounds

When searching conformational space to find structures consistent with a set of experimental data, it is necessary to have an estimate of the accuracy of the distance restraints. Error estimates on restraints are absolutely required for setting bounds in DG or flat-well size in rMD calculations. The refinement process should yield a final structure in the context of a conformational envelope reflecting the intrinsic limitations of the experimental data and the method. Estimates of the interproton distance error, reflected by the upper and lower bounds assigned, vary widely throughout the literature. Tighter distance bounds (smaller error bars) lead to a higher-resolution structure; however,

distance bounds made tighter than warranted by experimental accuracy mislead to a highly precise (small atomic root-mean-square-deviation, RMSD) but incorrect structure (Thomas et al., 1991). So one needs bounds as tight as possible but not so tight that the real distance can lie outside them. What this means is that for some purposes it may be perfectly reasonable to utilize ISPA to estimate distances, but the spread between upper and lower bounds should be made quite large.

More accurate distances obtainable via MARDIGRAS minimize the possibility of an estimated distance lying outside the bounds. MARDIGRAS can also aid our choice of bounds, set individually for each proton pair. In particular, it is recommended that bounds be set using the RANDMARDI option now incorporated into the program (Liu et al., 1995). This procedure uses an absolute noise error (e.g., conservatively using the size of the smallest NOE) and an additional relative error (e.g., 10% to 20%) in distance bounds determination. In addition, different motional models with varying correlation times are used in combination with different starting geometries. With this procedure, typical error bars amount to ~10% for 3-Å distances but may be >30% for 5-Å distances, which derive from cross-peaks with lower signal to noise. The compensation for the broader bounds is that distances up to 7 Å (entailing methyls) can be reliably determined (Liu et al., 1995).

Distances are determined from a single 2D NOE spectrum using MARDIGRAS rather than using buildup curves. It is still desirable to obtain spectra at a few different mixing times. Depending on the internuclear distance and proton environment, different mixing times will be optimum for different proton pairs. This enables distances determined independently from spectra acquired at different mixing times to be compared. Of course, the range of distance values measured from different spectra for any given proton pair can aid in the choice of bounds. MARDIGRAS calculates bounds for distances to protons undergoing motional or overlap averaging, i.e., methyl, methylene, and aromatic protons (Liu et al., 1992). With MARDIGRAS, distances are calculated from individual cross-relaxation rates in the converged rate matrix. Distances entailing protons averaged by motion or spectral overlap may be in serious error if the averaging is ignored. The cross-relaxation rates in these cases will depend on orientation as well as distance. MARDI-GRAS does a second level of iteration, varying the orientation and distances of all dipole-dipole interactions to find the best fit (Liu et al., 1992). Most importantly, however, MARDI-GRAS lists the distances corresponding to the worst-case geometries, enabling upper and lower bounds to be set for distances involving protons averaged by either overlap or internal motions. As noted above, the effects of chemical exchange should also be taken into account in setting bounds for exchangeable protons (Liu et al., 1993).

ACQUISITION OF TORSION-ANGLE RESTRAINTS

Bond torsion angles can be determined using Karplus correlations with vicinal coupling constants, which can be derived from various correlation spectroscopic techniques—e.g., E.COSY, PCOSY, and double-quantum-filtered COSY (2QF-COSY; Piantini et al., 1982; Marion and Bax, 1988). In principal, these should make it possible to determine deoxyribose-ring pucker conformations. A detailed description of the methodology (Schmitz and James, 1995) should be consulted by anyone planning on employing this technique. Broad lines prevent direct analysis of nearly all coupling constants in DNA oligomers greater than ~8 bp in length. Consequently, most labs determine only those that can be readily measured, or determine sums of coupling constants—e.g., $\Sigma H2'$, the sum of all coupling constants involving the H2' proton of a deoxyribose ring. We have found that limiting analysis to these can leave ambiguities, but fitting of experimental cross-peaks to cross-peaks simulated using the program SPHINX (Widmer and Wüthrich, 1987) enables extraction of vicinal coupling constants and, subsequently, torsion-angle restraints (Celda et al., 1989; Schmitz et al., 1990; Gonzàlez et al., 1994). This entails extensive peak-shape analysis, and typical errors range from ±0.3 Hz for the best-defined coupling constants, i.e., $J_{H1'H2''}$ and $J_{H1'H2'}$, to ±1 Hz or more for $J_{H3'H4'}$. The major difficulty is in establishing the correct linewidth to be employed. However, the choice of linewidth can usually be constrained such that limits (i.e., the bounds) on torsion angles describing sugar pucker can be established.

Dipolar effects can influence the scalar-coupling-constant values (Harbison, 1993), with the effects becoming significant at larger correlation times (Zhu et al., 1995). Calculations (most unpublished) have shown that the dipolar effects are small for correlation times <5 nsec—i.e., not significant compared with experimen-

tal errors. This means that for room-temperature studies of duplexes up to ~14 bp in length, dipolar effects on the sugar three-bond coupling constants will not be detectable. For larger nucleic acids, the temperature can be raised to keep τ_c <5 nsec (Conte et al., 1996). For larger nucleic acids, it is also possible to obtain more limited, but useful, information: the sum of the vicinal coupling constants for a particular proton, e.g., $\Sigma H1'$, is not affected by dipolar relaxation due to compensatory effects for individual coupling constants, $J_{H1'H2''}$ and $J_{H1'H2'}$ (Conte et al., 1996).

The torsion angles for deoxyribose rings are determined using a parameterization of the relationship between torsion angles and coupling constants (Rinkel and Altona, 1987; Wijmenga et al., 1993; Gonzàlez et al., 1994; Schmitz and James, 1995). In studies of about a dozen DNA and RNA·DNA duplexes, examining all scalar coupling–based cross-peaks, it has been found that a single conformer will rarely account for all measured coupling constants within a sugar ring, but reasonable fits are obtained with a two-state model representing a rapid interconversion between S- (i.e., C2'-endo) and N-type (i.e., C3'-endo) sugar puckers (Altona and Sundaralingam, 1972). It should be evident in considering the shape of a Karplus curve (vicinal coupling constant versus dihedral angle) that conformational exchange will not average all coupling constants the same, so that if one has measured them all they cannot simultaneously fit a single conformation. This highlights the utility of using the more time-consuming SPHINX fitting of spectra to extract more coupling-constant data; with just a few coupling constants from each sugar ring, the data can be fit to a single averaged conformer, but one that will not necessarily correspond to any of the exchanging conformers (Gonzàlez et al., 1994). To employ the simplest model to account for the data, one can employ the two-state model. The pseudorotational phase angles of the major S- and minor N-conformers (P_S and P_N, respectively), as well as the relative populations and amplitudes of these conformers, are typically derived. For nonterminal nucleotides, we generally find that the S-conformer dominates, being populated 70% to 95% of the time with some exceptions.

STRUCTURE REFINEMENT

The ability to determine solution structures by NMR is limited by the quantity, accuracy, and distribution of distance and torsion-angle restraints that can be extracted from the NMR data. Obtaining enough restraints is a paramount requirement and will significantly counterbalance a lack of precisely determined restraints. As noted, interproton distances are essential for structure determination and a complete relaxation matrix approach will enable more numerous, more accurate, and more precise distances to be used.

In principle, one should try to determine all structures that will satisfy the structural restraints taking into account experimental error—i.e., upper and lower bounds are specified, so that the accessible conformational space can be mapped out. Basically, a large number of structures are derived, all of which will fit the data. This is a source of irritation for scientists accustomed to the single structure typically reported in x-ray or modeling studies. (With arm-twisting, however, it is possible to come up with a single structure that will satisfy the biochemist or modeler.) In some sense, the family of structures gives a fuzzy picture of the structure. With more restraints, more accurate restraints, and tighter (but not unreasonable) bounds, the structural picture becomes sharper.

Methods entailing systematic searches of conformational space have been advocated, but for biopolymers they are computationally too expensive at present. So an intelligently restricted search of conformational space is performed using a conformational search engine, which attempts to find a global minimum in fitting a structure with the requisite primary sequence connectivities, bond lengths, and bond angles and experimental restraint bounds; it is also common to include additional chemical force-field information as well. While it is possible simultaneously to refine restraints and structure iteratively, I advocate separating the restraint-determination step and the structure-generation step. This should diminish any possibility of being trapped in a local energy minimum in the vicinity of early-iteration structures that are trying to satisfy inaccurate restraints. Also, individual assessment of the bounds for each structural restraint can be performed and then used in the structure-refinement procedure. And, as noted, one may also discern conformational exchange via inconsistent but accurate distance values.

DG and rMD (also known as simulated annealing) calculations are commonly used for biopolymer structure generation and refinement. Although DG has been employed for nucleic acid structure determination, it typically produces structures that need subsequent refinement to account for molecular energetic

considerations, generally via rMD calculations, to obtain energetically feasible structures.

Restrained Molecular Dynamics Simulations

The most commonly employed search technique is rMD (e.g., AMBER, GROMOS, or XPLOR). It should be understood that an rMD simulation is not the same as an MD simulation, which is generally concerned with following the molecular motions of a molecule over the femtosecond-to-nanosecond time scale. Rather, rMD uses the mathematics of MD to search conformational space: the kinetic energy in an MD simulation allows energy barriers of amplitude approximately kT (the Boltzmann constant multiplied by the absolute temperature) to be surmounted in a search for the global energetic minimum that maintains a balance between the "classical" energy terms and the experimental restraints. In rMD, the empirical force field is modified to incorporate a penalty term for not matching experimental NMR restraints:

$$V_{total} = V_{bond} + V_{angle} + V_{dihedral} + V_{vander Waals} + V_{coulomb} + V_{Hbond} + V_{NOE} + V_{J coupling}$$

Equation 7.2.6

The first five terms monitor the classical potential energy of the molecule. There may or may not be an explicit term used to maintain hydrogen bonds; restraints may be experimentally justified, for example, by observation of particular imino proton signals in a spectrum. The final two terms serve as penalty functions, monitoring the NOE-derived distance restraints and scalar coupling–derived torsion-angle restraints, respectively. Some studies have not incorporated torsion-angle restraints, but all have utilized distance restraints. Different functional forms have been employed for these penalty functions, but the consensus is a flat-well potential with quadratic boundaries beyond the experimental upper and lower distance bounds; for example, that is found in AMBER. The bounds determined via MARDI-GRAS can thus determine the distance range for each individual proton pair over which no penalty will be exacted. More information regarding access to AMBER, X-PLOR, and GROMOS can be found at the web pages, respectively, *http://www.amber.ucsf.edu/amber/amber.html, http://xplor.csb.yale.edu/xplor-info.html,* and *http://igc.ethz.ch/gromos.*

While force constants to be used for the conventional energy terms are well established, that is not the case for the experimental restraint-energy terms. We have consequently examined the effects of changing the force constant for the restraint-energy terms in some detail for four different DNA duplexes. While the apparently optimum value varies somewhat for different duplexes, as long as the force constant is in the range 10 to 40 kcal mol^{-1} Å^{-2}, for simulations near 300 K the exact value chosen is not too important. On the basis of our experience, I would recommend that a value of 20 kcal mol^{-1} Å^{-2} be used in the absence of an evaluation for any particular nucleic acid; this seems to work fine for RNA as well. The force-constant value for torsion angle–restraint violations should be about four times larger than that for distance-restraint violations. Of course, in a simulated annealing protocol where the temperature is raised, the force constants should be commensurably increased.

For rMD simulations, starting structures should be chosen in different regions of conformational space: e.g., for a DNA duplex, one would use A-DNA, B-DNA, and possibly some other model structure. In addition, different random initial trajectories should be used for each starting structure. The "final" structure reported for any particular rMD run is in fact not from the final step in the calculation. A trajectory (rMD simulation) might typically run for 30 psec in 1-fsec steps. An average set of atomic coordinates can be obtained from the coordinates of say 200 structures in the last 5 psec of the run when the search has stabilized. As this average set of coordinates may not represent a reasonable structure, that structure is subjected to restrained minimization—i.e., the energy is minimized with inclusion of the experimental restraint violation terms. For any method, successful refinement requires convergence: i.e., essentially the same final structure is obtained from different starting models and starting trajectories using a reasonable search protocol (for a detailed discussion see Schmitz and James, 1995). In reality, the structures resulting from different rMD runs will not be identical, but they should be similar—i.e., with small atomic RMSD between the individual structures little bigger than atomic displacements from librational motions (~0.5 Å); generally <1.0 Å is considered adequate.

For rMD, it is typically difficult to obtain convergence in reasonable computational times if the starting structure lies far from the final structure. For proteins, it has been found most

efficient to use DG initially and then utilize rMD for subsequent refinement. For nucleic acid duplexes, using any right-handed helix, e.g., A or B form, is sufficient for rMD. Most DG programs do not work at all or do not work well for nucleic acids. For other RNA or DNA structures, we have found the program DYANA to work well and efficiently. DYANA performs torsion-angle dynamics, gaining its efficiency from using internal rather than Cartesian coordinates, which decreases the number of parameters by an order of magnitude. It is best to use rMD, e.g., with AMBER, for final refinement of each of the acceptable structures found by running DYANA. More information regarding access to DYANA can be found on the web page *http://www.mol.biol.ethz.ch/dyana*.

The envelope of individual structures resulting from the multiple conformational search runs, which appears as a "fuzzy" structure, is perhaps a better representation of reality, but it is typically simpler to display a single final structure. That is obtained from the "final" structures resulting from the different trajectories being averaged and that average structure subjected to restrained minimization.

There are methods other than those mentioned above for generating structures from NMR data, but there are few cases where different structure-refinement methods have been independently applied to the same data set and the resulting structures compared. A Monte Carlo search in torsion-angle space has been developed using generalized helical parameters, rather than Cartesian coordinates, to define DNA conformation for efficiency (Ulyanov et al., 1993). Using an idealized geometry for the aromatic rings eliminates the small distortions sometimes observed with structures emanating from rMD, which result from the force field permitting some distortion of bond angles and lengths in a compromise to fit the experimental restraints. Restrained Monte Carlo (rMC) calculations have been used on two DNA duplexes (Ulyanov et al., 1993; Tonelli et al., 1998). Convergence of final structures via rMC is easily achieved from A-DNA, B-DNA, or other right-handed DNA models (atomic RMSD <0.3 Å for all rMC simulations in one study) without use of torsion-angle restraints. For rMD, it is difficult (but sometimes possible) to achieve convergence to a global minimum starting from A-DNA without torsion-angle restraints. The structures resulting from rMC and rMD are in agreement—RMSD <0.5 Å for all structures generated—despite the different force fields used and despite the fact that the

rMD calculation used deoxyribose torsion-angle restraints. The structures generated with rMC and rMD were both in reasonable accord with experimental 2D NOE peak intensities and with experimental scalar-coupling data from 2QF-COSY spectra. Agreement with the scalar-coupling data is particularly satisfying, since those data were used in the rMD refinement but not the rMC refinement.

ASSESSMENT OF STRUCTURE QUALITY

It is clear that the methodology exists to determine the structure of a small nucleic acid. Beyond adhering to the cautionary notes cited in the sections above, how can one be confident that the structure is correct? A detailed discussion of this question has been presented (James, 1994).

Number of Restraints

We have examined the effect of the number of restraints available on the structure determined in one DNA duplex where we had on average 20 distance restraints and 5 torsion-angle restraints per residue (Weisz et al., 1994). While a structure is fairly well restrained by ~10 distance restraints and 5 torsion angle restraints per nucleotide, we have found that it is better defined by ~15 distance restraints per residue, along with the torsion angle restraints. More than ~15 distance restraints provides redundant information, and the structure determined is little affected by additional restraints. It is assumed, of course, that the restraints are fairly evenly distributed across the molecule. If one has much less than ten restraints per residue, "structure determination" is the wrong term to apply; "modeling" would be more appropriate, as the chemical nature of the force field begins to dominate the resulting structure, with the experimental restraint data only providing some limits on that structure. As noted earlier, however, a model created with limited experimental restraints may still provide very valuable insight.

Atomic Root-Mean-Square-Deviation

It is common to cite a value of the atomic RMSD among the ensemble of structures fitting the experimental data (as has been done earlier in this unit). One should be cautious, however, about interpreting the RMSD in terms of the quality of the structure derived or comparing RMSD values from one study to another. In fact, one should be careful about trying to push the value too low. It is possible to make the RMSD smaller by various means—e.g., by

the use of larger force constants for the restraint violation terms in rMD simulations, by inadequate sampling in DG calculations, and the selection of very tight restraint bounds. The structural inaccuracy engendered by restraint bounds made tighter than the experimental accuracy has already been mentioned. While it provides some insight, the atomic RMSD is definitely not a measure of accuracy and is only a modest descriptor of precision (Shriver and Edmondson, 1993).

Another limitation of using the overall atomic RMSD to characterize a structure is demonstrated in Figure 7.2.3. Here NMR has been used to determine the structure of an RNA construct corresponding to the genomic dimer initiation complex (Mujeeb et al., 1998), which is essential for packaging and replication of the HIV-1 virus. The two identical 23-nucleotide RNA strands each form hairpin structures with the loops of each hairpin containing the hexameric CGCGCG palindrome. The loops from two hairpins, as well the stem regions, can thus form a helix in creating a so-called kissing-loop structure. The 34 structures resulting from NMR structure determination are shown with the kissing-loop residues superimposed at top left in Figure 7.2.3 (RMSD of heavy atoms of bases 10 to 15 is 0.63 ± 0.17 Å for the 34 structures relative to a global average structure) and with the stem residues superimposed at top right (RMSD of heavy atoms of bases 1 to 6
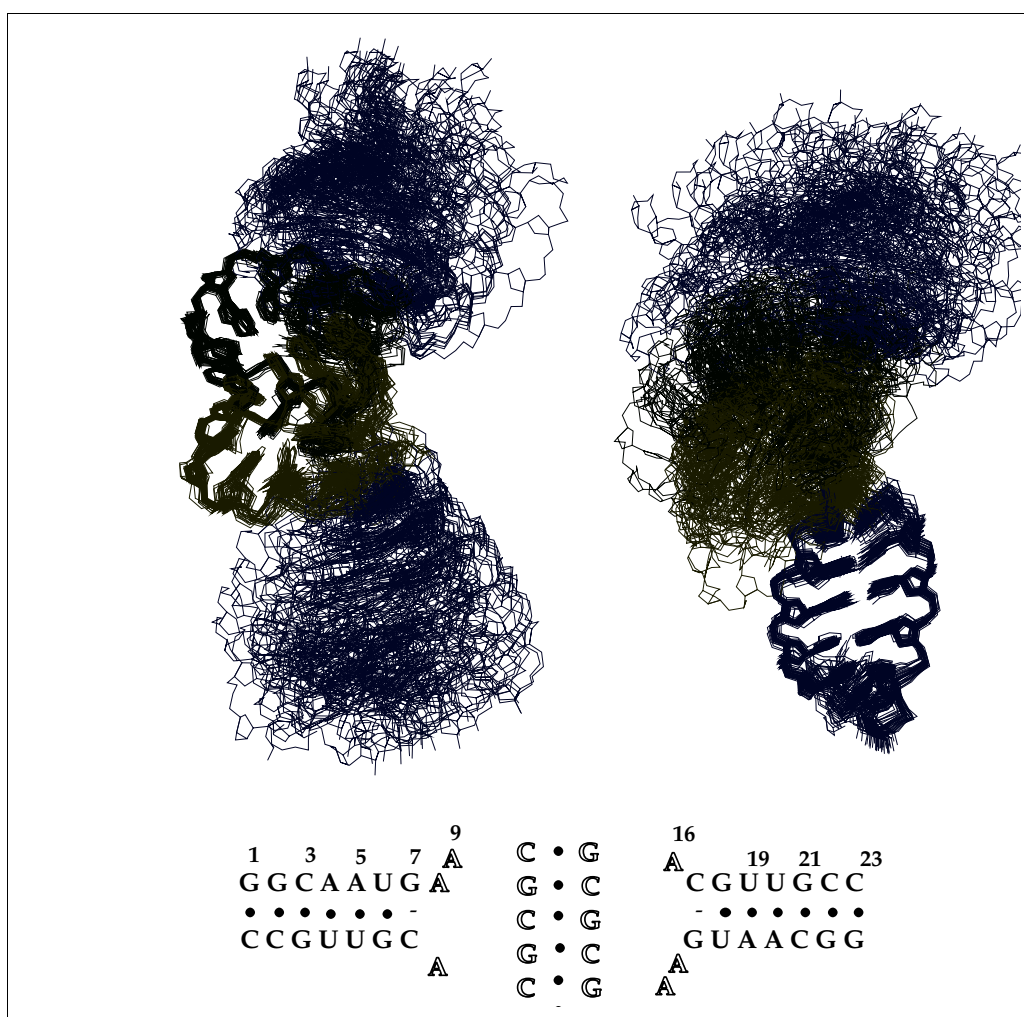

Figure 7.2.3 Structure of an RNA construct of the HIV-1 genome dimer initiation complex. The lower scheme shows the complex of 46 nucleotides composed of two identical 23-nucleotide RNA strands that each form a stem-loop. The two loops interact via the palindromic CG hexamer, creating a helix from the interacting loops—i.e., a kissing-loop structure is formed. The 34 structures resulting from NMR structure determination are shown with the kissing loop residues superimposed at top left and with the stem residues superimposed at the top right. As the dimer initiation complex is symmetric, the other stem region can also be superimposed.

and 18 to 23 is 0.87 ± 0.36 Å for the structures relative to a global average structure). As the dimer initiation complex is symmetric, the other stem region can also be superimposed with identical results. A casual glance at either the right or left ensemble of structures might lead one to conclude that a significant amount of the RNA was either unstructured or not determined very well. In fact, this appearance is due solely to the lower RMSD (1.63 ± 0.47 Å) in the region of the junction (G7, A8, A9, A16, and G17): some torsion angles in this junction region may possess some degree of flexibility or there may not be enough experimental restraints to define this region completely.

R-Factor Calculations

Other figures of merit can be used to assess NMR structures, however. For example, one can compare the final structure, or even interim structures, with the experimental NMR data. In x-ray crystallography, an evaluation of the fit of a derived structure with the original electron densities via a residual index (R factor) is requisite for any published structure. But such a comparison of the derived structure with the original data is done only occasionally for NMR structures. However, one can quantitatively compare calculated 2D NOE spectral intensities for any proposed molecular structure—obtained, e.g., with the program CORMA that is embedded within the MARDI-GRAS program package—with the experimental intensities. It would be possible to use a residual index analogous to the crystallographic R factors, but that gives equal weighting to all deviations between observed (a_o) and calculated (a_c) intensities such that stronger peaks (from very short distances, i.e., <2.5 Å) dominate. A sixth-root residual index R^x, however, permits longer interactions, e.g., ~4 to 5 Å, to have a role in the calculated R factor.

$$R^x = \frac{\sum_i |a_o^{1/6}(i) - a_c^{1/6}(i)|}{\sum_i a_o^{1/6}(i)}$$

Equation 7.2.7

Longer distance restraints are most important in structure determination, so it would be good to have their input into assessing structure quality. Unlike in x-ray crystallography, it is possible to define a subset of NOE cross-peaks and calculate R^x for some particular aspect of the structure, e.g., interresidue versus intraresidue or one selected region of the molecule. Such analysis is capable of indicating regions of good (or bad) fit between any model structure and the actual solution structure. In principle, one might wish to generate a structure with the lowest R factor possible. However, prudence is advisable in any efforts to push an R factor to its lowest value. R factors are also of limited value when assessing the accuracy of a refined model. A small restraint set can be easily overfit to a low R value, although the structure is only poorly defined by the data. Other complications arise from the nonrandom nature of errors in experimental intensities and the limited knowledge of molecular motions. Although the exact nature of the molecular motions is not a strong determinant in the calculations of the 2D NOE intensities, motions can still influence the intensities (Keepers and James, 1984). Values for NMR R factors should not be compared with those for x-ray R factors, since the total range of the x-ray R factor (with a maximum value of 0.83) is much more limited than for an NMR R factor (with a maximum value of ∞).

A free R factor R_{free} has been proposed as an unbiased indicator for assessing x-ray crystal structures and NMR structures (Brünger and Nilges, 1993). In the case of NMR, R_{free} measures the fit of a model structure's NOE intensities to a randomly selected set of experimental NOE intensities that are not used in structure refinement— ~10% of the total are in the test set; the other 90% are used in structure refinement. This avoids any model bias in calculating the R value. It is to be expected that R_{free} will be larger than the usual R factor, since the test data were not used in structure generation. However, discounting noise in the data, if the test data are consistent with the structure determined, R_{free} should not be too much larger. Determination of the exact amount that should be considered acceptable awaits more experience with real data. For one DNA decamer duplex, we found it to be 60% larger, and 50% to 100% larger has been reported for the few proteins examined so far.

Other Assessment Criteria

Consistency of the final structure with scalar coupling–based multidimensional NMR spectra may be assessed as well. For example, the RMSD between experimental and theoretical coupling constants can be calculated as

$$J_{\text{rms}} = (1/N)\sqrt{(J_{\text{exp}} - J_{\text{theor}})^2}$$

Equation 7.2.8

where the summation is over N, constituting all or any subset, of the coupling constants (Ulyanov et al., 1993). Of course, this assumes that a well-parameterized Karplus relationship has been established between the dihedral angles and three-bond coupling constants (Wijmenga et al., 1993).

A residual distance violation is the difference between a particular interproton distance in a structure and the closest of either the upper or lower bound. An evaluation of the structures resulting from any refinement technique should not exhibit substantial distance violations. Other criteria depend on the method of refinement. For structures resulting from rMD or rMC refinement, the restraints violation energies (V_{NOE} and $V_{J\ coupling}$ in Equation 7.2.6) should be low, and the total of the other terms should not be much greater than in the absence of any experimental restraints.

Any structure generated should, of course, be consistent with all other available experimental data on the oligomer in solution. This includes chemical as well as physical data.

ANALYSIS OF NUCLEIC ACID STRUCTURE

A detailed characterization of the resulting NMR structure should be performed, entailing determination of backbone torsion angles, sugar puckers, glycosidic torsion angles, and a series of helical (or base-orientation if not in a helix) parameters. The program CURVES 5.1 calculates structural parameters from the Cartesian coordinates of the atoms in a nucleic acid (Lavery and Sklenar, 1989, 1996). It calculates the structural disposition of individual bases—e.g., inclination, tip, x displacement, and y displacement—and uses that information to define a local helical axis. By building up from these local helical segments, the global helix axis

is constructed. That global axis may have kinks or curvature that can be examined and measured. Further information about use of CURVES can be found on the web page *http://plumber/csb.yale.edu/userguides/datamanip/curves/doc.html*.

Dials and Windows is a program package convenient for computing and graphing the structural parameters output by rMD (Ravishanker et al., 1989); it works ideally in concert with CURVES. Some information can be found on the Dials and Windows web page at *http://plumber.csb.yale.edu/userguides/graphics/dandw/dandw_descr*. Dials and Windows is one part of the Molecular Dynamics Toolchest developed at Wesleyan University.

An important reason for calculating the structural parameters is to assess the validity of structures generated by NMR. Unusual backbone torsion angles or sugar puckers could very well highlight inconsistencies among the restraints or even mistakes in the data that might have gone undetected by the assessment tools already mentioned. For example, helical parameters for duplex DNA solution structures should probably not deviate too much from canonical B-DNA or related crystal structures unless this is clearly justified by the original data. The same holds for double-helical regions of RNA, which should be very much like canonical A-form RNA. Table 7.2.1 lists some structural parameter values found for double-helical DNA; RNA should have the values of A-DNA. Extreme values of structural parameters or unusual conformations should be examined very carefully. It is necessary to scrutinize the restraints and the rMD protocol to be certain that the source of the unusual structural feature lies in reliable experimental data.

Table 7.2.1   Average Structural Parameters for DNA from Solution and Solid-State Data[a]

	Twist[b] (°)	Tilt (°)	Roll (°)	Shift (Å)	Slide (Å)	Rise (Å)	Propeller twist (°)
B-DNA[b]	36.1	0.0	−0.2	0.00	0.21	3.35	−13.8
	4.2	*3.6*	*5.6*	*0.55*	*0.75*	*0.24*	*6.6*
A-DNA[b]	30.8	0.0	7.9	0.00	−1.57	3.32	−9.8
	4.8	*3.3*	*5.6*	*0.52*	*0.38*	*0.31*	*5.6*
NMR[c]	35.3	−0.1	4.6	0.00	−0.34	3.15	−10.8
	4.2	*2.5*	*7.0*	*0.28*	*0.46*	*0.21*	*7.8*

[a]The first line of each entry gives the mean value, and the second line (in italics) gives the standard deviation in each individual set of data. Data are from a larger compilation that also lists sequence-dependence of the parameters (Ulyanov and James, 1995). The local helical parameters reported here conform to the Cambridge convention (Dickerson et al., 1989).
[b]From high-resolution crystal structures.
[c]From nine high-resolution NMR structures.

ACKNOWLEDGMENTS

LITERATURE CITED

Abragam, A. 1961. Principles of Nuclear Magnetism. Oxford University Press, Oxford.

Adams, B. and Lerner, L. 1992. A simple one-dimensional method for measuring proton exchange rates in water. *J. Magn. Reson.* 96:604-607.

Altona, C. and Sundaralingam, M. 1972. Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.* 94:8205-8212.

Boelens, R., Koning, T.M.G., van der Marel, G.A., van Boom, J.H., and Kaptein, R. 1989. Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach. Application to a DNA octamer. *J. Magn. Reson.* 82:290-308.

Borgias, B.A. and James, T.L. 1989. Two-dimensional nuclear Overhauser effect: Complete relaxation matrix analysis. *Methods Enzymol.* 176:169-183..

Borgias, B.A. and James, T.L. 1990. MARDI-GRAS—Procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure. *J. Magn. Reson.* 87:475-487.

Brünger, A. and Nilges, M. 1993. Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy. *Q. Rev. Biophys.* 26:49-125.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G., III, and Skelton, N.J. 1996. Protein NMR Spectroscopy: Principles and Practice. Academic Press, San Diego.

Celda, B., Widmer, H., Leupin, W., Chazin, W.J., Denny, W.A., and Wüfthrich, K. 1989. Conformational studies of d-(AAAAATTTTT)2 using constraints from nuclear Overhauser effects and from quantitative analysis of the cross-peak fine structures in two-dimensional ^{1}H nuclear magnetic resonance spectra. *Biochemistry* 28:1462-1470.

Conte, M.R., Bauer, C.J., and Lane, A.N. 1996. Determination of sugar conformations by NMR in larger DNA duplexes using both dipolar and scalar data: Application to d(CATGTGACGT-CACATG)2. *J. Biomol. NMR* 7:190-206.

Derome, A. 1987. Modern NMR Techniques for Chemistry Research. Pergamon Press, Oxford.

Dickerson, R.E., Bansal, M., Calladine, C.R., Diekmann, S., Hunter, W.N., Kennard, O., Lavery, R., Nelson, H.J.C., Olson, W.K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D.M., von Kitzing, E., Wang, A. H.-J., and Zhurkin, V.B. 1989. Definitions and nomenclature of nucleic acid structure parameters. *EMBO J.* 8:1-4.

Ernst, R.R., Bodenhausen, G., and Wokaun, A. 1987. Principles of Nuclear Magnetic Resonance in One and Two Dimensions. Clarendon Press, Oxford.

Gonzàlez, C., Stec, W., Kobylanska, A., Hogrefe, R., Reynolds, M., and James, T.L. 1994. Structural study of a DNA-RNA hybrid duplex with a chiral phosphorothioate moiety by NMR: Extraction of distance and torsion angle constraints and imino proton exchange rates. *Biochemistry* 33:11062-11072.

Harbison, G.S. 1993. Interference between J-couplings and cross-relaxation in solution NMR spectroscopy: Consequences for macromolecular structure determination. *J. Am. Chem. Soc.* 115:3026-3027.

James, T.L. 1994. Assessment of the quality of derived macromolecular structures. *Methods Enzymol.* 239:416-439.

James, T.L. (ed.) 1995. Nuclear Magnetic Resonance and Nucleic Acids. *Methods in Enzymology*, Vol. 261. Academic Press, New York.

Keepers, J.W. and James, T.L. 1984. A theoretical study of distance determinations from NMR. Two-dimensional nuclear Overhauser effect spectra. *J. Magn. Reson.* 57:404-426.

Landy, S.B. and Rao, B.D.N. 1989. Dynamical NOE in multiple-spin systems undergoing chemical exchange. *J. Magn. Reson.* 81:371-377.

Lane, A. N. 1995. Determination of fast dynamics of nucleic acids by NMR. *Methods Enzymol.* 261:413-35.

Lavery, R. and Sklenar, H. 1989. Defining the structure of irregular nucleic acids: Conventions and principles. *J. Biomol. Struct. Dyn.* 6:655-667.

Lavery, R. and Sklenar, H. 1996. CURVES 5.1. Helical Analysis of Irregular Nucleic Acids. Laboratoire de Biochimie Theoretique, Centre National de la Recherche Scientifique, Paris.

Leroy, J.L., Gehring, K., Kettani, A., and Guéron, M. 1993. Acid multimers of oligodeoxycytidine strands: Stoichiometry, base-pair characterization, and proton exchange properties. *Biochemistry* 32:6019-6031.

Liu, H., Thomas, P.D., and James, T.L. 1992. Averaging of cross-relaxation rates and distances for methyl, methylene and aromatic ring protons due to motion or overlap: Extraction of accurate distances iteratively via relaxation matrix analysis of 2D NOE spectra. *J. Magn. Reson.* 98:163-175.

Biophysical Analysis of Nucleic Acids

7.2.15

Liu, H., Kumar, A., Weisz, K., Schmitz, U., Bishop, K.D., and James, T.L. 1993. Extracting accurate distances and bounds from 2D NOE exchangeable proton peaks. *J. Am. Chem. Soc.* 115:1590-1591.

Liu, H., Spielmann, H.P., Ulyanov, N.B., Wemmer, D.E., and James, T.L. 1995. Interproton distance bounds from 2D-NOE intensities: Effect of experimental noise and peak integration errors. *J. Biomol. NMR* 6:390-402.

Liu, H., Tonelli, M., and James, T.L. 1996. Correcting NOESY cross-peak intensities for partial relaxation effects enabling accurate distance determination. *J. Magn. Reson. B* 111:85-89.

Macura, S. and Ernst, R.R. 1980. Elucidation of cross relaxation in liquids by 2D NMR spectroscopy. *Mol. Phys.* 41:95-117.

Marion, D. and Bax, A. 1988. P.COSY, a sensitive alternative for double-quantum-filtered COSY. *J. Magn. Reson.* 80:528-533.

Mujeeb, A., Clever, J.L., Billeci, T.M., James, T.L., and Parslow, T.G. 1998. Structure of the dimer initiation complex of the HIV-1 genomic RNA. *Nature Struct. Biol.* 5:432-436.

Piantini, U., Sørensen, O.W., and Ernst, R.R. 1982. Multiple quantum filters for elucidating NMR coupling networks. *J. Am. Chem. Soc.* 104:6800-6801.

Post, C.B., Meadows, R.P., and Gorenstein, D.G. 1990. On the evaluation of interproton distances for three-dimensional structure determination by NMR using a relaxation rate matrix analysis. *J. Am. Chem. Soc.* 112:6796-6803.

Ravishanker, G., Swaminathan, S., Beveridge, D.L., Lavery, R., and Sklenar, H. 1989. Conformational and helicoidal analysis of 30 ps of molecular dynamics on the d(CGCGAATTCGCG) double helix: "Curves," dials and windows. *J. Biomol. Struct. Dyn.* 6:669-699.

Rinkel, L.J. and Altona, C. 1987. Conformational analysis of the deoxyribofuranose ring in DNA by means of sums of proton-proton coupling constants: A graphical analysis. *J. Biomol. Struct. Dyn.* 4:621-649.

Schmitz, U. and James, T.L. 1995. How to generate accurate solution structures of double-helical nucleic acid fragments using nuclear magnetic resonance and restrained molecular dynamics. *Methods Enzymol.* 261:3-44.

Schmitz, U., Zon, G., and James, T.L. 1990. Deoxyribose conformation. *In* [d(GTATATAC)]$_2$: Evaluation of sugar pucker by simulation of double-quantum-filtered COSY cross-peaks. *Biochemistry* 29:2357-2368.

Shriver, J. and Edmondson, S. 1993. Defining the precision with which a protein structure is determined by NMR. Application to motilin. *Biochemistry* 32:1610-1617.

Thomas, P.D., Basus, V.J., and James, T.L. 1991. Protein solution structure determination using distances from 2D NOE experiments: Effect of approximations on the accuracy of derived structures. *Proc. Natl. Acad. Sci. U.S.A.* 88:1237-1241.

Tonelli, M. and James, T.L. 1998. Insights into the dynamic nature of DNA duplex structure via analysis of nuclear Overhauser effect intensities. *Biochemistry* 37:11478-11487.

Tonelli, M., Ragg, E., Bianucci, A.M., Lesiak, K., and James, T.L. 1998. NMR structure of d(GCATATGATAG).d(CTATCATATGC): A consensus sequence for promoters recognized by σK RNA polymerase. *Biochemistry* 37:11745-11761.

Ulyanov, N.B. and James, T.L. 1995. Statistical analysis of DNA duplex structural features. *In* Nuclear Magnetic Resonance and Nucleic Acids, Vol. 261 (T.L. James, ed.) pp. 3-44. Academic Press, New York.

Ulyanov, N.B., Schmitz, U., and James, T.L. 1993. Metropolis Monte Carlo calculations of DNA structure using internal coordinates and NMR distance restraints: An alternative method for generating high-resolution solution structure. *J. Biomol. NMR* 3:547-568.

Ulyanov, N.B., Schmitz, U., Kumar, A., and James, T.L. 1995. Probability assessment of conformational ensembles: Sugar repuckering in a DNA duplex in solution. *Biophys. J.* 68:13-24.

van de Ven, F.J.M. and Hilbers, C.W. 1988. Resonance assignments of non-exchangeable protons in B type DNA oligomers, an overview. *Nucl. Acids Res.* 16:5713-5726.

Weisz, K., Shafer, R.H., Egan, W., and James, T.L. 1994. Solution structure of the octamer motif in immunoglobulin genes via restrained molecular dynamics calculations. *Biochemistry* 33:354-366.

Widmer, H. and Wüthrich, K. 1987. Simulated two-dimensional NMR cross-peak fine structures for ^{1}H spin systems in polypeptides and polydeoxynucleotides. *J. Magn. Reson.* 74:316-336.

Wijmenga, S.S., Mooren, M.M.W., and Hilbers, C.W. 1993. NMR of nucleic acids; from spectrum to structure. *In* NMR in Macromolecules (G.C. Roberts, ed.) pp. 217-288. IRL Press, Oxford.

Zhu, L., Reid, B.R., and Drobny, G.P. 1995. Errors in measuring and interpreting values of coupling constants J from PE.COSY experiments. *J. Magn. Reson. A* 115:206-212.

Contributed by Thomas L. James
University of California
San Francisco, California