Dipolar Couplings in Partially Aligned Macromolecules - New Directions in Structure Determination using Solution State NMR.

Recently developed methods for the partial alignment of macromolecules in dilute liquid crystal media such as phospholipid bicelles (Sanders et al 1994, Tjandra & Bax 1997) filamentous phage (Hansen et al 1998) or purple membrane fragments (Sass et al 1999) provide a more general solution to the limitations of classical structure determination.

If a macromolecule experiences restricted orientational sampling, due to the presence of a liquid crystal or due to the paramagnetic properties of the molecule, strong first-order interactions, such as chemical shift anisotropy or dipolar coupling, are no longer averaged to zero as is the case in isotropic solution (Saupe & Englert 1963; Gayathri et al 1982).

The static dipolar Hamiltonian depends on the orientation $\theta$ of the internuclear vector between the coupled spins, relative to the magnetic field. This is a second order Legendre polynomial dependence ($P_2 \cos \theta$). In solution state NMR, the measured dipolar coupling, is described by the time, and ensemble, average of the dipolar Hamiltonian over all sampled orientations;

$$D_{ij} = -\frac{\gamma_i \gamma_j \mu_0 h}{8\pi^3} \left\langle P_2 \left(\cos \theta(t)\right) \right\rangle$$

Averaging of first order interactions in aligned and isotropic media

Dipolar coupling

$$H_{DD}(t) = I_k D I_l$$

CSA

$$H_{CSA}(t) = I_k \delta B_0$$

$D_{ik} = 0$

$D_{ik} = f(\theta, r^{-3})$
This averaging, denoted by the brackets $< >$, reduces the measured coupling to zero under conditions of purely isotropic averaging. Under conditions of partial ordering, where a preferential orientation of the molecule exists relative to the static field direction, this averaging is no longer reduced to zero.

While partial alignment will effect any first order phenomenon, the most powerful application of non-isotropic averaging is the measurement of residual dipolar coupling (RDC) (Tjandra & Bax 1997, Tolman et al 1995). The inherent strength of the dipolar coupling interaction allows readily measurable effects to be induced under conditions of weak ($10^{-3}$) alignment. If the alignment is weak enough, the solution properties necessary for high resolution NMR can be retained, while still allowing significant non-averaged first-order interactions to be measured under conditions of solution state spectroscopy.

**Inducing order in the sample.**

Molecules in solution can align naturally, due to their inherent paramagnetic properties, or due to the large anisotropic magnetic susceptibilities found for example in extended, regularly structures nucleic acid molecules. Recent measurements of residual dipolar couplings in partially aligned proteins were based on such systems (e.g. Tolman et al 1995, Tjandra et al
The main drawback to natural alignment is that the method is difficult to apply generally. There are however distinct advantages to studying systems containing a natural paramagnetic center with anisotropic magnetic susceptibility; in these systems complementary, distance dependent interactions can be measured between the observed spins and the paramagnetic spin (e.g. pseudo-contact dipolar shifts, curie-dipole cross correlation). In combination, these parameters, measured for NH sites in a 128 amino acid cytochrome c', have recently been shown to be sufficient to determine the overall fold of the protein.

More generally, order can be induced in the sample by immersing the molecule in a dilute liquid crystal solution. Examples include phospholipid bicelles, filamentous phage or purple membrane fragments (references above). Prof Grzesiek will describe some physical aspects of these alignment media in his presentation, so I will not describe these in any detail here.

Important aspects, which we will need to remember, are that the degree of order can be tuned as a function of concentration to achieve a suitable compromise between measurable couplings and spectral quality. This will eventually be reduced as the degree of order increases due to increased rotational diffusion of the solute and spectral complexity from multiple dipolar interactions. The nature of the interaction between the solute and the liquid crystal will also play an important role - as we will see later the ability to measure couplings in media which align the molecule differently is of fundamental importance. This implies non-steric interactions for the additional media (for example electrostatic repulsion/attraction).

**Residual Dipolar Couplings in Partially Aligned Macromolecules**

The simple averaging described in the equation (1) contains information about the average orientation of vectors over the ensemble and over time. The average is therefore a convolution of the restricted motion of the solute molecules, and the orientation of the vector with respect to the molecule. It is convenient to describe the measured couplings in terms of their
orientation to the fixed molecular frame, and describe the orientational averaging of the molecule in terms of an order matrix, or alignment tensor:
Assuming a fixed molecular shape, the resultant residual coupling can be expressed in terms of the orientation \( \{ \theta, \phi \} \) of the inter-nuclear vector relative to a common alignment tensor attached to the molecule;

\[
D_{ij} = -S \frac{\gamma_i \gamma_j \mu_0 h}{16\pi^3 r_{ij}^3} \left( A_a \left( 3\cos^2\theta - 1 \right) + \frac{3}{2} A_r \sin^2\theta \cos2\phi \right)
\]  

(1)

where \( A_a \) and \( A_r \) are the axial and rhombic components of the alignment tensor, \( r_{ij} \) the internuclear distance, and \( S \) the order parameter. In the case of coupling measured between covalently bound spins where inter-nuclear distance is constant, the geometric dependence is purely orientational, providing correlated directional information relative to the molecular coordinate frame. The ability to measure orientations for vectors positioned throughout the molecule represents a powerful tool, supplying conformational restraints of a different nature to the local constraints available from classical methods.

The second order alignment tensor is traceless and can be describing using 5 independent parameters, defining the orientation \( (\alpha, \beta, \gamma \) relative to the molecular frame) and degree \( (A_a \) and \( A_r \) of the average alignment experienced by the solute molecule. Molecular alignment is often described in terms of the Saupe matrix, with 5 independent elements \( (S_{xx}, S_{zz}, S_{xy}, S_{xz}, S_{yz}) \). This description is entirely equivalent and the elements can be linearly combined to determine the former description, which will be preferred here.
Orientational Dependence of Residual Dipolar Couplings

The available orientations of an interaction vector for a single measured residual dipolar coupling in the presence of a known tensor are shown on the left. There is clearly a strong angular degeneracy - a measured coupling with an intermediary value can be aligned either along the $\pm y$-axis, in the $\pm x\pm z$ plane, or an infinity of orientations in between. This degeneracy can be raised if we can measure more couplings in a domain of known structure, and whose relative orientation in the domain is known.

Estimation of the Alignment Tensor.

One additional point, which will need to be addressed at some stage in any structure calculation designed to exploit RDCs, is that the magnitude and orientation of the alignment tensor will not be known \textit{a priori}. If the structure of the molecule is known approximately the tensor may be estimated from the shape of the molecule (Ferrarini et al 1992, van der Est et al 1987, Zweckstetter et al 1999), assuming that the order induced exists only due to repulsive steric interactions with the liquid crystal surface. If the angular sampling of different vectors for which RDC's have been measured is sufficient then $A_x$ and $A_y$ can be estimated from a histogram of the distribution of values (this is essentially the Pake pattern representing all possible orientations of the interaction with respect to the $B_0$ field).
**Orientational degeneracy of RDC - Introduction of structural coherence.**

The orientational degeneracy continuum for a single coupling can be raised by measuring multiple couplings in a structural domain whose conformation is known. This is illustrated here:

On the right hand side of the figure shown above, we have sketched the equivalent orientations for an imaginary sub-structure consisting of differently oriented vectors. There are now four equivalent orientations of the differently valued couplings (colour coded in the figure) which are in agreement with measure values.

This four-fold degeneracy is inherent to the orientation of any three-dimensional structure relative to a molecular alignment tensor, and derives from simple symmetry operations (180° rotations around $A_{xx}$, $A_{yy}$ and $A_{zz}$).

Despite this inherent orientational degeneracy, the ability to determine domain orientation is a very powerful complement to classical structure determination and forms the basis of many recent studies of the molecular architecture of multidomain systems (Cai et al 1998, Skrynnikov et al 1999, Mollova et al 2000), and protein-ligand complexes (Weaver and Prestegard 1998, Olejniczak et al 1999 et al).
For example in the case of a two-domain complex shown here; the structures of the component parts of the molecule are assumed to be known, but not their relative orientation in the complex. If we measure sufficient couplings in each of the two domains, the alignment tensors for the two domains can be determined independently. {Note - Sufficient couplings here can theoretically be as few as 5; as this is the number of unknown parameters required to define the alignment tensor/Saupe matrix (see above). In reality, in the presence of noise, more couplings will be required to have sufficient accuracy.}

Once the tensors have been determined, assuming that the two domains experience identical alignment forces and are in a stable complex, we can simply align the molecules such that the component axes are coaxial. The four degenerate orientations of one of the domains are shown relative to the second, fixed domain in the above example.

We can of course check the validity of the assumption that the molecules experience similar alignment, by comparing the magnitude of the rhombic and axial components of the tensor.

These data are complementary to many readily available sources of structural constraint currently used to build models of molecular assemblies, whether these are experimental, such as intermolecular nOe measured between interacting surfaces (Clore 2000) or predicted from existing structural information, for example electrostatic or hydrophobic surface calculations.
Module - An Interactive Tool for Rigid-Body Modelling of Multi-Domain Macromolecules using Residual Dipolar Couplings

In the simplest applications, for example the study of the relative orientation of multi-domain systems and the investigation of molecular complexes, where individual component structures are known, RDC-driven rigid-body molecular mechanic or dynamic methods are therefore particularly appropriate.

Interpretation of residual dipolar couplings for the determination of relative domain orientation requires tools specifically developed for the manipulation of sub-structures within a reference calculation frame. As currently available graphical molecular modelling programs are not yet adapted to handling this kind of specific analysis, we have developed a program (Module), to facilitate this kind of analysis. The graphical interface allows the user to choose the couplings to be taken into consideration, to define modules from the primary sequence - these can be contiguous (like alpha helices) or non-contiguous (beta sheets) - and to check the quality of the fit from correlation diagrams.

Most importantly; for multi-domain systems, the program determines the relative orientation of individual structured domains. Module also provides graphical user-driven rigid-body modelling of the different modules relative to the common tensorial frame.

Translational freedom in the common frame, and equivalent rotations about the diagonalized (x,y,z) axes, can be used to position the different modules in the common frame to find the correct model in best agreement with experimentally measured couplings.

Covalent and non-bonded interactions can be tested and are used to propose the most likely structure, and distance restraints can be used in combination with the orientational data to find an optimal geometry of the oriented structural motifs.
Some Details About Module.

The alignment tensor will be calculated for each unit, and the unit considered structurally intact throughout the procedure. This region is not necessarily contiguous in primary sequence, for example in domain-swapped assemblies, nucleic acid structures (where paired strands may be taken as structurally inseparable) or multi-partner molecular complexes. This choice is performed using a simple cursor selection in the graphical interface.

Tensor eigenvalues and eigenvectors are then extracted using least-squares minimization of the target function over all couplings associated with a given domain:

\[ \chi^2 = \frac{\sum_n \left( D_{ij}^{\text{exp}} - D_{ij}^{\text{calc}} \right)^2}{\sigma_{ij}^2} \]

where \( \sigma_{ij} \) is the uncertainty in the experimentally measured coupling. The minimization algorithm searches the \( \{ A_x, A_y, \alpha, \beta, \gamma \} \) parametric space by random variation of these parameters, using a combination of simulated annealing (Metropolis et al. 1953), temperature regulation using fuzzy logic (Leondes 1997), and Levenberg-Marquardt minimization (Press et al. 1988) which we have previously developed for the determination of the rotational diffusion tensor from heteronuclear relaxation measurements (Dosset et al. 2000). The couplings are calculated with the appropriate pre-factors in equation (1), including the gyromagnetic ratio and the inter-nuclear distance, which can be chosen to be a either a standard fixed distance from an interactive table, or the actual distance (Å) present in the coordinate file.

The traceless molecular alignment tensor has an inherent degeneracy if \( A_x \) and \( A_y \) are allowed to take any values — to avoid confusion Module applies relevant transformations to place the minimum within the reference frame (\(|A_{xx}| < |A_{yy}| < |A_{zz}|, -\pi < \alpha, \gamma < +\pi, 0 < \beta < \pi\)). The three axes of these tensors are then superimposed graphically on the structural motifs and correlation plots presented for each different coupling type, as well as \( \chi^2 \) value for the fit of the RDC data for each module.

If we then assume that the different domains present in the molecule or complex experience negligible mobility relative to the each other, they will experience the same interaction with
the liquid crystal, and consequently the same aligning forces, and will therefore be governed by the same alignment tensor A. If the eigenvalues of the tensors determined for the separate domains are significantly different, the amplitude of the relative domain motion can no doubt be estimated, although an appropriate analysis is beyond the scope of this presentation (Fischer et al 1999).

Assuming similar eigenvalues, the relative orientation of the different sub-structures can be determined by aligning the domains such that all tensors are collinear (it should be remembered that we cannot exclude interdomain motion even if the eigenvalues are similar, and that in all cases inter-domain orientation representing the averaged couplings will be determined).

The program Module simply reorients each domain, and associated tensors, into a common graphical display frame (this can be considered to be the frame in which all tensors are diagonal). There is an inherent degeneracy of relative orientation present, due to the equivalence of any combination of vectors with respect to 180°; rotations about any of the alignment tensor axes (A_{xx}, A_{yy} and A_{zz}) (Al-Hashimi et al 2000).

These equivalent orientations can be viewed by the user, who can then position the different modules using the graphical interface (cursor-controlled) with respect to each other using only these equivalent orientations and three-dimensional translational freedom with respect to the diagonalized frame.

The entire coordinate space available with these degrees of freedom is equivalent with respect to the sum of the target functions (equation 2) for the different modules. In the case of an axially symmetric alignment tensor (i.e. negligible rhombicity), it is possible to select a specific mode allowing rotation of the molecule about the unique axis A_{zz}, as all of these positions are equivalent in this case.

In the case of a covalently bonded multimer, the program highlights the bonded partners at the junction between the selected modules and indicates the distance between the bonded atoms, so that the user can gauge the most likely relative positioning of the different domains.
Automatic domain positioning is performed by the program to provide an initial model, by minimizing the function

$$E_{\text{cov}} = \sum_{n} \left( d_{ij} - d_{ij}^{\text{cov}} \right)^2$$

with respect to the relative positions of the different oriented modules. $d_{ij}$ are the distances between the covalently bound atoms at each module junction. The positions can also be manually adapted to find a more intelligent solution.

Here is an example which illustrates the main features of the program -

**Modelling the tertiary structure of the Hammerhead Ribozyme from RDC.**

It has recently been demonstrated that residual dipolar couplings can contribute important information to the determination of RNA global fold determination (Mollova et al 2000) precisely because of the complementarity of this long-range structural order, with the local secondary structure which can often be identified from well-established experimental procedures.

Similarly, in this example we have simulated dipolar couplings measured in both sugars and bases (assuming a $^{13}$C labelled sample to be available) from the hammerhead ribozyme, by calculating C-H couplings from the crystallographic structure (pdb code 1mmh) and adding 8% stochastic noise to the simulated values. The alignment tensor was assumed to be

$$\frac{\gamma_c \gamma_h \mu_H \rho_h}{16\pi\hbar^3 T_{ij}} A_a = 12.2 \text{Hz};$$
$$\frac{\gamma_c \gamma_h \mu_H \rho_h}{16\pi\hbar^3 T_{ij}} A_t = 1.5 \text{Hz};$$

and $S$ assumed to be equal to 1 throughout the molecule. The molecule was then "unwound" using the Discover-derived program SCULPTOR (Hus et al 2000) using a high temperature restrained molecular dynamics calculation, such that the orientation of the helices was no longer native, but the secondary structural regions remained intact.
Comparison of the X-ray crystallographic structure of RNA/DNA ribozyme inhibitor and the structure used as initial model for the simulated experiment using Module (right). The native structure was partially unfolded using high-temperature restrained molecular dynamics as described in the text. The three stem regions are shown in blue (I), orange (II) and red (III), while the core is shown in grey. The heavy atoms from the core region were used for the superposition of the two structures. The rmsd of the heavy atoms between the two models is 10.5Å.

The different regions of the molecule to be treated as individual domains are selected from the primary sequence. The three stem regions are shown in blue (I), orange (II) and red (III), while the core is shown in yellow. This non-native structure (heavy atom rmsd of 10.5Å compared to the initial, correct structure) and the simulated couplings were then used to reconstruct a model of the molecule using Module. Comparison of noise — simulated and fitted data from the 3 stem regions of the ribozyme — The blue data correspond to points from stem I, orange from stem II and red from stem III.
I - The alignment tensors of the different modules are determined and their eigenvectors superposed on the structures in their original (unwound) orientation

II - The modules are then oriented so that the tensors all have the same alignment in the frame indicated by the tensor directions. The dotted lines indicate the distances between the covalently bound atoms. The substructures can then be manipulated individually on the screen, using only translational degrees of freedom and 180°; rotations about $A_{xx}$, $A_{yy}$ and $A_{zz}$, to find the most feasible model.

III - The optimal position of the different modules can also be calculated automatically, as described in the text, and this, or the manually adjusted orientation, can then be fixed and written in standard coordinate format.

The final structure was calculated automatically using MODULE

The final model has a backbone rmsd of 2.5 Å compared to the initial crystal structure.