NMR spectroscopy is now established as the most effective method for the determination of three-dimensional structure of biomolecules in the solution state (Wüthrich 1986). The success of this technique relies primarily on the measurement of cross-relaxation effects (nOe) providing distance restraints between protons in close proximity in space (<6Å) supplemented by 3J scalar couplings which contribute local dihedral angle information (Clore & Gronenborn 1997). In combination with simple modeling of covalent and through-space interactions, these short-range constraints allow the accurate reconstruction of the secondary and tertiary molecular geometry of globular proteins.

The strength of nOe-based structure determination is also the cause of its principal limitation; the measurement of insufficient distances, for example at interfacial or hinge regions, can result in ill-defined relative positioning of distant regions of extended or modular macromolecules.
The accurate determination of long-range order using nOe is thus limited by the absence of correlated structural information from parts of the molecule which are far apart in both primary sequence and three-dimensional coordinate space. This can compromise our understanding of the structure, not simply due to low structural precision, but also because disorder due to insufficient constraints cannot easily be distinguished from real dynamic averaging. The difficulty is compounded by the fact that nOe becomes difficult to measure in large protonated molecules due to prohibitive relaxation effects, making the determination of structure beyond 30kD a challenging problem using classical techniques. These parameters can of course be improved by deuterium labelling, although the number of available inter-proton distances is then concomitantly reduced. Selective $^1$H labelling of specific sidechain methyl groups is then necessary to provide important long-range NOESY connectivity.

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 potential solution to these limitations:

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;

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

This averaging, denoted by the brackets $\langle \rangle$, 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.**

**Long-Range Orientational Constraints.**

Residual Dipolar Couplings

Incomplete averaging due to partial molecular orientation.

**Sources of Molecular Alignment**

(Tjandra & Bax 1997, Science 278,1111; Tolman et al;1995, P.N.A.S. 92, 9279.)
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 1997). 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:

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

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 ±y-axis, in the ±xz 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 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 below:
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:

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

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

\[
\gamma_C \gamma_{CH} h^3 \frac{1}{16 \pi \frac{1}{3} r_{ij}^3} A_a = 12.2 \text{ Hz};
\]

\[
\gamma_C \gamma_{CH} h^3 \frac{1}{16 \pi \frac{1}{3} r_{ij}^3} A_t = 1.5 \text{ Hz};
\]
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 $180^\circ$; 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.
**Orientational degeneracy in the presence of one alignment tensor.**
Rotation about the x, y and z axes can be performed in the aligned mode of the program to examine the four degenerate orientations of two modules as shown below -

The $180^\circ$ rotations about the axes of the alignment tensor generate multiple solutions for the relative orientations of the two modules. The four different orientations of module 1 relative to module 2 are shown above. The dotted line indicates the covalent junction between the modules. This covalent information may in certain cases be sufficient to select the most probable conformation.

The residual dipolar couplings considered here (between spins from covalently bound nuclei) contain purely orientational information; this means that there is complete translational freedom in all Cartesian (x,y,z) directions. Where possible either covalent, experimental or non-bonded interactions can of course be used to place different modules with respect to each other. The program Module will calculate, and localise the steric contacts due to atom-overlap between modules (left).
RDC for refining homology-based models.

A number of groups have recognised the obvious power of RDC's to either verify or refine structural models. It is something of a revolution for the solution state structural biologist to have access to coherent structural information from throughout the molecule, as soon as backbone assignment has been completed. The increasing number of experimentally determined structures available in databases have increased the possibility of structure modelling from primary sequence alone, on the basis of sequence similarity. RDC's measured in secondary structural elements (so that the local geometry can be assumed to be relatively good, and tensors can be determined from sufficient couplings) can be used to either verify, or refine, the predicted folds. This idea is illustrated here using the program Module, although the approach has been suggested by numerous authors (Meiler et al 2000, Chou et al 2000, Annila et al 1999) in recent years.

In this example the global fold is assumed to be known, but the relative orientation of secondary structural elements is not precisely determined (this may be the case for example when a homology-based model is available).

In the first case we have selected all of the secondary structural elements to be part of the same module. If we determine a single tensor for this selection, assuming therefore that the relative orientation is correct, the fit is already quite good, as shown from the correlation plots for the different coupling types. Nevertheless there are significant differences between calculated and experimental values.
We now divide the secondary structural elements into four separate modules - the three helices and the central beta-sheet;

If we fit these modules separately to independent tensors we get better correlations between experimental and calculated values.

If we look at the orientation of the tensors relative to the separate structural elements in the pdb frame we can see why this might be -

The alignment tensors of the different motifs are differently oriented with respect to the pdb frame. The axial and rhombic components have similar values, suggesting that the orientation of the secondary structural elements was not quite right in the initial model. In particular the C-terminal helix, (blue) seems to be tilted significantly relative to the others.
The program can now be used to orient the motifs into a common frame so that each of the alignment tensors is coaxial. All four modules are then aligned with the same tensor axes. In this case a large number of possible solutions exist for the different orientations of the modules, due to the 4-fold degeneracy for each motif. The program can test these different possibilities to find the geometry, which is in best agreement. The structure will then need to be refined in a molecular dynamics type calculation to ensure that non-bond and covalent considerations are respected.

**Structural Homology from Database Comparisons with RDC.**

An obvious development of this kind of analysis is the direct comparison of measured sets of RDC with expected couplings from structural motifs present in databases. It has indeed recently have shown that RDC can also be used to rapidly identify complete or partial homologous fold by comparison with structures of different characteristic length, present in the available databases (Annila et al 1999, Meiler et al 2000, Andrec et al 2001). Delaglio et al 2000, have also shown that in the case where extensive data are available from two alignment media, it is possible to identify 7-amino acid segments from the structural database which are in best agreement with a sliding window of experimental RDC's along the primary chain of the protein of interest, and superimpose the common regions to define the fold of the molecule (Molecular Fragment Replacement). The improved structural definition available from two different alignment media will be discussed below.
Structure Refinement using RDC in a hybrid energy function.

The definition of an experimental RDC residual term in a hybrid molecular dynamics energy function has also been popularly used in the refinement stage of structure determination protocols to drive an existing structure into a conformation in agreement with measured couplings (Clore & Gronenborn 1998, Garrett et al 1999, Tsui et al 1999, Hus et al 2000, Huang et al 2000, Clore et al 2000). Nevertheless the directional ambiguity of RDC’s, for example resulting in equivalent energetic minima at 180° rotations about each axis of the alignment tensor, makes the energetic landscape particularly obscure. Despite the obvious complementarity of RDC and $^1$H-$^1$H nOe it is therefore not yet clear that ab initio structure calculation protocols developed for predominantly short-range restraints are the most appropriate for use with orientational restraints. Meiler et al (2000) have recently overcome this problem by transforming measured dipolar couplings into angular restraints between vectors (these are correlated because related to the same tensor). These projection angle restraints can be numerous, because of the number of possible cross-correlations, but provide a much simpler energy landscape.

As noted above, another problem which must be addressed by any structure calculation protocol is the determination of tensor eigenvalues ($A_a$ and $A_r$) and axis orientations (defined by the Euler angles $\alpha$, $\beta$, $\gamma$) are unknown a priori and must somehow be calibrated during any proposed calculation strategy. The Grzesiek group have recently shown that the linearity of the alignment matrix allows the magnitude to be eliminated from the target function (Moltke & Grzesiek 1999).

A number of structure calculation packages now contain an RDC term; the details of each are no-doubt quite similar. I will again describe the program developed in our laboratory, simply because we are most familiar with this, rather than claiming because we feel it is any more efficient than other available program. We have written a suite of programs to incorporate the residual dipolar coupling constraint into the code of the program Fdiscover as an explicit target potential in addition to the classical potential energy function of the AMBER4 force field.

$$E_{RDC} = k_{RDC}(D_{ij}^{\text{calc}} - D_{ij}^{\text{exp}})^2 / \sigma_{ij}^2$$

In order to facilitate reorientation of complete structural motifs, and to penalize local violations of planarity, the force-field can have higher force constants than normally used to constrain peptide bond or aromatic ring planarity and reinforce local valence angle geometry.

This module is called SCULPTOR (Structure Calculation Using Long-range, Paramagnetic, Tensorial and Orientational Restraints). SCULPTOR was specifically written to allow maximal flexibility in the development of conformational search algorithms using long-range structural restraints: Tensor parameters are treated as independent pseudo-molecules, read with the coordinates of the molecule of interest. Eigenvalues and eigenvectors are treated separately, such that one two-point molecule represents $A_a$, one $A_r$ and a third, 3 point orthogonal axis system represents the orientation $\alpha_{\alpha,\beta,\gamma}$ of the tensor, so that each term could
be manipulated independently. The residual dipolar coupling term can be treated either using a standard distance term from a look-up table, or using the actual distance between the atoms during the calculation. Two independent alignment tensor functions are available in SCULPTOR; these can be used separately for restraints from different alignment media, or to refine two regions independently to test the similarity of tensors for the different regions.

The advantage of using residual dipolar couplings as restraints in a hybrid molecular dynamics energy function is that the final model is already in agreement with covalent and non-bonded interactions. An additional advantage lies in the flexibility of the approach - local structures can for example be refined to fit measured RDC, thus increasing the accuracy of both local conformation and alignment parameters, before determining the long-range order of the molecule.

This is illustrated in our recent study of the theophylline binding RNA, whose classical NMR structure is shown below, calculated using nOe and 3J couplings (Zimmerman et al 1997). The three ensembles represent superpositions on the two helical regions of the molecule (middle and right), and whole sequence (left). A restrained molecular dynamics calculation was performed using two independent tensors for these two regions of the molecule. During this calculation, which included nOe, J-couplings and RDC, the tensors were allowed to float freely for the two regions simultaneously.
The evolution of rhombic and axial components during a typical calculation are shown below. Over the ensemble of initial structures both tensors converged to very similar eigenvalues. The local helical, and helical/tetraloop structure is refined in the two regions, but their relative orientation is not yet determined, as the refinement was carried out using independent tensors. The similarity of the tensor eigenvalues implies that significant differential domain motion is not taking place, although this cannot be ruled out. The rest of the analysis nevertheless assumes that this is not the case.
We then proceeded to determine the overall structure, which would be agreement with one single alignment, by applying a high temperature, restrained molecular dynamics calculation in the presence of one tensor. The local structures of the helical regions were constrained to their refined structures using artificial distance constraints, to facilitate the reorientation of the separate domains. The relative orientation of the different regions can be seen to sample a broad region of conformational space, before annealing to the final topology, in agreement with nOe and RDC. The snapshot structures are superimposed on the helix/tetraloop region (lower part of the molecule in this figure).

The final structure is of course more ordered than the initial nOe-based ensemble:
Residual Dipolar Couplings Measured in Two Alignment Media.

We have already seen that the orientational degeneracy inherent to a single measured coupling can be raised by measuring different directions in a rigid body. We will now turn our attention to the further gain in orientational definition achieved by measuring in the presence of a differently aligned tensor. This will of course require that we have access to liquid crystal media whose interactions with the solute molecules are different. If these are the same, then the alignment tensors will also be the same. A number of groups have observed significantly different alignment characteristics between, for example bacteriophage and bicelles, probably due to differences in the electrostatic and steric contributions, or by changing the pH or ionic strength. The effect of the orientational definition for a single vector is sketched here: The continuum of solutions in the presence of a single tensor is mapped as distorted cones of orientational isocontours on the surface of the sphere. If we add data from the same coupling in the presence of a second, differently aligned, tensor the distorted cones intersect to provide a maximum of eight equivalent solutions (four in this example) (Ramirez & Bax 1998).

If we now look at the previously considered case, where a three-dimensional, or chiral, motif was found to have four equivalent orientations in the presence of a single alignment medium; Four equivalent orientations exist for the sub-structure in the presence of the two different alignment tensors, related in both cases to 180° rotations about the respective axes of the two
different alignment tensors, but only one of the four solutions in each case (the correct orientation) is common (Al-Hashimi et al 2000). This is very important information and implies that structure calculation using only residual dipolar couplings can be achieved assuming that the following criteria can be satisfied -

1) The molecule of interest can be divided into sub-structures whose local geometry is accurately known.
2) Sufficient RDCs can be measured from these sub-structures to determine their relative orientation.
3) These couplings can be measured in two significantly differently alignment media.

With the development of more and more powerful techniques for the measurement of $^1J$ and $^2J$ RDC's along the peptide chain, it is evident that, under favourable conditions, these criteria can be fulfilled for the study of proteins.

In the final part of this chapter, I will therefore describe an approach we have developed to determine the structure of the protein backbone using extensive sets of heteronuclear residual dipolar couplings measured from throughout the peptide plane and the tetrahedral junctions between planes. The approach has been named MECCANO (Molecular Engineering Calculations using Coherent Association of Non-averaged Orientations), and is applied to the dataset measured from ubiquitin in bicelles and charged bicelles by the Bax group at NIH (Cornilescu et al 1998, Ottiger & Bax 1998). These are the same data, which were used in the Molecular Fragment Replacement approach described earlier.

The protein backbone can be described as a chain of planar motifs, connected by tetrahedral junctions. We will therefore consider the peptide plane as a sub-unit of known structure. The RDC's measured in this sub-unit are shown here. Two sets of 63 N-H⁸, 61 C'-H⁸, 61 and 63 C'-N and 59 and 54 C⁸-C' couplings defining the peptide plane orientations, in addition to two sets of 62 C⁸-H⁰ and one set of 39 C⁸-C⁰ couplings were used in the calculation.

Meccano is essentially a two-step algorithm. As mentioned earlier, in an *ab initio* calculation the alignment tensors will not be known *a priori*. The first step is therefore designed to calibrate both tensors, in the absence of any structural information concerning the fold of the molecule.

We have developed a least-squares-based search algorithm to determine the alignment tensors, described by 7 parameters in the calculation frame $(A_1^1, A_1^2, A_2^2, A_2^3, \alpha, \beta, \gamma)$, where $\alpha, \beta, \gamma$ describe the orientation of $A_2^2$ with respect to $A_1^1$, taken to be diagonal in the calculation frame.
Simultaneously, the orientation of each peptide plane \((\theta_1, \theta_2, \theta_3)\) is determined with respect to the calculation frame. The number of parameters to be determined is then \((3N+7)\) describing each plane orientation and the common reference frame, while the number of RDC’s available is \(8N\), assuming we only include peptide planes with complete datasets in this stage of the calculation. This algorithm reliably finds the global minimum of the target function:

\[
\chi^2 = \sum \left( \frac{(D^\text{exp} - D^\text{calc})}{\sigma} \right)^2
\]

over all measured couplings, requiring no \textit{a priori} estimation of the alignment tensors. The second stage of the algorithm then constructs the molecule in this reference frame.

It is important to note here that the planarity of the peptide plane reintroduces an orientational ambiguity between the mirror image and the correct alignment. It is possible to construct the folded peptide chain from known orientation of individual peptide planes (plane \(i\) is defined here as \(C'_{i-1}, C_{i-1}, N_i, C_i\)). The correct plane orientation can be distinguished from the mirror image by tetrahedral geometry requirements at the junctions connecting peptide planes, although for the general experimental case this is not always sufficient.

The combination of RDC measured in the peptide plane and tetrahedral junctions (\(C_{i-1} - H_{i-1}, C_{i-1} - C_{i-1}\)) however allows unambiguous positioning. When no peptide plane orientation is available for plane \((i+1)\), \(\phi/\psi\) values are optimized to reproduce \((C^a - H^a, C^a - C^a)\) from \((i)\) and \((i+1)\) and peptide plane data from \((i+2)\). This allows the unambiguous continuation of the peptide chain, albeit less precisely than for the more complete peptide plane data-set.

Two sequential proline residues are present in the ubiquitin sequence\((37,38)\), for which no orientational plane RDC information is available. In this case a four parameter minimization is performed to optimize the target function defining all relevant (shown in red here) measured vector orientations with respect to the dipeptide angles \(\phi_{36}/\psi_{56}, \phi_{37}/\psi_{37}\) using a robust multi-parametric minimization algorithm. The constructed (red) and refined (orange) structures in this region are shown below, in comparison to the nOe-based structure (blue).
The final structure (1.3Å backbone rmsd from the structure determined using 2727 nOe, in combination with RDC, $^3$J couplings and hydrogen bonding restraints) is therefore defined as a combination of oriented peptide planes and alpha-carbon junctions. The ability to determine protein structure *ab initio* simply on the basis of residual dipolar couplings measured along the protein backbone combined with rudimentary covalent considerations concerning local peptide plane conformation is clearly an exciting development for the application of this kind of restraint. The dataset used is nevertheless extensive, and apparently highly accurate, raising the question as to whether this kind of analysis can be generalised to larger, and more difficult systems. We should however remember that a great deal of free information was not included in this calculation, for example non-bonding terms were ignored, and secondary structural information, available from chemical shift analysis, was not taken into account. The future exploitation of RDC's for *ab initio* calculation of protein structure will no doubt combine all of the available structural information to produce molecular models.

The approach described here is conceptually similar to work carried on aligned molecules in bicelles (Cross & Opella 1994, Ketchem et al 1996) and recently described work in partially aligned proteins (Prestegard et al 2000).

An example of how this method has been applied to determine the backbone conformation of the active site of a somewhat larger molecule is now given. In this system, the Methionine Sulfoxide Reductase from *Erwinia chrysanthemi*, a crystallographic structure of a highly homologous molecule from *E.coli* was found, using RDCs and the program MODULE, to have very similar structure throughout the molecule except for one highly localised region. Using backbone RDCs and the approach presented above the local conformation was determined ab initio and reinserted into the common scaffold of the protein.


Huang, K; Louis, J.M.; Donaldson, L.; Lim, F-L.; Sharrocks, A.; Clore, G.M. EMBO.J. 2000, 19, 2615-2628


Tjandra, N. (1999), Structure, 7, R205-R211.
Direct Structure Determination Using Residual Dipolar Couplings: Reaction-Site Conformation of Methionine Sulfoxide Reductase in Solution

Sabine Béraud,† Beate Bersch,‡ Bernhard Brutscher,‡ Pierre Gans,† Frédéric Barras,‡ and Martin Blackledge*†

Contribution from the Institut de Biologie Structurale, Jean-Pierre Ebel, Centre National de la Recherche Scientifique–Commissariat à l’Energie Atomique UJF, 41 rue Jules Horowitz, 38027 Grenoble Cedex, France, and the Laboratoire de Chimie Bactérienne, Centre National de la Recherche Scientifique, Institut Biologie Structurale et Microbiologie, 31 Chemin Joseph Aiguier, 13402, Marseille, France

Received May 13, 2002

Abstract: Residual dipolar couplings (RDC) from partially aligned molecules provide long-range structural data and are thus particularly well adapted to rapid structure validation or protein fold recognition. Extensive measurements in two alignment media can also provide precise de novo structure from RDC alone. We have applied a novel combination of these approaches to the study of methionine sulfoxide reductase (MsrA) from Erwinia chrysanthemi, a 27 kDa enzyme essential for repairing oxidative stress damage. The tertiary fold was initially validated by comparing backbone RDC to expected values from the crystal structure of the homologous MsrA from Escherichia coli. Good agreement was found throughout the chain, verifying the overall topology of the molecule, with the exception of the catalytically important peptide P196–L202, where strong and systematic RDC violation was observed. No evidence for local differential mobility in this region was detected, implying that the structure of the strand differs in the two molecules. We have therefore applied the de novo approach meccano to determine the conformation of this peptide using only RDC. A single conformation is found that is in agreement with all measured data. The aligned peptide can be docked onto the expected covalence of the rest of the template molecule while respecting its strictly defined relative orientation. In contrast to the structure of MsrA from E. coli, the reactive side chain of Cys200 is oriented toward the interior of the molecule and therefore closer to the catalytic Cys53, obviating the need for previously proposed conformational reorganization prior to formation of this disulfide intermediate. This analysis requires only backbone assignment and uses unambiguously assigned and readily measurable structural data, thereby greatly economizing investigation time compared to established nuclear Overhauser effect- (nOe-) based structure calculation methods.

Classical NMR structure determination requires extensive assignment of backbone and side-chain resonances followed by unambiguous identification of nuclear Overhauser effect (nOe) correlations between these assigned frequencies, before even low-resolution structural models can be derived.1 Complete structural analysis can thus be particularly time-consuming for large proteins. In contrast, residual dipolar couplings (RDC) in macromolecules aligned in dilute liquid crystalline media2–4 can be measured routinely from pairs of nuclei distributed throughout the molecule immediately following assignment of the backbone resonances. RDC provide coherent, long-range structural data, allowing entirely new approaches to structural biology in solution.5–8 This coherence makes RDC particularly appropriate for comparison with available structural homologues, enabling rapid validation of initial molecular models or identification of folds from database structures9–11 and subsequent refinement of these initial models.12 Although this approach requires the presence of a structural homologue in accessible databases, the identification of even a low-resolution molecular fold from primary data is a major breakthrough for solution-state NMR.

10.1021/ja0268783 CCC: $22.00 © 2002 American Chemical Society
In cases where sufficient measurements can be made from throughout the peptide chain in the presence of two alignment media, protein structure can be determined from RDC alone, either in comparison with short fragments from structural databases or by sequentially building the peptide chain from unambiguously oriented peptide units. The requirement of continuous data from peptide planes and/or tetrahedral junctions has so far limited the number of applications of these de novo methods to particularly well-behaved systems. In the absence of continuous RDC data, locally well-defined substructures may be isolated in the primary sequence and cannot easily be placed with respect to the rest of the molecule unless additional structural information, for example long-range N\(^2\)Oes, is also available.

Methionine sulfoxide reductase (MsrA) catalyzes the reduction of free and protein-bound oxidized methionine residues and is consequently required for the repair of important enzymes. The enzyme, found in nearly all living organisms, is present at high concentration in tissues and cells susceptible to oxidative stress damage and has thus been linked to aging processes, lifespan determination, and diverse degenerative pathologies such as Alzheimer’s disease. Two crystal structures of MsrA have recently been elucidated: MsrA from Esherichia coli and Bos taurus (MsrA\(^{Ecoli}\)). Although very similar, these structures differ significantly in the C-terminal strand involved in freeing the catalytic site for further function. To further understand the mechanism by which MsrA provides protection against oxidative stress, we have studied the solution structure of reduced MsrA from Erwinia chrysanthemi (MsrA\(^{Echmi}\), 221 amino acids), a plant pathogen that requires such enzyme for full virulence, using RDC measured in partially aligned liquid crystal media.

To characterize the catalytically important C-terminal loop domain of MsrA\(^{Echmi}\), we have developed a novel combination of the two RDC-based approaches presented above: The tertiary fold is initially validated by RDC, providing low-resolution structural information and positioning the molecule in the common alignment frames. We then focus on the region of interest, which is determined from RDC alone by the meccano approach. This region can be unambiguously positioned relative to the remaining molecular scaffold by use of the graphic modeling tool Module, a program specifically developed for RDC analysis to facilitate the manipulation of oriented molecular domains relative to a common alignment frame.

This is a highly efficient form of conformational analysis via NMR: unambiguously assigned structural data, requiring only backbone resonance assignment and derived from comparatively simple experimental techniques, provide precise structural detail in regions of specific interest in the molecule. In the example shown here, the characterization of the reaction site of MsrA provides potentially important insights into the molecular mechanism of the catalytic cycle of the enzyme.

### Methods

Samples of \(^{15}\)N-labeled and \(^{13}\)C,\(^{15}\)N,\(^{2}\)D-labeled MsrA\(^{Echmi}\) were prepared as previously described. Both samples were studied under reducing conditions with 10 mM dithiothreitol.

#### Residual Dipolar Coupling Measurement

All NMR experiments were performed on a Varian Inova 600 spectrometer, equipped with a triple-resonance (\(^1\)H, \(^{15}\)N, \(^{13}\)C) probe and shielded z-gradients. Sample temperature was set to 298 K for all studies. RDC were collected on a 1.0 mM uniformly \(^{13}\)C, \(^{15}\)N, and 82% \(^{2}\)H-labeled sample suspended in a liquid crystalline medium consisting of 20 mg/mL of the filamentous phage P1 (Asla Ltd., Riga, Latvia) in 80 mM potassium phosphate buffer, pH 7.5. RDC were also collected on an identical sample suspended in a liquid crystalline medium consisting of 5% C12E6/hexanol in 25 mM potassium phosphate buffer, pH 7.5. Four different types of dipolar couplings were measured as previously described and described in detail elsewhere. Data were analyzed by the model-free Lipari–Szabo approach with the program TENSOR2.

#### Relaxation Experiments and Data Analysis

The \(^{15}\)N R\(_1\) and R\(_2\) relaxation and \(^1\)H–\(^{15}\)N N\(^2\)Oe measurements were performed on the \(^{15}\)N-labeled sample (1.4 mM concentration in 25 mM potassium phosphate buffer) at \(^1\)H frequency of 600 MHz and temperature of 298 K with the classical \(^1\)H-detected pulse sequence based on established methods and described in detail elsewhere. Data were analyzed by the model-free Lipari–Szabo approach with the program TENSOR2.

#### Fold Validation

The crystal structure of MsrA\(^{Echmi}\) (molecule A in PDB file 1FF3) was used to analyze and validate the solution structure of MsrA\(^{Echmi}\). Secondary structure was identified from experimental \(^13\)C chemical shifts from random-coil values (chemical shift index, CSI), and the intensity distribution of short- and medium range \(^{15}\)N\(^2\)Oes involving \(^1\)H\(^2\) protons.

The program Module was used to evaluate the accordance between the MsrA\(^{Echmi}\) structure and the experimental RDC data. Residual dipolar couplings can be expressed in terms of the orientation \((\theta, \phi)\) of the internuclear vector relative to a common alignment tensor for the molecule:

\[
D_\theta = \frac{2\mu_0 h}{16\pi r^3} (A_0 (3 \cos^2 \theta - 1) + \frac{3}{2} A_4 \sin^2 \theta \cos 2\phi)
\]

where \(A_0\) and \(A_4\) are the axial and rhombic components of the alignment tensor, \(r_0\) is the internuclear distance, and \(S\) is the order parameter.
The alignment tensor is characterized by five parameters: the axial and rhombic components, $A_a$ and $A_r$, measure the extent of residual alignment due to the restricted orientational sampling in the anisotropic medium, whereas the Euler angles, $(\alpha, \beta, \gamma)$, define the nonaveraged orientation of the molecule relative to an external reference frame. For each entity, these five parameters were determined by nonlinear least-squares minimization of the target function over all couplings associated with a given domain:

$$\chi^2 = \sum_n (D_{ij} - D_{ij}^{\text{calc}})^2 / \sigma_{ij}^2$$

(2)

where $D_{ij}$ are the residual dipolar coupling between spins $i$ and $j$ and $\sigma_{ij}$ is the uncertainty in the experimentally measured coupling. The average estimated $\sigma_{ij}$ is on the order of 2 Hz. Alignment tensor parameters were determined and visualized relative to the three-dimensional atomic coordinates.

The quality of the fit to experimental RDC was inspected qualitatively with correlation plots of the measured and calculated couplings from the best-fit alignment tensor and quantitatively with the total $\chi^2$ target function. Amplitude and orientation of the individual alignment tensors for the different structural elements were compared for evidence of differential flexibility. In a rigid molecule, $A_a$ and $A_r$ should be identical for each structural entity, and the individual alignment tensors should be coaxial if the relative orientation of the secondary structure elements in the model structure is correct.

De Novo Structure Calculation. The recently published meccano approach14 was slightly modified. The molecular dynamics (MD) program Sculptor15 was incorporated into the sequential positioning algorithm designed to place the peptide units. This provides randomized sampling of conformational space, which facilitates inspection of the conformational precision in the final ensemble. The use of a molecular dynamics force field also allows the straightforward introduction of repulsive nonbonding interactions. In this case a specific force field dynamics force field also allows the straightforward introduction of conformational precision in the final ensemble. The use of a molecular sampling of conformational space, which facilitates inspection of the algorithm designed to place the peptide units. This provides randomized positioning of different modules using only the equivalent orientations and three-dimensional atomic coordinates.

The experimental RDC for each peptide unit [from C\text{\textsuperscript{\textalpha}} junction ($i$) to C\text{\textsuperscript{\textgamma}} junction ($i + 1$)] were successively applied to place the unit in the calculation frame. $A_a$, $A_r$, $A'$, and $A''$ and their relative orientations $(\alpha, \beta, \gamma)$ are known from the analysis of the core region of the molecule and are fixed throughout the calculation. A short RDC-restrained molecular dynamics calculation consisting of 1000 0.1-fs heating steps (to a nominal 500 K) and 1000 0.1-fs cooling steps was applied for each peptide unit. One structure calculation takes 15 s on a Linux 1 GHz PC. Two thousand conformers were calculated by sequential positioning of peptide planes and C\text{\textsuperscript{\textgamma}} junctions. The final ensemble was chosen on the basis of the total residual in the RDC target function $\chi^2$ (eq 2). Covalent strain in the peptide planes or tetrahedral junctions in the final peptide structures was measured from the residual terms compared to a standard molecular force field (AMBER4).35

Docking the Meccano Peptide to the Protein Scaffold. The MsrA\text{Ecol} P196-L202 peptide was positioned relative to the crystal coordinates of MsrA\text{Ecol} with the molecular modeling tools available in the program Module. This program allows positioning of different oriented modules using only the equivalent orientations and three-dimensional translational freedom available in a common alignment frame. The crystal structure and peptide were aligned relative to their crystal structure conformation by use of the additional force-field potential:

$$F_{\text{eth}} = k_{\text{eth}} \sum_i \sqrt{(x_i - x_i^{\text{mol}})^2 / N}$$

(3)

with a $k_{\text{eth}}$ value of 20 kcal mol\textsuperscript{-1} Å\textsuperscript{2}. All measured RDCs were used as standard restraints in the presence of the two independent tensors.

Results and Discussion

We have assigned the backbone resonances36 and measured extensive dipolar couplings (N–H, C–H, C\text{\textsuperscript{\textalpha}}–C, and C\text{\textalpha}–C\text{\textsuperscript{\textgamma}}) from the reduced form of $^{13}$C, $^{15}$N, and $^2$H-labeled MsrA\text{Echmi} in two liquid crystalline solutions. These couplings were initially compared to expected values from the recently published crystal structure of the 75% identical primary sequence homologue MsrA\text{Ecol} by use of the alignment tensor optimization and molecular modeling program Module. The reduced form of MsrA\text{Ecol} consists of an $\alpha/\beta$ roll core structure, comprising 80/123 residues in secondary structural elements, and predominantly coil N- and C-terminal regions wrapped around this core. The 45 amino acid C-terminal region is of particular interest as two of the three cysteines, Cys200 and Cys208, present in the proposed catalytic cycle36,37 are found in this strand (MsrA\text{Echmi} numbering is used unless otherwise stated). This cycle involves initial nucleophilic attack of Cys53 on Met-SO, followed by a disulfide cascade implicating Cys53–Cys200 and then Cys200–Cys208 disulfide bridges, the latter step freeing the catalytic sulfur (Cys53) for further function.

Validation of the Backbone Fold in Solution. Initial comparison, concentrating on the secondary structural elements of the molecule, reveals that the central core is folded very similarly in MsrA\text{Echmi} in solution and in the MsrA\text{Ecol} crystal structure (Figure 1 and Table 1). Alignment tensor axial and rhombic components from nonmobile sites in the molecule are reproduced similarly in the different structural motifs, and the relative orientations of the alignment tensor axes in the combined helical and combined $\beta$-sheet regions are in agreement with the MsrA\text{Ecol} crystal structure. This analysis was performed for RDC measured in both alignment media, further validating the expected fold of the core of the molecule. The best-fitting structural element is the combined helical region, which has a total $\chi^2$ of 279 for 294 couplings in the two media ($\chi^2 / N = 0.95$). Alignment parameters were then determined for the C-terminal loop alone and in combination with the core structure. This again shows satisfactory agreement with experi-

Comparison of measured RDC from MsrA\textsubscript{E.coli} with expected values for the crystal structure coordinates from MsrA\textsubscript{E.coli}. (a) Related orientation of the alignment tensors for the different secondary structure elements in MsrA\textsubscript{E.coli} for the RDC from the C12E6/hexanol aligned sample. In this figure the alignment tensors of the central \(\beta\)-sheet (46–50, 68–75, 95–103, 139–141, 174–176, and 182–184), and \(\alpha\)-helices (55–62, 85–89, 110–119, and 145–164) are compared relative to the crystal structure. The axial and rhombic components and \(\varphi\) of the five secondary structure elements are given in Table 1. The different secondary structure elements were also fitted as a single domain using all couplings simultaneously as shown in panels b–d. (b, top left) Correlation between \(\Delta_{\text{DCCS}}\) and \(\Delta_{\text{HC}}\) for all core secondary structural elements shown in panel a. (c, bottom left) Correlation between \(\Delta_{\text{DCCS}}\) and \(\Delta_{\text{HC}}\) for all core secondary structural elements. (d, top right) Correlation between \(\Delta_{\text{DCCS}}\) and \(\Delta_{\text{HC}}\) for all core secondary structural elements.

The observed systematic disagreement could be explained by differential dynamics in this region, producing time-averaged RDC values that are in disagreement with a single conformational model, or by a different local conformation in MsrA\textsubscript{E.coli} compared to the MsrA\textsubscript{E.coli} crystal structure. \(^{15}\)N relaxation data measured at 600 MHz \(^1\)H frequency present no evidence for large amplitude motion on the rapid (pico- to nanosecond) or intermediate (micro- to millisecond) dynamic time scale in this region. By use of the Lipari–Szabo approach, the lowest order parameter in this region is 0.75 and the average is 0.84 (Figure S1, Supporting Information). None of the residues in the strand segment from Pro196 to Leu202, where strong disagreement is found (Figure 2), this local inconsistency is systematically found for each individual coupling type, in both alignment media. Removal of these data from the fit reproduces very similar alignment tensor values for the coil region alone (Table 1) as were found for the core region and verifies the relative orientation of these regions in the solution state.
De Novo Structure Determination with *meccano*. This analysis was performed via the *meccano* approach, previously demonstrated for the de novo determination of the backbone conformation of the protein ubiquitin. In the case of MsrA *Echmi*, the remainder of the protein structure is assumed to be identical to MsrA *Ecoli*, as suggested from the RDC analysis in the core region. Only the structure of the P196–L202 peptide is determined de novo. The *meccano* method requires that the eigenvalues and relative orientation of the alignment tensors are known, or can be determined, and that sufficient RDC are available from two differently orienting media to sequentially build the backbone conformation in the molecular calculation frame. In the case of MsrA *Echmi* the two tensors are determined from the analysis of the secondary structural core of the molecule, as described above, in comparison with the coordinates of MsrA *Ecoli*. Despite the incompleteness of the RDC dataset over the whole sequence (only 90% of all possible RDC could be measured), nearly all potential RDC (47 from 48 potential couplings) were available from the peptide region, comprising six peptide planes and seven Cα junctions (Table S2, Supporting Information).

Conformational sampling of the algorithm is illustrated in Figure 3, where the root-mean-square difference of the structural ensemble compared to the lowest χ² structure is plotted with respect to each individual χ². Only conformers that do not violate any expected covalent angle by more than 10° are shown in the figure. Structures whose χ² falls below 62 form the ensemble shown in Figure 4.

**Figure 3.** Conformational sampling of the structure calculation algorithm. The root-mean-square difference of the structures in the ensemble compared to the lowest χ² structure (χ² = 40) is plotted with respect to each individual χ². Only conformers that do not violate any expected covalent angle by more than 10° are shown in the figure. Structures whose χ² falls below 62 form the ensemble shown in Figure 4.

**De Novo Structure Determination with *meccano***. This analysis was performed via the *meccano* approach, previously demonstrated for the de novo determination of the backbone conformation of the protein ubiquitin. In the case of MsrA *Echmi*, the remainder of the protein structure is assumed to be identical to MsrA *Ecoli*, as suggested from the RDC analysis in the core region. Only the structure of the P196–L202 peptide is determined de novo. The *meccano* method requires that the eigenvalues and relative orientation of the alignment tensors are known, or can be determined, and that sufficient RDC are available from two differently orienting media to sequentially build the backbone conformation in the molecular calculation frame. In the case of MsrA *Echmi* the two tensors are determined from the analysis of the secondary structural core of the molecule, as described above, in comparison with the coordinates of MsrA *Ecoli*. Despite the incompleteness of the RDC dataset over the whole sequence (only 90% of all possible RDC could be measured), nearly all potential RDC (47 from 48 potential couplings) were available from the peptide region, comprising six peptide planes and seven Cα junctions (Table S2, Supporting Information).

Conformational sampling of the algorithm is illustrated in Figure 3, where the root-mean-square difference of the structural ensemble compared to the lowest χ² structure is plotted with respect to each χ² (for simplicity only those structures with valid covalent angles are included in the figure). The 20 lowest target function (χ²) conformers, selected on the basis of the total χ² only, are shown in Figure 4 (average pairwise rms difference in this alignment frame is (1.2 ± 0.3) Å over all atoms). The highest target function conformer in the final ensemble has χ² = 62, and the best-fitting structure has a χ² of 40, close both to the expected experimental error and to the average from the core region (χ²/N = 0.85 compared to χ²/N < 0.95 for the helical regions of the crystal structure). While it is difficult to quantify the degree of confidence in the proposed model, the data appear to define a unique conformation as all structures with χ² < 70 have the same fold as the minimum χ² conformer (all-atom rms difference of the aligned conformer is less than 1.9 Å). Nevertheless over the whole 2000-conformer ensemble, χ² ranges to a maximum of 640, and conformational space is sampled very broadly (pairwise backbone rmsd of aligned conformers 10.5 ± 4.2 Å), illustrating that algorithm efficiency is low.

The experimental and calculated values from the lowest target function model are shown in Figure 5. The covalent distortion present in the molecule due to possible overrestrained vector orientations appears to be minimal in the final ensemble as illustrated from the distances and dihedral and covalent angles.
shown in Table 2. Total RDC violation ($\chi^2$) and molecular strain (as measured by the residual energy in the covalent terms of the molecular force field) are correlated over all structures ($r = 0.89$, data not shown), illustrating that the experimental data are coherent with conformers of expected peptide chain geometry and in disagreement with conformers of incorrect geometry. All dihedral angles are in most-favored or additionally-allowed regions of the Ramachandran plot.

**Table 2.** Interatomic Distances and Covalent and Planar Dihedral Angles in the Lowest Energy Meccano Conformer

<table>
<thead>
<tr>
<th>atoms</th>
<th>distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N–H$^b$</td>
<td>1.02 ± 0.01</td>
</tr>
<tr>
<td>C$<em>{\alpha}$–C$</em>{\alpha}$</td>
<td>1.53 ± 0.01</td>
</tr>
<tr>
<td>C–C$_{\alpha}$</td>
<td>1.52 ± 0.02</td>
</tr>
<tr>
<td>C$_{\alpha}$–N</td>
<td>1.450 ± 0.003</td>
</tr>
<tr>
<td>C–N</td>
<td>1.335 ± 0.002</td>
</tr>
</tbody>
</table>

*No deviations of angles greater than 2.6° from the expected covalence were found.*

Figure 6. Representation of the positioning of the P196–L202 meccano-peptide relative to the crystal structure of Msra$^{Ecoli}$ by use of the program Module to place the fragment with respect to the alignment tensor in phage. Only transverse degrees of freedom are available in the common coordinate system. The peptide can be easily accommodated at the C-K195 and N-G203 positions without significantly violating known covalence.

couplings measured in a single alignment medium. It may be instructive to underline the differences between our analysis and this kind of study: In contrast to restrained structure refinement, which is subject to stringent constraints, for example on the terminal positions of the peptide of interest, our approach is de novo, so that the conformation of the peptide is determined only by RDC and local covalence. Once the local structure has been determined, the ability to replace the oriented peptide in the molecular scaffold then provides an independent measure of the probability that the conformation is realistic. This de novo approach is of course only made possible because the available data is sufficient to unambiguously define the molecular conformation (eight RDC per peptide unit). This is not true if RDC from a single alignment medium are available, where multiple conformations can exist for the same measured data. In such cases molecular dynamics-based refinement of an available structure becomes the most appropriate method. The different conformational sampling characteristics of the two methods are evident.

**Comparison with the Available Crystal Structures.** The differences between the meccano conformation of Msra$^{Ecnu}$ and the X-ray crystal Msra$^{Ecoli}$ structure in this region provide potentially important information concerning the proposed functional cycle of Msra. While the Leu202 C$^{\alpha}$–C$^{\beta}$ vector direction is the same in both conformers, implying that the hydrophobic interactions for this side chain are conserved, the peptide chain around Cys200 has a very different conformation. Most importantly, the C$^{\alpha}$–C$^{\beta}$ vector is pointing away from the core in the X-ray structure and points into the core in the solution model (Figure 7, top). The Cys200 and Cys53 C$^{\gamma}$ atoms are therefore 3.2 Å closer together in the chimeric model ($dc_{\gamma} = 7.9$ Å) of Msra$^{Ecnu}$, and the disulfide-forming sulfurs are concomitantly closer, apparently obviating the need for previously suggested large-scale conformational rearrangement before the formation of this bond. Note that no information concerning the exact position of the S$^{\gamma}$ is available, as the $\chi^4$ angle is undefined in our study.

It may be relevant that the crystal lattice contains three monomers, of which only one monomer (A) contains coordi-
Echmi is also close to that predicted from the MsrA conformation. Interestingly, the more distantly related MsrA reveals that the Cys200 S of the common main chain of monomer A on monomer B nates for the C-terminal region beyond Lys194. Superposition model, compared to {De no...}

**Figure 7.** De novo RDC derived and crystal structures of the P196–L202 peptide. (a, top) Comparison of the native MsaE coded model (yellow) after positioning of the meccano peptide onto the MsaE crystal structure. The backbone dihedral angles are {\( \phi, \psi = (-129^\circ, 1^\circ)_{V197}, (-43^\circ, 163^\circ)_{G198}, (-72^\circ, 170^\circ)_{Y199}, (-150^\circ, 98^\circ)_{C200}, (-86^\circ, 18^\circ)_{K201}, (-34^\circ, \text{undefined})_{L202} \)} for the MsaE crystal model, compared to {\( \phi, \psi = (-78^\circ, 166^\circ)_{V197}, (-71^\circ, 133^\circ)_{G198}, (-137^\circ, 120^\circ)_{Y199}, (-131^\circ, 46^\circ)_{C200}, (-116^\circ, 17^\circ)_{K201}, (-120^\circ, -1)_{L202} \)} for the MsaE crystal structure. Note that the MsaE numbering is used here for both molecules. The principal implication of the different backbone conformations involves the direction of the Cys200 C–C' vector, and the consequent inter-S distances between Cys53 (shown in the MsaE crystal configuration) and Cys200. The S' position is unknown from the RDC-defined structure and is shown here placed on the available cone with the assumption of a \( \phi = -60^\circ \) conformation. (b, bottom) Comparison of the Pro196–Leu202 peptide backbone conformation in MsaE determined from the direct structure calculation with RDC (yellow) and the equivalent conformation in MsaA (green). For clarity only the (C', C, C', N) atoms are shown.

NMR structure determination of large molecules by classical methods can be time-consuming due to the need for extensive assignment of backbone and side-chain resonances and the unambiguous identification of nOe correlations between these assigned frequencies. Structural genomics projects are currently providing an immense database, relating primary sequence to expected protein fold, an effort that still remains largely unexploited by the biomolecular NMR community. The routine measurement of RDC has recently been shown to provide a promising tool for low-resolution protein fold validation by comparatively simple experimental methodology. More precise structure determination is also possible with RDC but remains elusive for large molecules, where RDC data may be harder to measure extensively. In this study we propose a combination of these RDC-based techniques, initially for fold validation by use of a primary sequence homologue, followed by a focused structure determination of the site of interest, de novo, by only RDC.

This analysis required only backbone assignment and used only unambiguously assigned structural data, thereby greatly economizing investigation time and effort in comparison to established nOe-based structure calculation techniques, and demonstrates the enormous potential of using RDC for solution-state structural biology. Using the reaction site of the methionine sulfoxide reductase as an example, we have shown that RDC can provide precise local structure in functionally important regions of relatively large, highly deuterated molecules. This approach vastly simplifies the characterization of backbone structure in large molecules in solution, suggesting that this kind of analysis will have a significant impact on functional studies of biomolecules by NMR in the future.

**Acknowledgment.** This work was supported by the Commissariat à l’Energie Atomique and the Centre National de la Recherche Scientifique.

**Supporting Information Available:** Graph of order parameters from the region of interest (Figure S1) and a table of RDC data used in the analysis (Table S2). This material is available free of charge via the Internet at http://pubs.acs.org.