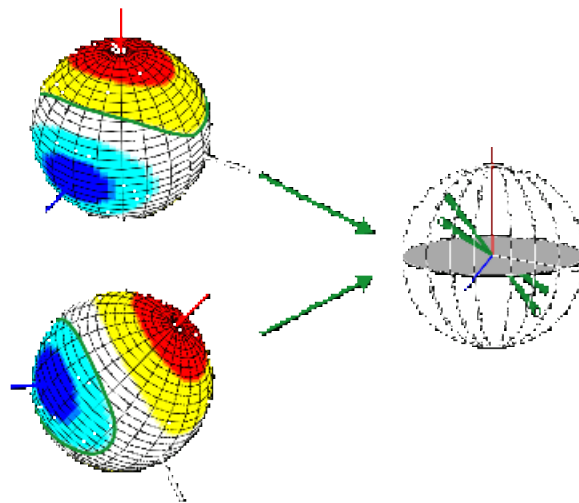


Residual Dipolar Couplings Measured in Multiple Alignment Media.

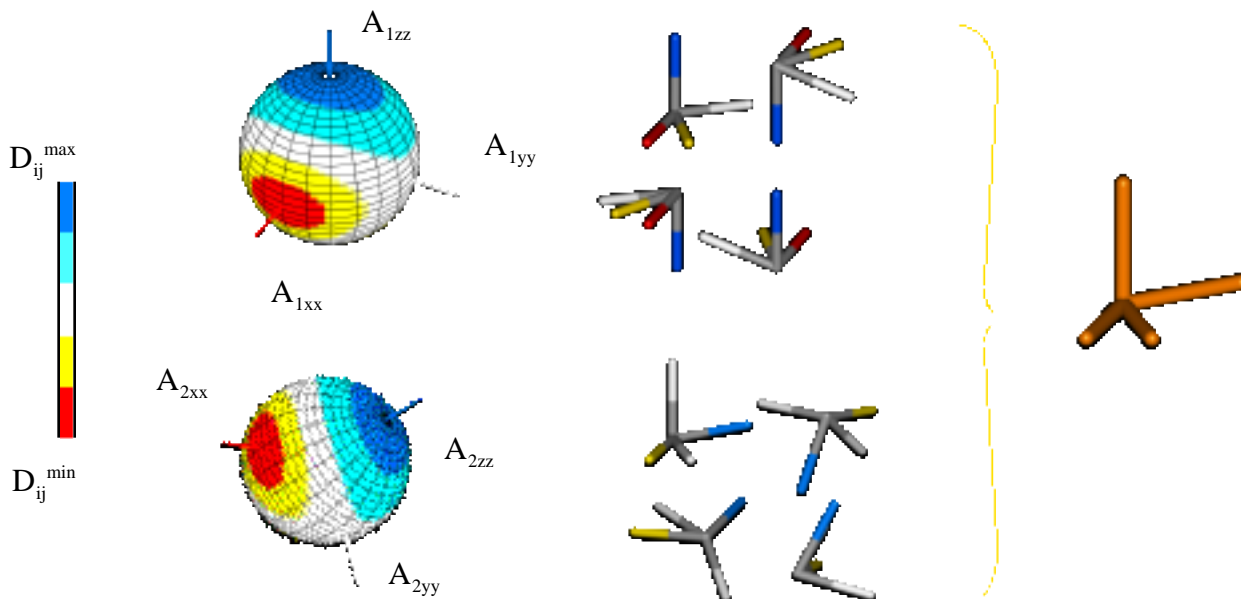
We have already seen that the orientational degeneracy inherent to a single measured coupling can be raised by measuring different directions in a rigid body. We will now turn our attention to the further gain in orientational definition achieved by measuring in the presence of a differently aligned 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;

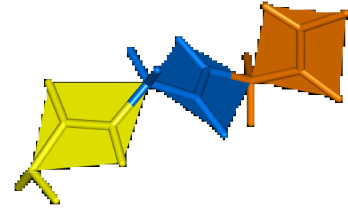


Four equivalent orientations exist for the sub-structure in the presence of the two different alignment tensors, related in both cases to 180° rotations about the respective axes of the two different alignment tensors, but only one of the four solutions in each case (the correct orientation) is common (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 -

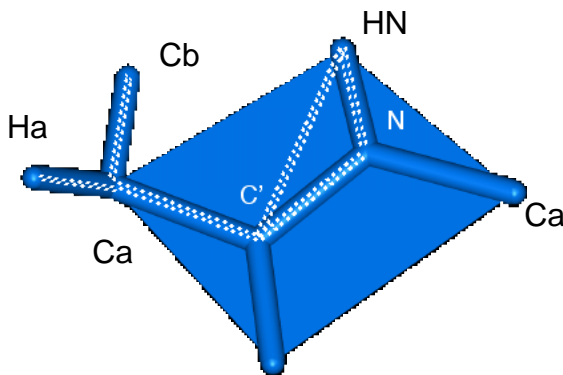
- 1) The molecule of interest can be divided into sub-structures whose local geometry is accurately known.
- 2) Sufficient residual dipolar couplings can be measured from these sub-structures to determine their relative orientation.
- 3) These couplings can be measured in two significantly differently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



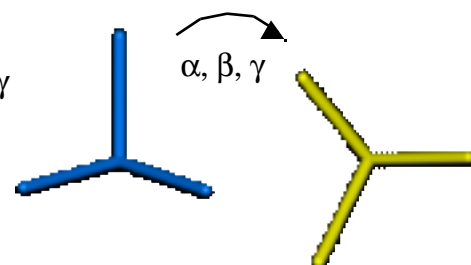
RDC's measured in this sub-unit are shown here. Two sets of 63 N-H^N, 61 C'-H^N, 61 and 63 C'-N and 59 and 54 C^α-C' couplings defining the peptide plane orientations, in addition to two sets of 62 C^α-H^α and one set of 39 C^α-C^β couplings

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 $(A^1_a, A^1_r, A^2_a, A^2_r, \alpha, \beta, \gamma)$ where α, β, γ describe the orientation of A^2 with respect to A^1 , taken to be diagonal in the calculation frame.

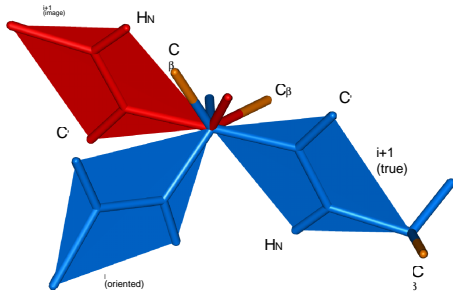


Simultaneously, the orientation of each peptide plane $(\theta_1, \theta_2, \theta_3)_i$ is determined with respect to the calculation frame. The number of parameters to be determined is then $(3N+7)$ describing each plane orientation and the common reference frame, while the number of RDC's available is $8N$, assuming we only include peptide planes

with complete datasets in this stage of the $\chi^2 = \sum ((D_i^{\text{exp}} - D_i^{\text{calc}}) / \sigma_i)^2$

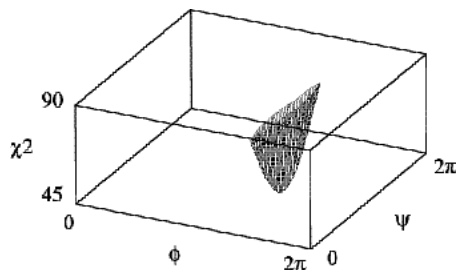
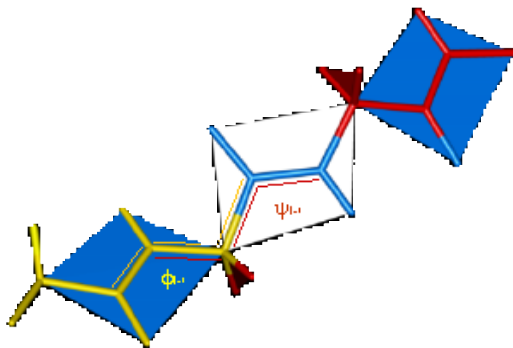
calculation. This algorithm reliably finds the global minimum of the target function over all measured couplings, requiring no *a priori* estimation of the alignment tensors. The second stage of the algorithm then constructs the molecule in this reference frame.

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

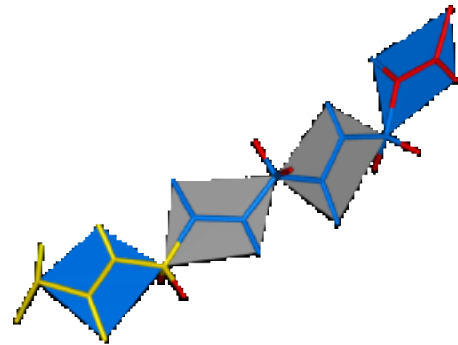


here as $C_{i-1}^\alpha, C'_{i-1}, N_i, C_i^\alpha$). The correct plane orientation can be distinguished from the mirror image by tetrahedral geometry requirements at the junctions connecting peptide planes, although for the general experimental case this is not always sufficient.

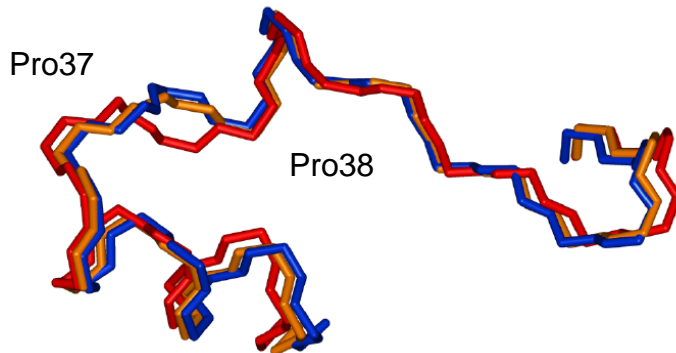
The combination of RDC measured in the peptide plane and tetrahedral junctions ($C_{i-1}^\alpha - H_{i-1}^\alpha$, $C_{i-1}^\alpha - C_{i-1}^\beta$) however allows unambiguous positioning. When no peptide plane orientation is available for plane $(i+1)$, ϕ_i / ψ_i values are optimized to reproduce $(C^\alpha - H^\alpha, C^\alpha - C^\beta)$ 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 ϕ_{36}/ψ_{36} , ϕ_{37}/ψ_{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).

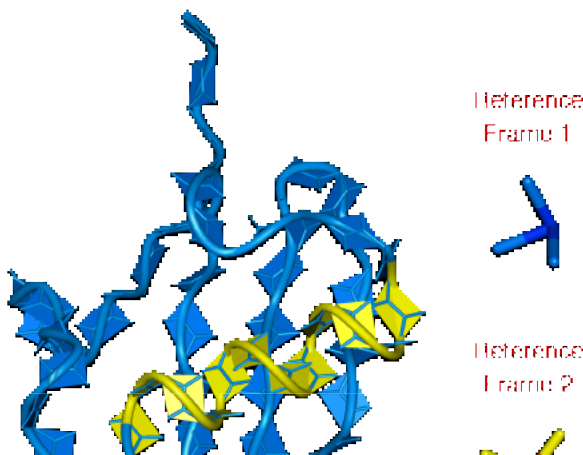


This example demonstrates an important aspect of RDC constraints, which provide coherent long-range structural information from throughout the molecule

relative to a fixed reference frame; this dependence on an external frame significantly reduces the problems of error propagation along the peptide chain.

The final structure, (1.3Å backbone rmsd from the structure determined using 2727 nOe, in combination with RDC, 3J 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 interactions were ignored completely, 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 by the groups of Cross and Opella (Cross & Opella 1994, Ketchem et al 1996) and recently described work by the Prestegard group in partially aligned proteins (Prestegard et al 2000).

Residual dipolar couplings also provide otherwise inaccessible information concerning intramolecular and inter-domain mobility, due to the dynamic averaging of the dipolar interaction. RDC are sensitive to motions over a much longer timescale than those sampled by heteronuclear relaxation measurements, and can therefore be used to complete our understanding of macromolecular dynamics. While the contribution of dynamics to measured couplings has been recognized since the beginning of their exploitation for structure determination, this potentially rich source of information has so far only rarely been exploited (Fischer *et al* 1999, Tolman *et al* 2001, Meiler *et al* 2001).

Annala, A.; Aitio, H.; Thulin, E.; Drakenberg, T. *J. Biomol. NMR* **1999**, *14*, 223-230.

Al-Hashimi, H., Valafar, H., Terrell, M., Zartler, M., Eidsness, M. and Prestegard, J.H.. *J. Magn. Reson.* **2000**, *143*, 402-406.

Andrec, M.; Du, P.; Levy, R.M. *J. Am. Chem. Soc.* **2001**, *123*, 1222-1229

Cai, M., Huang, Y., Zheng, R., Wei, S., Ghirlando, R., Lee, M., Craigie, R., Gronenborn, A.M. & Clore, G.M. (1998) *Nat. Struct. Biol.* **5**, 903-909.

Clore, G.M. (2000) *Proc. Natl. Acad. Sci.* **97**, 9021-9025.

Clore, G.M.; Gronenborn, A.M. *Nature Struct. Biol* **1997**, *4*, 849-853.

Clore, G.M. and Gronenborn, A.M. (1998) *Proc. Natl. Acad. Sci.* **95**, 5891-5898.

Clore, G.M., Gronenborn, A.M. and Tjandra, N. (1998) *J. Magn. Reson.* **131**, 159-162.

Clore, G.M.; Starich, M.R.; Bewley, C.A.; Cai, M.; Kuszewski. *J. Am. Chem. Soc.* **1999**, *121* 6513-6514.

Cornilescu, G.; Marquardt, J.L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836-6837.

Cross, T.; Opella. S. *Ann. Rep. NMR. Spectr.* **1994** *29*, 123-167.

Delaglio, F.; Kontaxis, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 2142-2143.

Dosset, P., Hus, J.-C., Blackledge, M. and Marion, D. (2000) *J. Biomol. NMR.* **16**, 23-28.

Ferrarini, A., Moro, G., Nordio, P.L. Luckhurst, G.L. (1992) *Mol Phys* **77**, 1-15.

Fischer, M.W.F., Losonczy, J.A., Weaver, J.L. and Prestegard, J.H. (1999) *Biochemistry* **38**, 9013-9022.

Garrett, D.S.; Seol, Y-J; Peterkovsky, A.; Gronenborn, A.M.; Clore, G.M. *Nature Struct. Biol* **1999**, *6*, 166-173.

Gayathri, C.; Bothner-by, A.A.; van Zijl, P.C.M.; Maclean, C. *Chem. Phys. Lett.* **1982**, *87*, 192-196.

Hansen, M.R.; Müller, L.; Pardi, A. *Nature Struct. Biol.* **1998**, *5*, 1065-1074.

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

Hus, J.-C., Marion, D. and Blackledge, M. (2000) *J. Mol. Biol.* **298**, 927-936.

Ketchum, R.; Lee, K.; Huo, S.; Cross, T. *J. Biomol. NMR*, **1996**, *8*, 1-14.

Leondes, C.T. (1997) *Fuzzy Logic and Expert Systems Applications*. Academic Press, San Diego, U.S.A.

Meiler, J.; Peti, W.; Griesinger, C. *J. Biomol. NMR* **2000**, *17*, 283-294.

Meiler, J.; Blomberg, N.; Nilges, M.; Griesinger, C. *J. Biomol. NMR* **2000**, *16*, 245-252.

Meiler, J.; Peti, W.; Prompers, J; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953) *J. Chem. Phys.* **21**, 1087-1094.

Mollova, E.T., Hansen, M.R. and Pardi, A. (2000) *J. Am. Chem. Soc.* **122**, 11561-11562.

Moltke, S.; Grzesiek, S. *J. Biomol. NMR* **1999**, *15*, 77-82.

Olejniczak, E.T., Meadows, R.P., Wang, H., Cai, M., Nettlesheim, D.G. and Fesik, S. (1999) *J. Am. Chem. Soc.* **121**, 9249-9251.

Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 12334-12341.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1988) *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge University Press, Cambridge, U.K.

Prestegard, J.; Tian, F.; Tolman, J. *Q. Rev Biophys.* (2000) **33**, 371-424.

Ramirez, B.; Bax, A. (1998) *J. Am. Chem. Soc.* **120**, 9106-9107.

Sanders, C.R.; Hare, B.J.; Howard, K.P.; Prestegard, J.H. *Progr. N.M.R. Spectr.* **1994**, *26*, 421-444.

Saupe, A. ; Englert, G. *Phys.Rev.Lett.* **1963**, *11*, 462-465.

Sass, J.; Cordier, F.; Hoffman, A.; Cousin, A.; Omichinski, J.G.; Lowen, H.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 2047-2055.

Skrynnikov, N., Goto, N.K., Yang, D., Choy, W.-Y., Tolman, J.R., Mueller, G.A. and Kay, L. (2000) *J.Mol.Biol.* **295**, 1265-1273.

Tjandra, N. (1999), *Structure*, **7**, R205-R211.

Tjandra, N. and Bax A. (1997), *Science*, **278**, 1111-1114.

Tjandra, N., Omichinski, J., Gronenborn, A.M., Clore, G.M. and Bax, A. *Nat. Struct. Biol.* **1997** *4*, 732-738.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc, Natl. Acad. Sci.* **92**, 9279-9283.

Tolman, J.R.; Flanagan, J.M.; Kennedy, M.A.; Prestegard, J.H. *Nature Struct.Biol* **1997**, *4*, 292-297.

Tolman, J.R.; Al-Hashimi, H.M.; Kay, L.E.; Prestegard, J.H. *J. Am. Chem. Soc.* **2001**, *123*, 1416-1424.

Tsui, V., Zhu, L., Huang, T.-H., Wright, P.E. & Case, D.A. (2000) *J. Biomol. NMR* **16**, 9-21.

Weaver, J.L. and Prestegard, J.H. (1998) *Biochemistry*, **37**, 116-128.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.

Zimmerman, G.R.; Jenison, R.D.; Wick, C.L.; Simorre, J.-P.; Pardi, A. *Nature Struct.Biol* **1997**, *4*, 644-649.

Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122* 3791-3792.