Journal of Biomolecular NMR, **20:** 223–231, 2001.
KLUWER/ESCOM

A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings

Patrice Dosset, Jean-Christophe Hus, Dominique Marion & Martin Blackledge*
Institut de Biologie Structurale-Jean-Pierre Ebel, C.N.R.S.-C.E.A., 41 rue Jules Horowitz, F-38027 Grenoble Cedex, France

Received 17 January 2001; Accepted 2 April 2001

Key words: alignment tensor, liquid crystal, modular domains, partial alignment, residual dipolar coupling, rigid body modeling

Abstract

Residual dipolar couplings (RDC), measured by dissolving proteins in dilute liquid crystal media, or by studying naturally paramagnetic molecules, have rapidly become established as routine measurements in the investigation of the structure of macromolecules by NMR. One of the most obvious applications of the previously inaccessible long-range angular information afforded by RDC is the accurate definition of domain orientation in multi-module macromolecules or complexes. In this paper we describe a novel program developed to allow the determination of alignment tensor parameters for individual or multiple domains in macromolecules from residual dipolar couplings and to facilitate their manipulation to construct low-resolution models of macromolecular structure. For multi-domain systems the program determines the relative orientation of individual structured domains, and provides graphical user-driven rigid-body modeling 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 are used to position the different modules in the common frame to find a model in best agreement with experimentally measured couplings alone or in combination with additional experimental or covalent information.

Introduction

While NMR spectroscopy is now successfully established as the most important technique for the high resolution structure determination of small to medium sized, compact macromolecules in the solution state (Wüthrich, 1986; Clore and Gronenborn, 1998), the method is severely limited for more complex molecular systems. The basic experimental parameter used for the determination of molecular structure (nuclear Overhauser effect – NOE) becomes difficult to measure in large protonated molecules due to prohibitive relaxation effects, making the determination of structure beyond 30 kDa unrealistic using classical techniques (Gardner and Kay, 1998). Moreover, modular or elongated proteins, and large RNA superstructures, encounter the serious problem of ill-defined relative orientation of different domains, due to inadequate local structural information at interfacial or hinge regions. The relative orientation of different domains is, however, known to be closely correlated to physiological function while the characterization of the exact nature of molecular interaction in reaction complexes clearly holds the key to understanding macromolecular function.

The last five years have seen a rapid acceleration in the search for viable, alternative sources of structural information for the resolution of long-range orientation in systems of more complex geometry (Tjandra, 1999). In particular the dependence of heteronuclear relaxation rates on the orientation of the relevant interaction is, under certain experimental conditions, sufficiently precise to determine the alignment of individual structural domains relative to the molecular diffusion tensor (Brüschweiler et al., 1995; Tjandra

*To whom correspondence should be addressed. E-mail: martin@rmn.ibs.fr

et al., 1997; Fushman et al., 1999; Hus et al., 1999). More generally, weak alignment of proteins prevents complete averaging of the dipolar interaction, while retaining the solution properties necessary for high resolution NMR. This alignment can exist naturally, due to the paramagnetic properties of the molecule (Tolman et al., 1995), or can, more generally, be induced by solvation in liquid crystal media (Tjandra and Bax, 1997). The residual dipolar coupling (RDC) measured under these conditions provides geometric information relative to the common alignment frame of the form

$$D_{ij} = \quad (1)$$

$$-S\frac{\gamma_i\gamma_j\mu_0 h}{16\pi^3 r_{ij}^3}\left(A_a(3\cos^2\theta - 1) + \frac{3}{2}A_r\sin^2\theta\cos 2\varphi\right)$$

A_a and A_r are the axial $1/3(A_{zz} - (A_{xx} + A_{yy})/2)$ and rhombic $1/3(A_{xx} - A_{yy})$ components of the alignment tensor, and $\{\theta, \varphi\}$ is the vector orientation relative to this tensor, r_{ij} is the internuclear distance and S the local order parameter. RDC have been shown to provide previously inaccessible structural definition in 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).

Residual dipolar couplings provide the first example of routine collection of coherent long-range structural information from throughout the molecular system using solution state NMR. This opens up exciting possibilities for rapid low-resolution global-fold screening by comparison of measured experimental couplings with expected distributions from proteins of known high resolution structure present in conformational databases (Annila et al., 1998). Such methods of protein fold identification from readily available experimental data are highly complementary to recently developed database-mining algorithms designed to predict structure and function from protein sequences alone (Jones, 2000; Simmerling et al., 2000). The development of appropriate algorithms to optimally exploit RDC data for the investigation of long-range order represents a major challenge for the NMR community, whose principal interests in the past have been concerned with short-range geometric constraints. Novel tools which automatically search databases for complete homologous structures to predict fold (Meiler et al., 2000), or reconstruct peptide chain structure from known molecular fragments (Delaglio et al., 2000), have recently contributed to the diverse approaches currently under investigation. It has also recently been shown that in the presence of sufficient RDC measured from throughout the peptide chain, it is possible to construct the backbone of the protein ubiquitin using only these data (Hus et al., 2001).

In the case of multimeric, extended macromolecules or molecular complexes, RDC provide previously inaccessible information concerning the orientation of different regions of the macromolecule. These data are again complementary to alternative 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. Interpretation of residual dipolar couplings for the determination of domain orientation requires tools specifically developed for the manipulation of sub-structures within a reference calculation frame. As currently available molecular modeling packages are not yet adapted to handling this kind of specific analysis, we have developed a program (Module) which determines alignment tensor parameters and graphically displays the tensor relative to the three-dimensional atomic coordinates, as well as correlation plots of the measured and calculated couplings for the selected datasets. Module also provides graphical user-driven, rigid-body modeling of the individual modules of multi-domain assemblies by simple cursor-driven manipulation, for the determination of the relative position of structural motifs with respect to the common alignment tensor **A**.

Methods

Module requires two sources of input information: the measured residual dipolar coupling values D_{ij}, their associated uncertainty σ_{ij} and an estimation of the order parameter S (for many applications this will of course be assumed to be 1), and a standard coordinate file from the Brookhaven data bank containing the structure under investigation (protein or nucleic acid).

The program allows the user to define regions to be taken into consideration as separate structural entities. The alignment tensor will be calculated for this 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

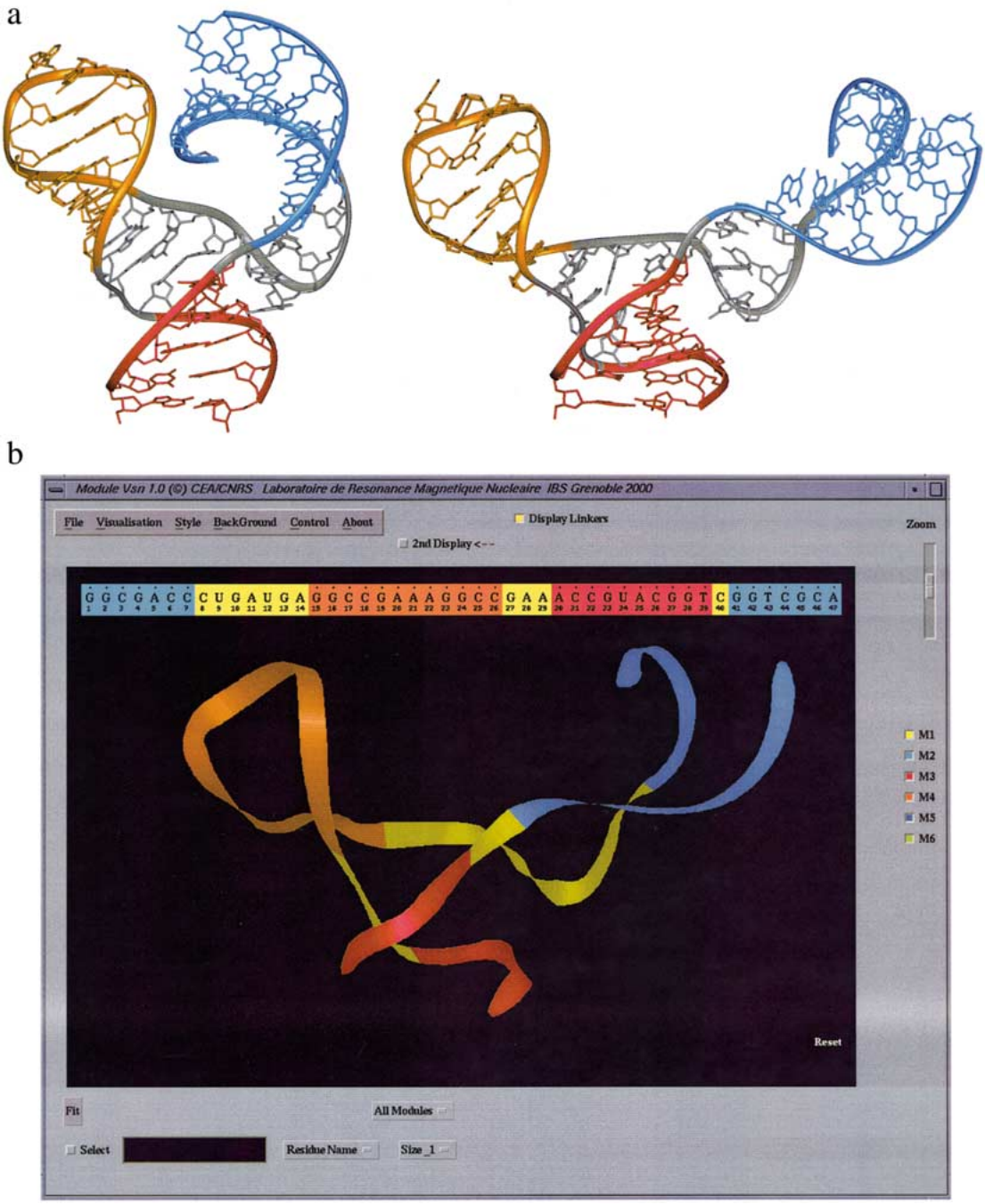


Figure 1. (a) Comparison of the X-ray crystallographic structure of RNA/DNA ribozyme inhibitor (left, pdb code 1mmh) and the structure used as initial model for the simulated experiment using Module (right). The native structure was partially unfolded using high-temperature restrained molecular dynamics 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 10.5 Å. (b) Presentation of the user interface of the program MODULE. 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.

(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:

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

where σ_{ij} is the uncertainty in the experimentally measured coupling. The minimization algorithm searches the $\{A_a, A_r, \alpha, \beta, \gamma\}$ 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 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_a and A_r 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 are presented for each different coupling type, as well as the χ^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 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 paper (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 inter-domain 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

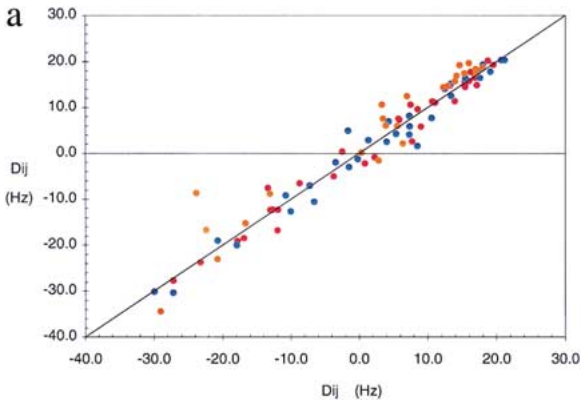


Figure 2. Determination of the relative orientations of the secondary structural elements stems I–III in the hammerhead ribozyme using simulated Residual Dipolar Couplings and Module (this figure was produced entirely using the graphic interface of Module). (a) Noise-simulated and fitted data from the three stem regions of the ribozyme. The blue data points are from the fit to stem I, orange from stem II and red from stem III. (b) Top left: The alignment tensors of the different modules are determined and their eigenvectors superposed on the structures in their original (unwound) orientation. Bottom left: 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. Top right: 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. Bottom right: The final structure calculated automatically has a backbone rmsd of 2.5 Å compared to the crystal structure.

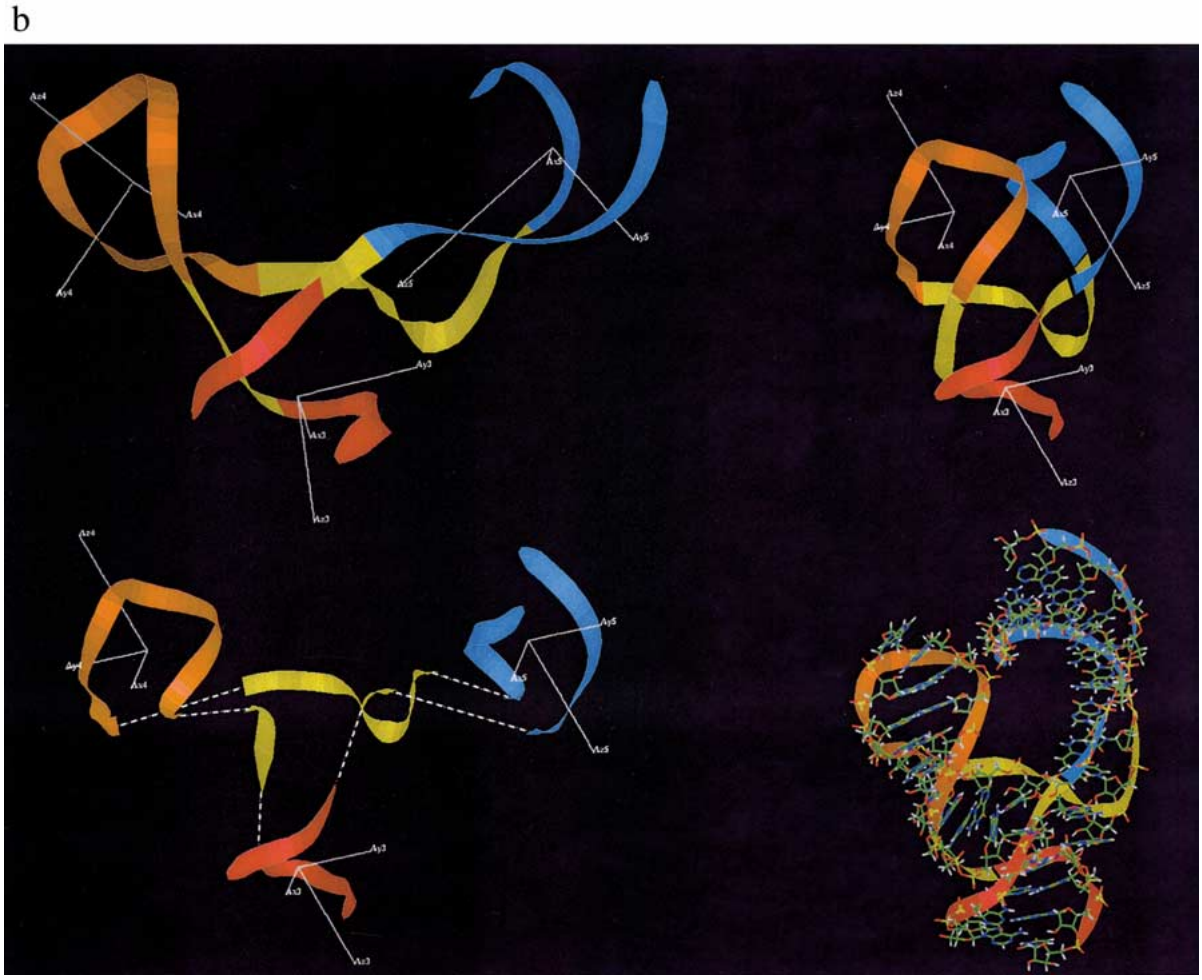


Figure 2. (continued).

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

$$E_{cov} = \sum_{n} \{d_{ij} - d_{ij}^{cov}\}^2 \quad (3)$$

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. Once the preferred orientation has been found, the model can be fixed, and the coordinates written to a standard format coordinate file, or transferred to a standard molecular dynamics package for further refinement under RDC constraint forces (Clore et al., 1998; Hus et al., 2000; Tsui et al., 2000).

Results

Two examples have been chosen to illustrate the use of Module: in both cases we have simulated data from theoretical alignment tensors in systems where orientational information would be particularly valuable. The first is the hammerhead ribozyme, whose three-dimensional structure has been determined using

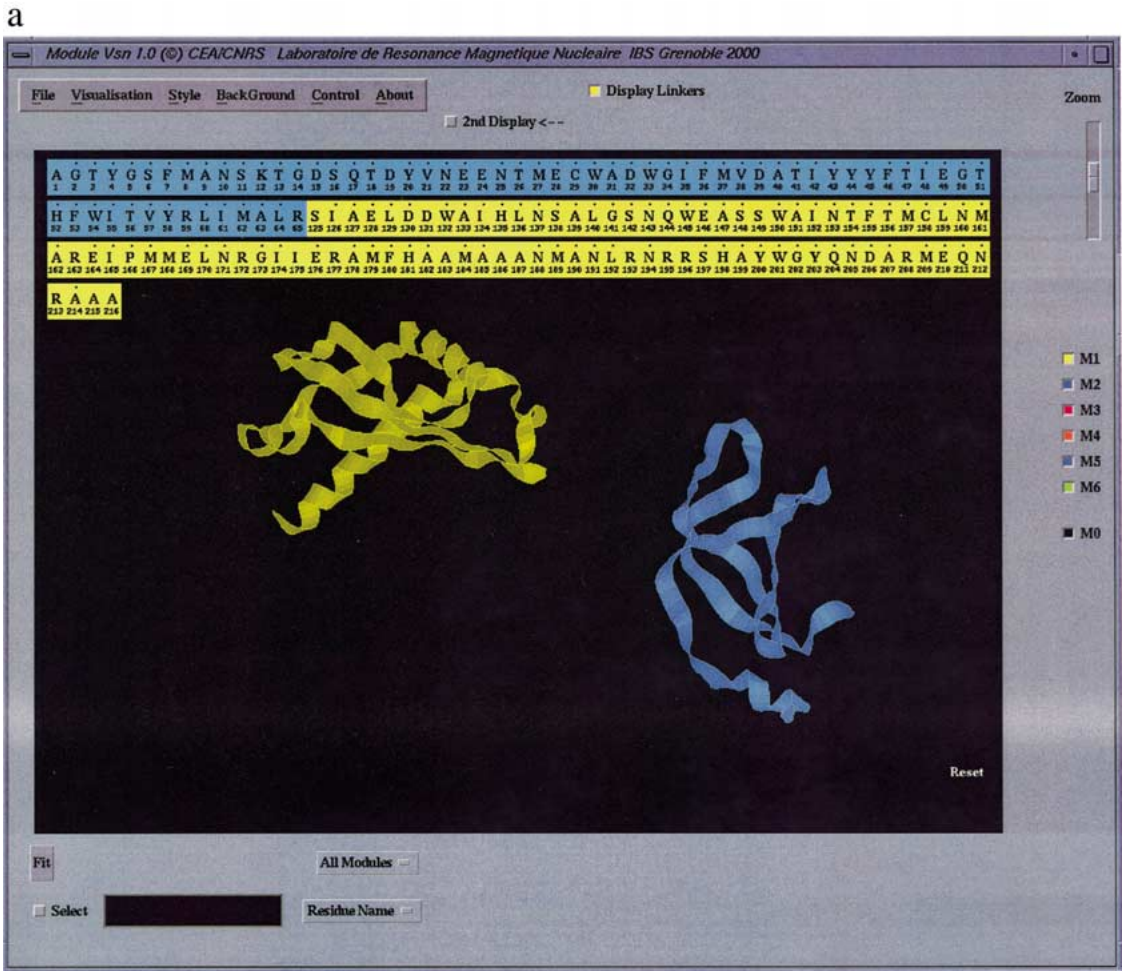
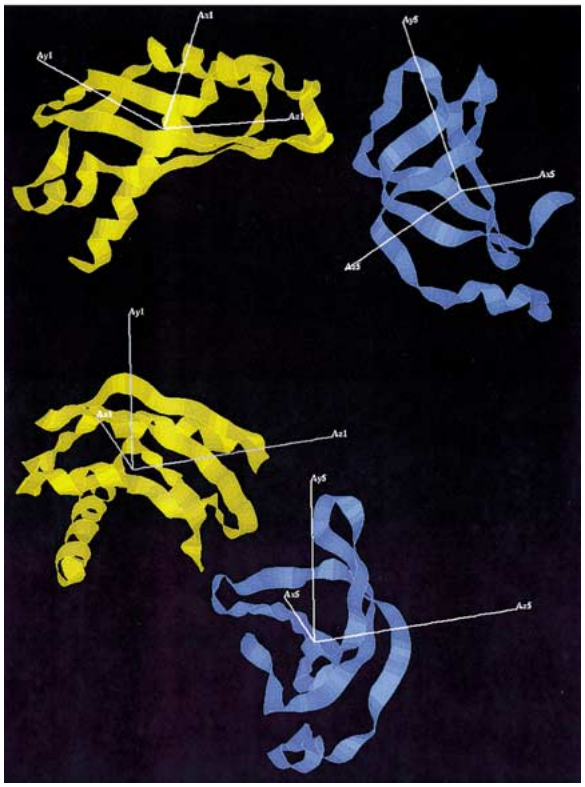


Figure 3. Determination of the relative orientations of the complexed proteins Tol-a III and GP3 (this figure was produced entirely using the graphic interface of Module). (a) Presentation of the user interface of the program MODULE. The two proteins are treated as individual domains, again selected from the primary sequence (blue = GP3, yellow = Tol-a III). (b) Top: The alignment tensors of the proteins are determined and their eigenvectors superposed on the structures in their original (pdb) orientation. Bottom: The proteins are rotated so that the tensors all have the same orientation in the alignment tensor frame indicated by the tensor directions. The proteins can be manipulated separately in this frame, using only translational freedom and rotations about A_{xx}, A_{yy} and A_{zz}. (c) There is no covalent interaction between the proteins, so in order to select between the multiple possible solutions, the user can select points on preferred interaction surfaces of the two molecules to aid the model building. These may be experimental, such as intermolecular NOE measured between interacting surfaces, or predicted from existing structural information, for example electrostatic or hydrophobic surface calculations. Here we have kept the orientation of the Tol-a constant and show the four equivalent orientations of the GP3 protein, due to the inherent degeneracy of π rotations about A_{xx}, A_{yy} and A_{zz}. (d) Again the preferred conformation, respecting the known interaction surface, can be stored for further refinement using molecular dynamics or more specific modelling procedures.

X-ray crystallography (Pley et al., 1994). This small catalytic RNA comprises three canonical regions of consensus secondary structure in the form of A-type helices, with, in the case of stem II, an additional GAAA tetraloop configuration, folded around the central core of the molecule. It has recently been demonstrated that residual dipolar couplings can contribute important information to the determination of RNA global fold (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 (Saenger, 1984). 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

b



c

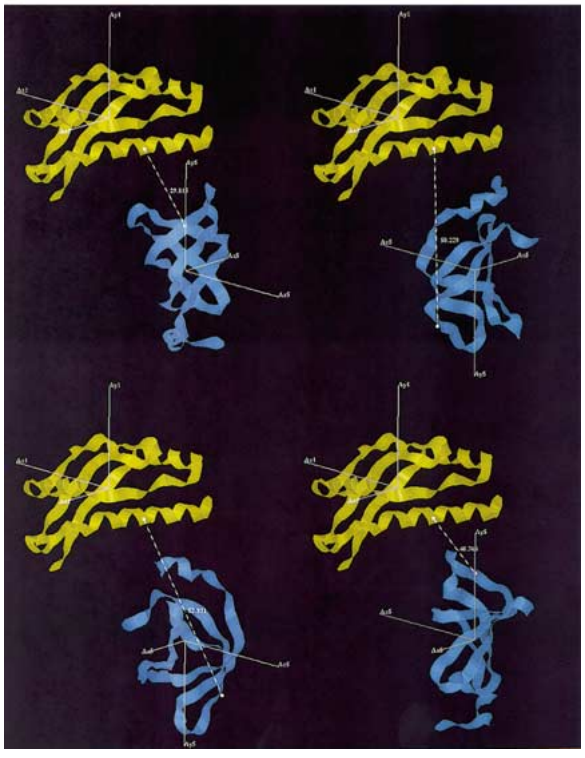


Figure 3. (continued).

d

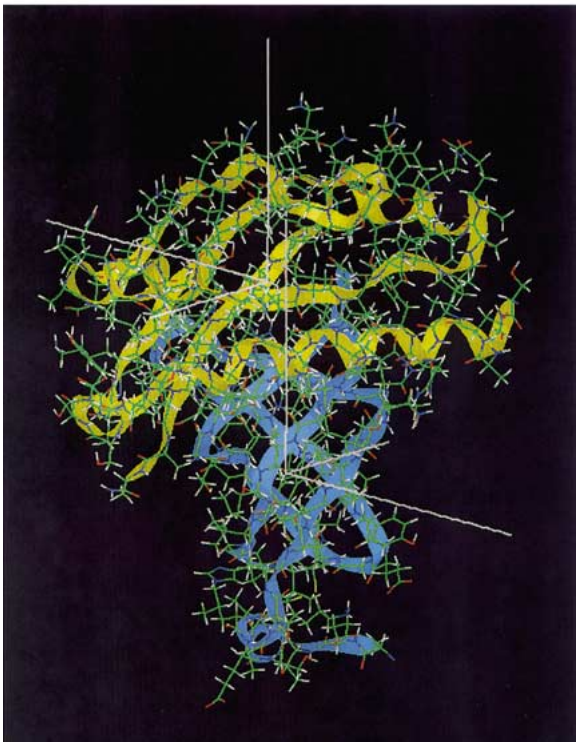


Figure 3. (continued).

noise to the simulated values. The alignment tensor was assumed to be

$$\frac{\gamma_C \gamma_H \mu_0 h}{16\pi^3 r_{ij}^3} A_a = 12.2 \text{ Hz}$$

$$\frac{\gamma_C \gamma_H \mu_0 h}{16\pi^3 r_{ij}^3} A_r = 1.5 \text{ Hz}$$

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) with 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 (Figure 1). The structure of the core region was conserved using a tethering force on the initial positions of the heavy atoms in the relevant residues by incorporating the additional energy term

$$E_{teth} = K \sum_i \sqrt{(x_i - x_i^0)^2 / N} \qquad (4)$$

into the potential energy function. x_i are the Cartesian coordinates of the atoms to be tethered and x_i^0 the target coordinates. A force constant of $K_{teth} = 100.0$

kcal · mol^{-1}Å^{-1} was used to restrain the N = 278 atoms to the coordinates of the crystal structure. The secondary structural regions were restrained using distance restraints from the canonical A-form helices and the crystal structure tetraloop. This non-native structure (heavy atom rmsd of 10.5 Å compared to the initial, correct structure) and the simulated couplings were then used to reconstruct a model of the molecule using Module (Figure 2). The alignment tensors are fitted for the stem regions (I–III), which are then automatically aligned in the reference frame of a common tensor. In this case the tensors are virtually identical, as the data are all calculated assuming the same simulated system. It is then possible to organise the three oriented domains relative to the core, either manually or automatically (the example shown in Figure 2 was positioned by minimising Equation 3), to find a model in agreement with the orientational data and preserving the known covalence (heavy atom rmsd of 2.5 Å compared to the initial, correct structure).

The second example concerns a recently published molecular complex between the minor coat protein from Gene III in phage M13 (G3P) (86 amino acids) and the C-terminal domain of *E. coli* protein TolA (126 amino acids). Again this complex has recently been crystallised, and its structure has been determined using X-ray diffraction (Lubkowski et al., 1999). This structure was used to simulate experimental residual dipolar coupling data from NH sites distributed throughout the two molecules and 5% stochastic noise added to these simulated values. The tensor used in this case has eigenvalues

$$\frac{\gamma_N \gamma_H \mu_0 h}{16\pi^3 r_{ij}^3} A_a = -12.4\,\text{Hz}$$

$$\frac{\gamma_N \gamma_H \mu_0 h}{16\pi^3 r_{ij}^3} A_r = -5.4\,\text{Hz}$$

and S was again assumed to be equal to 1 in all cases. The individual protein structures were then aligned using Module, as shown in Figure 3. In this case the degeneracy of relative orientation plays a more significant role, as there is no covalence between the two partners. Sparse experimental data, derived from chemical shift mapping or intermolecular NOE, or indications derived from the physical or chemical properties of the reaction partners, can be used to select the interaction faces of the two molecules, and thereby propose the most likely relative molecular orientation. An interactive mode allows the user to define atom pairs and to display these distances throughout while positioning the modules. In this case atoms chosen on the hydrophobic faces of both proteins were selected to ensure that this face was involved in the contact surface. An additional tool allows the user to read in distances between atoms in different domains, which the program will then use to automatically propose a model in best agreement with RDC data and the measured distances.

Conclusions

We have developed an interactive tool for the determination of alignment tensors derived from residual dipolar coupling measurements for partially oriented molecules. In particular, the program has been designed to allow molecular modeling of multi-domain macromolecules in the context of orientational restraints in a common alignment tensor frame, incorporating the inherent angular degeneracy and axial symmetry where applicable. The facility with which Module can be used to define preliminary models of segmental or modular systems has been illustrated for a complex nucleic acid assembly and protein–protein interaction system, but we feel sure that the utility of the program will be manifest in many diverse applications. The program is available free of charge from the address given below.

Software availablility

MODULE is currently available from our web-site at www.ibs.fr/ext/labos/LRMN/welcome_en.htm#software.

Acknowledgements

This work was supported by the Commisariat à l'Energie Atomique and the Centre National de la Recherche Scientifique.

References

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

Annila, A., Aitio, H., Thulin, E. and Drakenberg, T. (1999) *J. Biomol. NMR*, **14**, 223–230.

Brüschweiler, R., Liao, X. and Wright, P. (1995) *Science*, **268**, 886–889.

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

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

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

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

Delaglio, F., Kontaxis, G. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 2142–2143.

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

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

Fushman, D., Xu, R. and Cowburn, D. (1999) *Biochemistry*, **38**, 10225–10230.

Gardner, K. and Kay, L. (1998) *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357–406.

Hus, J.-C., Marion, D. and Blackledge, M. (1999) *J. Am. Chem. Soc.*, **121**, 2311–2312.

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

Hus, J.-C., Marion, D. and Blackledge, M. (2001) *J. Am. Chem. Soc.*, **123**, 2311–2312.

Jones, D.T. (2000) *Curr. Opin. Struct. Biol.*, **10**, 371–379.

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

Lubkowski, J., Hennecke, F., Pluckthun, A. and Wlodawer, A. (1999) *Structure*, **7**, 711–722.

Meiler, J., Peti, W. and Griesinger, C. (2000) *J. Biomol. NMR*, **17**, 283–294.

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.

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

Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994) *Science*, **372**, 68–74.

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.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, NY.

Simmerling, C., Lee, M.R., Ortiz, A.R., Kolinski, A., Skilnick, J. and Kollman, P. (2000) *J. Am. Chem. Soc.*, **122**, 8392–8402.

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., Garrett, D.S., Gronenborn, A.M., Bax, G.M. and Clore, G.M. (1997) *Nat. Struct. Biol.*, **4**, 443–449.

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

Tsui, V., Zhu, L., Huang, T.-H., Wright, P.E. and 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, NY.