Available online at www.sciencedirect.com
SCIENCE DIRECT®
C. R. Physique 5 (2004) 359–375
COMPTES RENDUS PHYSIQUE

Highly polarized nuclear spin systems and dipolar interactions in NMR/Systèmes de spins nucléaires fortement polarisés et interactions dipolaires en RMN

Orienting molecular fragments and molecules with residual dipolar couplings

Nikolai R. Skrynnikov

Department of Chemistry, Purdue University, West Lafayette, IN 47907-2084, USA

Available online 15 April 2004

Presented by Guy Laval

Abstract

Many proteins and nucleic acids, as well as their complexes, consist of structural units with a known internal structure. However, relative position of these units within the entity is often a subject of uncertainty. This problem can be effectively addressed with the help of residual dipolar couplings (RDCs) measured by solution-state NMR in weakly oriented samples. Using a rigid-body approach, the relative orientation of the structural units can be adjusted to match the experimental RDC data. This paper explains why the rigid-body approach is well suited for application with the RDC data and discusses the general algorithm suitable for such studies. The examples of RDC-based rigid-body treatment involving protein backbone fragments, RNA helices, multidomain proteins, protein-ligand systems, and protein complexes are described in detail. The discussion includes systems where multiple structures co-exist in a dynamic equilibrium. Various degeneracies associated with the RDC data are considered and the methods that can assist in resolving the ambiguities are described. ***To cite this article: N.R. Skrynnikov, C. R. Physique 5 (2004).***

Résumé

Orienter des fragments moléculaires et des molécules grâce aux couplages dipolaires résiduels. Les protéines et les acides nucléiques, tout comme leurs complexes, sont constitués de différents domaines dont la structure est souvent connue. Toutefois, la connaissance de la position relative de ces éléments reste souvent peu précise. Une bonne solution à ce problème consiste à exploiter les couplages dipolaires résiduels (RDCs) mesurés par des techniques de RMN en solution, en utilisant des échantillons faiblement orientés. En s'appuyant sur une approche basée sur des fragments rigides, l'orientation relative des différents éléments structuraux peut être optimisée afin d'être en accord avec les données de RDCs. Cet article explique pourquoi cette approche est bien adaptée à leur exploitation et discute son implémentation. Des exemples d'une telle approche appliquée à des fragments de protéines, à des hélices d'ARN, à des protéines multidomaines, à des systèmes ligand-protéine ou encore à des complexes de protéines sont décrits en détails. La discussion comprend également le cas des systèmes en équilibre conformationnel. Les différentes dégénérescences inhérentes aux couplages dipolaires résiduels sont considérées ainsi que les méthodes permettant de lever ces ambiguités. ***Pour citer cet article : N.R. Skrynnikov, C. R. Physique 5 (2004).***

Keywords: Residual dipolar couplings; Solution-state NMR

Mots-clés : Couplages dipolaires résiduels ; RMN en solution

E-mail address: nikolai@purdue.edu (N.R. Skrynnikov).

1631-0705/$ – see front matter
doi:10.1016/j.crhy.2004.02.006

1. Introduction

Magnetic dipole–dipole interaction plays a major role in spin spectroscopy. Whereas in the solid-state NMR dipolar interaction manifests itself directly through the fine structure of spectral lines (which is typically complex and not well resolved), in solution these effects are canceled because of random molecular tumbling. On the time scale of the solution-state NMR experiment the dipolar interactions average to zero. However, over a very short time interval when the molecules in solution can be considered approximately static, dipolar interactions have a non-vanishing effect on the evolution of spin magnetization. This ultimately leads to line broadening and other relaxation phenomena associated with solution-state NMR, including nuclear Overhauser effect (NOE).

If a molecule in solution develops preferential orientation with respect to the external frame of reference then dipolar interactions no longer average to zero. In this situation spectral lines acquire multiplet structures identical to solid state. The splitting between the components of the multiplet depends on the magnitude of the time-averaged dipolar interaction and, consequently, on the degree of molecular orientation (degree of alignment). In practice, molecular alignment can be achieved using the interactions between the molecule and external electromagnetic fields or steric interactions involving anisotropic media. The degree of alignment can be adjusted to ensure that the spectrum looks like a regular solution-state spectrum except for multiplet patterns reflecting a limited number of strongest dipolar interactions. The multiplet splittings observed in this manner became known as residual dipolar couplings (RDCs). Similar to solid-state data, RDCs encode a wealth of useful structural information.

In biological macromolecules, residual dipolar couplings were first measured by Tolman, Prestegard, and co-workers using the effect of paramagnetic alignment in a strong external field [1]. Later, Tjandra and Bax used dilute liquid crystals to induce the alignment in diamagnetic proteins [2]. Residual dipolar coupling between spins I and M can be expressed as:

$$D_{IM} = D^0_{IM} A_a S \left\{ \left(3\cos^2\theta - 1\right) + \frac{3}{2} R \sin^2\theta \cos 2\phi \right\}, \tag{1}$$

where $D^0_{IM} = -(1/2\pi)(\mu_0/4\pi)\hbar\gamma_I\gamma_M\langle r^{-3}_{IM}\rangle$ is the dipolar interaction constant, S is the order parameter reflecting partial averaging due to local dynamics, A_a is the axial component of the molecular alignment tensor which is usually referred to as $\Delta\chi$ in case of direct alignment via magnetic susceptibility, R is the rhombicity parameter of the allignment tensor that is confined to the range from 0 to 2/3, and θ and ϕ are polar angles that specify the orientation of the IM internuclear vector with respect to the molecular alignment frame. The angles $\{\theta, \phi\}$ in the alignment frame are related to $\{\theta', \phi'\}$ in a given molecular frame via:

$$\begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix} = \begin{pmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & \sin\alpha\cos\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\beta\cos\gamma \\ -\cos\alpha\cos\beta\sin\gamma - \sin\alpha\cos\gamma & -\sin\alpha\cos\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\beta\sin\gamma \\ \cos\alpha\sin\beta & \sin\alpha\sin\beta & \cos\beta \end{pmatrix} \begin{bmatrix} \sin\theta'\cos\phi' \\ \sin\theta'\sin\phi' \\ \cos\theta' \end{bmatrix}, \tag{2}$$

where α, β, and γ are the three Euler angles.

Given a data set of 5 or more RDCs and a set of atomic coordinates, the alignment parameters $\{A_a, R, \alpha, \beta, \gamma\}$ can be determined by least-square fitting of these data to Eqs. (1) and (2). A number of non-linear minimization methods have been used toward this goal. Alternatively, the equations can be linearized with respect to the elements of the alignment tensor and quickly solved by means of singular-value decomposition [3]. The RDC input for the fitting procedure is often edited based on, for example, ^{15}N relaxation data to exclude highly mobile residues [4,5]. An iterative approach that eliminates the worst outliers has also been used in this context [6]. The quality of the fit can be assessed using the quality factor $R_{\rm dip}$ [7,8]:

$$R_{\rm dip} = rms\left(D^{\rm obs}_{IM} - D^{\rm calc}_{IM}\right) / \left(2\sqrt{2/5}\right) D^0_{IM} S\left\{A_a\sqrt{1 + (3/4)R^2}\right\}, \tag{3}$$

where the term in curly brackets is the 'general degree of order' [9] which characterizes the strength of the alignment. When RDCs are fit using the appropriate high-resolution X-ray structures $R_{\rm dip}$ values typically fall in the range of 15–25%.

Due to the finite precision of the atomic coordinates ('structural noise' [10]) A_a is usually underestimated by fitting procedure [11]. This can be corrected by analyzing the 'powder pattern' distribution of $D^{\rm obs}_{IM}$ [12–14] which leads to estimates of A_a and R, but not of α, β, γ. Conversely, RDCs can be transformed into frame-independent angular restraints which are free of α, β, γ but depend on A_a and R [15–17]. Structure calculation programs based on restrained Molecular Dynamics (rMD) usually optimize alignment parameters (either the 3 angles, or all 5 parameters) in parallel with structure refinement [12, 18,19]. In doing so, the alignment parameters are encoded in coordinates of pseudo-atoms; the masses of these pseudo-atoms are empirically adjusted to ensure a reasonably slow variation. Finally, in a number of situations the alignment tensors can be

determined from first principles, i.e. by modeling interactions between biomolecules and external field [20] or alignment media [21,22], which often provides a useful check.

The alignment tensor has a number of symmetry properties. Importantly, the tensor associated with a molecule or molecular fragment is invariant with respect to translation of this molecule or fragment. Furthermore, since the principal axes system (PAS) of the alignment tensor has no 'sense of direction' there are four different right-handed coordinate frames that can be associated with the PAS. As a consequence, a 180° rotation of the molecule about any of the PAS axes leaves the alignment and all RDC values invariant. While this 4-fold degeneracy is unimportant when a single alignment tensor is considered, it causes difficulties when two local alignment tensors are defined for different parts of a molecule [23] (see below).

In the case where the rhombicity parameter is zero, $R = 0$, there is an additional invariance with respect to any rotation about the z-axis of the PAS. Conversely, when $R = 2/3$ there is an invariance described by the following permutation of the PAS axes, $\{x, y, z\} \rightarrow \{x, z, -y\}$ [24]. If the molecule or complex posses a point group symmetry, it is also reflected in their alignment tensors [25]. For example, if a molecule has C_2 symmetry then one of the alignment PAS axes, x, y, or z, must lie along the symmetry axis. If furthermore $R = 0$, the unique alignment axis should be either parallel or perpendicular to the symmetry axis. The presence of C_3 or a higher order symmetry axis dictates that the alignment tensor is axially symmetric with its unique axis along the symmetry axis [25,26]. Finally, a poor sampling of orientations on $\{\theta, \phi\}$ sphere by internuclear vectors leads to difficulties in determination of the alignment tensor. For example, if all RDC data are from dipolar vectors that lie in a same plane then no more than 3 alignment parameters can be determined from the analyses of these data.

2. Distinctive properties of residual dipolar couplings invite a 'rigid body' approach

It can be noted that RDC data are *weakly coupled* to each other, in contrast to NOE data that are *strongly coupled*. To illustrate this point consider a hypothetical mistake in spectral assignment of a protein resulting in a single incorrect RDC value. This erroneous value can often be accommodated by reorienting a single peptide plane which does not affect the flanking regions and leaves RDCs from other peptide planes unchanged. Hence, one incorrect RDC value does not necessarily produce a contradiction with the rest of the data set. In contrast, one misassigned NOE is very likely to be in contradiction with a network of distance constraints based on the rest of NOEs.

In structure calculations, a trial structure is generated and tested against a set of experimental RDCs. The test begins with determination of the 5 alignment parameters (see above) followed by evaluation of $\chi_{\mathrm{RDC}} = rms(D_{IM}^{\mathrm{obs}} - D_{IM}^{\mathrm{calc}})$ which is added to the overall target function. The residual χ^2_{RDC} has two unfavorable properties:

(i) χ^2_{RDC} can be significantly reduced by introducing local changes into the trial structure (e.g., by subtle adjustment of individual peptide planes or bonds) without altering the global structure;
(ii) χ^2_{RDC} surface has no appreciable gradient toward the global minimum if the trial structure is far removed from the real structure (illustrated in Fig. 1).

The combination of these factors is detrimental for structure calculation. Consider, for example, the trial set of coordinates for a 2-domain protein where the intradomain structure is accurate while the relative orientation of the two domains is not. In the context of the rMD structure calculation, feature (i) causes unwarranted 'fluidity' of the local structure, while feature (ii) means that the program may fail to find the path toward the correct global conformation of the protein. Technically, the rMD search algorithm becomes trapped in multiple local minima on the 'rugged flatland' type of the χ^2_{RDC} surface [27] (by contrast, strong coupling between NOE data radically simplifies the χ^2_{NOE} surface so that it does not suffer as much from the features (i) and (ii)).

In order to avoid this pitfall, a 'rigid body' approach has been widely used. In this approach the local coordinates are obtained elsewhere (X-ray crystallography or NOE-based NMR structures, possibly in conjunction with homology modeling and database mining) before the RDC data are employed. The local structure so defined is treated as rigid. For instance, the internal structure of domains in a multidomain protein is 'frozen' and not allowed to change during further refinement. This approach completely eliminates problem (i). It also drastically reduces the dimensionality of the phase space by eliminating all degrees of freedom associated with intradomain structure. This, in turn, dramatically simplifies the χ^2_{RDC} surface and improves the situation with local minima. In this case, even a small gradient is enough to guide the search algorithm toward the global minimum (alternatively, a grid search can be undertaken in the low-dimensionality phase space). Hence 'rigid body' optimization approach provides a natural answer to the deficiencies of the RDC data.

3. Rigid-body optimization procedure: general scheme

The concept of rigid-body optimization in biomolecular NMR can be traced to the seminal relaxation study by Brüschweiler, Liao, and Wright [28]. While rotational diffusion tensors determined from relaxation measurements are much less accurate

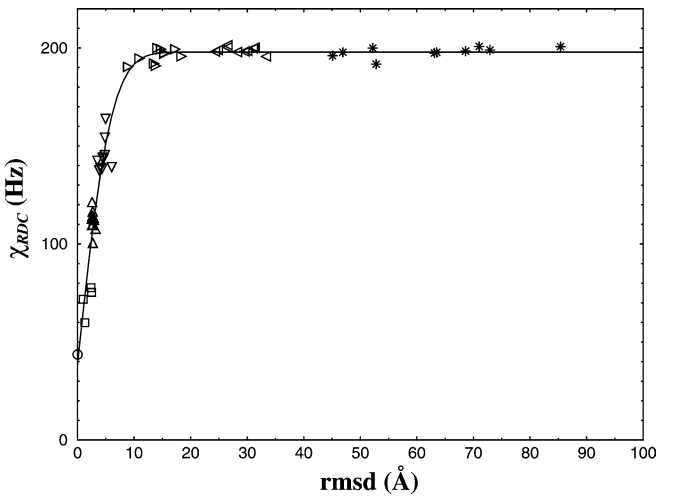

Fig. 1. Relationship between χ_{RDC} and the quality of structural model. Based on experimental RDC data from maltodextrin-binding protein [61] and a series of structures including X-ray, NOE, and random-coil structures. Reproduced from [16] with permission from Kluwer Academic Publishers.

than the alignment tensors, they have the advantage of being measured in an isotropic solution. It has been demonstrated on a number of occasions that the results of relaxation- and RDC-based procedures are consistent within their respective error margins [29–31]. The main ideas in developing rigid-body protocols for RDC studies came from Prestegard and co-workers. The list of novel concepts originating from this group includes: rigid-body treatment of α-helices [32], protein domains [33,3], multimers [25], and protein-ligand systems [34], analyses of systems with internal motions [35], use of point group symmetry in determination of the alignment tensor [25], use of multiple alignment media for resolving orientational degeneracy [23], and use of artificial restraints in rMD implementation of rigid-body protocol [36]. It is worth noting that conventional structure calculation protocols usually involve a 'soft' version of the rigid-body approach where at some stage during the calculations certain experimental and generic restraints are strengthened with the goal to prevent distortions of local geometry and to facilitate global conformational changes [37]. In this review, however, we discuss only the methods where essential rigid-body approximation is made.

A number of different implementations of the RDC rigid-body algorithm have been developed and tested. The method proved to work well irrespective of the details of implementation [38,39]. However, in many practical situations where the number of experimentally measured RDCs is small and/or the accuracy of underlying coordinates is problematic the design of the computational procedure becomes important. In addition, there are a number of important considerations that must be taken into account when interpreting the results of the RDC rigid-body procedure that have been often overlooked in the past. The scheme of the rigid-body procedure that is outlined below is an attempt to summarize the experience accumulated by many research groups in this field.

As a first step, an underlying set of coordinates must be chosen. X-ray crystallography is the primary source of well-defined 'rigid' structures [40]. In fact, high-resolution X-ray structures are so accurate that they can only marginally benefit from RDC refinement [7] and, therefore, can be used as is (following the standard protonation step). Usually, the choice of 'rigid' structural elements is evident, as in the case of protein domains or monomeric units in a multimer. In situations where the choice is not obvious, RDC consistency maps [16] can be used to identify fragments of the protein structure that are internally self-consistent with respect to experimental RDC data, but inconsistent with other fragments. The definition of a rigid fragment may vary: for example, it can extend only to the main chain of the protein, with side chains allowed full flexibility [41].

In the next step, the RDC data are analyzed against the coordinates of the selected fragments. Five alignment parameters, A_a, R, α, β, and γ, are fitted and the *rms* deviation between the experimental and best-fit RDC values, χ^2_{RDC}, is determined. Importantly, this step tests the consistency between the model coordinates and the experimentally measured RDCs which can be quantified using a quality factor based on χ^2_{RDC}. The requirements for a comprehensive quality factor, hereafter called Z_{dip}, can be summarized as follows. Firstly, Z_{dip} should not depend on the strength of the alignment. Quality factor R_{dip}, Eq. (3), satisfies this requirement. In addition, Z_{dip} should reflect the sampling of different orientations by dipolar vectors in the underlying structure (note that good R_{dip} score alone has little significance if all dipolar vectors are near-parallel). A measure for orientation sampling has been suggested by Fushman et al. [42]. Finally, Z_{dip} should also take into account a size of RDC data set (note that good R_{dip} score has little significance if, for example, only 6 couplings have been measured).

At this stage it is important to carefully analyze the agreement between the alignment parameters A_a and R obtained from the different fragments. Specifically, it is important to determine the margin of error for both parameters. A primary source of error is the 'structural noise' which reflects the limited accuracy of the underlying atomic coordinates. Other contributions arise from experimental errors in RDC measurements and from local dynamics. A good handle on all sources of error can be obtained by means of comprehensive Monte Carlo simulations [11]. After A_a and R are evaluated together with their associated margins of uncertainty, it is necessary to compare the values obtained for different fragments. If A_a and R from different regions are in agreement within their respective error margins, it is concluded that the entire molecular entity aligns as a rigid unit and that a single set of A_a and R can be used to represent the magnitude of the alignment. Discussion of the alternative scenario, where so-called 'differential alignment' is observed, is deferred until later.

In fact, the situation can be more complex than described above. Specifically, good agreement between A_a and R from different fragments does not necessarily indicate the absence of internal motion in the system. Consider, for example, disulphide-bonded symmetric dimer. Although there may be a considerable amount of motion in this system, with one monomeric unit moving with respect to the other, the symmetry dictates that A_a and R are the same in both units. Likewise, in two-domain protein similar A_a, R values can be expected from two domains even in presence of interdomain motion provided that the mechanism of alignment is steric and the two domains are similar in size [10]. The same argument can be made for electrostatic alignment so long as the distribution of charges on the surface of the two domains is similar. Therefore, generally speaking, similarity in A_a and R values from different fragments is not sufficient to claim that the system aligns as a rigid entity. However, the presence of near-symmetry or symmetry in this context (such as in the case of disulphide-bonded dimer) means that RDC couplings can accurately describe the average conformation in such systems. Thus, the structures generated using rigid-body approach should be often viewed as *average* structures of dynamic entities, rather than static constructs. With this caveat in mind, we continue the discussion of the algorithm.

The next stage is the core of the algorithm. At this point, a number of trial structures are generated by reorienting the rigid fragments. Each trial structure is tested to assess its agreement with complete set of RDC data assuming a single set of alignment parameters. The proposed algorithm is, essentially, a nested optimization: the outer loop searches for the optimal conformation by reorienting the fragments, while the inner loop fits 5 alignment parameters based on the current set of atomic coordinates and returns the χ^2_{RDC} value. In the end, the structure that provides the best match for the complete set of RDC data is identified. This procedure is best implemented using the restrained molecular dynamics platforms such as XPLOR [43], Discover [44], and DYANA [45]. The advantages of rMD in this context are two-fold: first, a realistic set of coordinates is produced which includes the linkers between the rigid fragments; second, the false solutions arising from four-fold degeneracy in RDC-derived orientations are automatically discarded as they typically cause steric clashes, NOE violations, etc. If rMD implementation is chosen, the rigidity of selected molecular fragments is maintained throughout the course of simulations by means of tight artificial restraints such as distance, angle, or crystallographic symmetry restraints. The computational load for this hypothetical procedure remains relatively low since the alignment tensor can be rapidly fitted using the singular value decomposition procedure [3] and molecular dynamics trajectory can be generated very efficiently when a large fraction of torsional angles are held fixed [46,47]. After the set of low-energy structures is generated in this fashion, the quality factor Z_{dip} should be calculated. A properly applied rigid-body procedure should lead to Z_{dip} value similar to the values previously determined for individual fragments if one assumes that the quality of the final structure is limited mainly by finite resolution of the underlying X-ray coordinates [48]. Note that a single set of A_a, R is employed at this stage. This is, generally speaking, preferable to a popular approach where separate alignment tensors are determined for individual fragments and the overall structure is subsequently reconstructed by matching the respective alignment frames. So long as the alignment parameters from different fragments agree within their respective error margins (see previous step), there is no need to retain multiple versions of A_a and R as they effectively become redundant fitting parameters.

At the final stage, it is recommended that artificial constraints be removed and the structure refined in order to relieve a small degree of strain that may have been introduced during the rigid-body optimization (for example, at points where the flexible linker is attached to the rigid fragment). This step does not lead to any significant changes in the structure and should be viewed as optional. However, if the coordinates of rigid fragments are based on a low-resolution structure, such as majority of NOE structures, then the refinement procedure using RDC data becomes important. At this stage RDC data should be interpreted strictly as the source of local structural information and any global changes should be disallowed (see recent work by McCallum and Pardi [49] for insightful discussion).

Finally, we shall briefly discuss the alternative scenario when the alignment parameters A_a and R evaluated for the different fragments do not agree. Before the rigid-body approach can be applied in this situation, it is necessary to choose a motional model. In particular, the popular 'diffusion in a cone' model [50] for the relative motion of two fragments leads to simple rescaling of A_a and permits the use of rigid-body treatment. The structure calculated with a rigid-body approach should be interpreted in this case as a bona fide average structure.

In what follows, we present several selected examples illustrating the applications of the RDC rigid-body approach to various biomolecular systems.

4. Applications of rigid-body approach

4.1. Small backbone fragments

Elementary structural units such as peptide planes or nucleotide bases are essentially rigid and can be reoriented using the RDC-driven rigid-body approach [33,51–53]. These elements, however, are automatically treated as rigid or near-rigid in any structure calculation protocol and, therefore, do not require any specialized approach. Larger fragments consisting of several peptide units were used by Delaglio, Contaxis and Bax [54] to reconstruct the fold of the small protein ubiquitin. The distinction of this method is that the structures of 7-residue rigid fragments that are used to assemble the protein are not known a priori and are selected instead from the pool of 350 000 fragments from Protein Data Bank. The selection is made on the basis of the lowest χ^2_{RDC} score using the set of experimentally measured RDCs.

The program developed by Delaglio et al. uses a 'sliding window' principle as it consecutively finds structural models for residues 1–7, 2–8, etc. The results of the search are stored as a sequence of $\{\phi, \psi\}$ angles that defines the entire backbone fold. The fact that a series of dihedral angles in the 1–7 stretch is consistent with the series of angles in 2–8 reflects the intrinsic rigid-body character of this algorithm: the peptides are effectively placed in a common alignment frame when RDC fitting is performed.

Dipolar couplings from a given 7-residue fragment can often be fitted equally well using coordinates from a number of peptides. As a result, the process generates multiple trajectories in $\{\phi, \psi\}$ space. Most of the trajectories are clustered around the path which corresponds to the correct protein fold. In some cases, however, ambiguities have been encountered. To understand the origin of these ambiguities consider, for example, a peptide that faithfully reproduces the structure of Gly10–Glu16 fragment from ubiquitin. Assume further that one of the principal axes of the alignment tensor accidentally coincides with $\mathrm{N}\text{–}\mathrm{C}^\alpha$ bond in Thr14. In this situation the model peptide can be transformed by performing a 180° rotation about the said bond ($\phi \to \phi + 180^\circ$) causing reorientation of Leu15 and Glu16. The transformed peptide has a different topology, but satisfies the set of experimental dipolar couplings equally well compared to the original peptide. The resulting ambiguity ultimately limits the efficiency of this method.

Using an extensive set of data (the total of 542 NH^{N}, CN, CH^{N}, and $\mathrm{C}^\alpha\mathrm{H}^\alpha$ couplings from measurements in two different bicelle media) the backbone fold of the entire small protein, ubiquitin, was determined by Delaglio and co-workers with an accuracy of 7 Å. The calculation involved a certain amount of serendipity since dihedral angles in one of the residues were highly ambiguous. Although very crude, this protein fold can be efficiently refined using RDC data (cf. Fig. 1). Using a specially written refinement procedure operating on torsional angles, the authors were able to refine the structure from 7 to 1.2 Å.

A similar method has been independently developed by Andrec, Du, and Levy [55]. Rohl and Baker [56] incorporated this approach into the structure prediction program Rosetta that operates on protein fragments obtained from structural databases. Finally, Wang, Zuiderweg, and co-workers [57] considered the local alignment tensors for 5-peptide fragments with the aim to characterize local dynamics.

4.2. RNA helices

Mollova, Hansen and Pardi [58] applied a rigid-body reorientation algorithm to *E.coli* $\mathrm{tRNA}^{\mathrm{Val}}$. Two arms of the L-shaped tRNA were treated as rigid structural units. Since no high-resolution structure is available for *E.coli* $\mathrm{tRNA}^{\mathrm{Val}}$, the initial set of coordinates was produced by making base replacements in the X-ray structure of the related yeast tRNA. Based on this set of coordinates and 24 NH couplings measured in Pf1 bacteriophage media, the alignment tensors were determined for the two arms of tRNA leading to very similar A_a and R values. Rigid-body reorientation was subsequently performed on one of the arms using the previously described algorithm [10]. The resulting arrangement was found to be liable to the standard 4-fold degeneracy. However, two possible conformations could not be accommodated using reasonable linker geometry, while the third contradicted experimental NOE data. The remaining solution was close to the starting X-ray conformation as characterized by a 13° difference in the inter-arm angle. Furthermore, a similar result, 15°, was obtained when the stem regions in the starting structure were replaced with canonic A-form helices. This latter result indicates that RDC data can be used to determine the global conformation of nucleic acids when little or no prior structural information is available.

An improved version of this approach was applied by Bondensgaard, Mollova and Pardi [39] to a 35-nucleotide hammerhead ribozyme (complexed with a 13-nucleotide stabilizing substrate). The hammerhead ribozyme motif consists of three helical stems converging on a flexible region known as catalytic domain. A subset of data which was used to determine relative orientations of the three stems consisted of 17, 23, and 11 values (NH and CH couplings) for stems I, II, and III, respectively. As a first step, the alignment parameters A_a and R were carefully analyzed on per stem basis using a set of suitably modified X-ray structures as well as the canonic A-form helices. It was found that the alignment parameters for all stems are identical within the error bounds, with errors mainly reflecting 'structural noise' in the underlying atomic coordinates and limited accuracy of the measurements. It was concluded that RDC data do not present any evidence for mobility of the stems, although such mobility

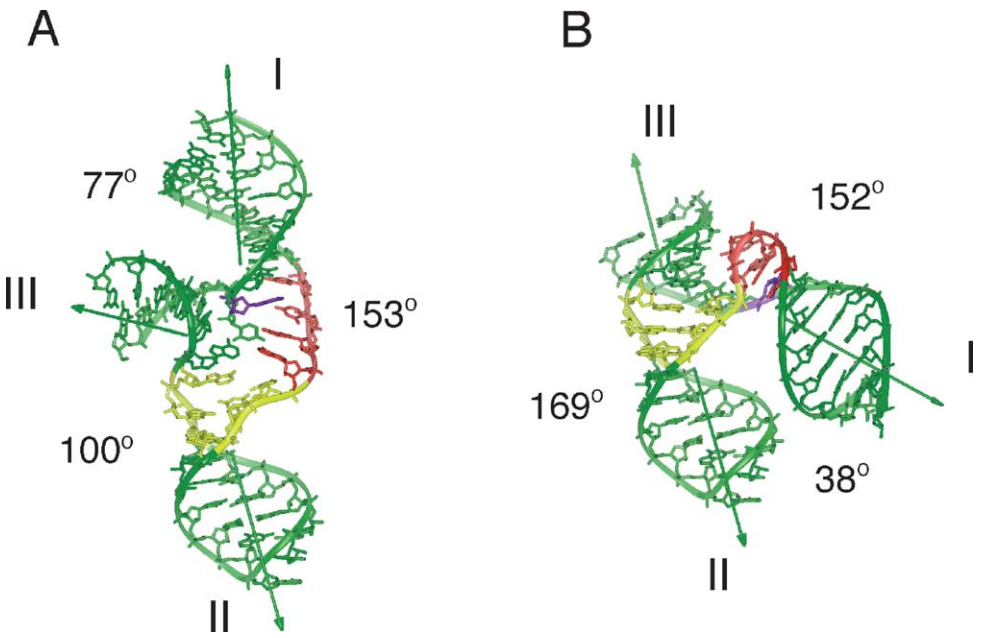

Fig. 2. (A) Solution conformation of hammerhead ribozyme in the absence of Mg^{2+} obtained with RDC data. (B) X-ray structure (obtained in the presence of Mg^{2+}; a very similar conformation has been observed in the absence of Mg^{2+} at high ionic strength). Reproduced with permission from [39]. Copyright 2002 American Chemical Society.

has been suspected on the basis of other data. (In this regard, it is important to keep in mind that similar A_a, R values can be observed even in dynamic situation provided that the amplitudes of motion are similar between the stems.)

After establishing that the alignment in the hammerhead ribozyme is controlled by a single effective tensor, Bondensgaard and co-authors applied the rigid-body reorientation procedure. In doing so, the orientation of the stem II was fixed, and the two other stems were reoriented. The orientation of each helical stem was described by three Euler angles: two polar angles specifying the orientation of helical axis, plus one angle specifying the twist about this axis. Given the 4-fold degeneracy in the orientations of stems I and III, a total of 16 potential solutions was obtained in this manner. Thus, the first step was the standard rigid-body procedure that was applied to isolated stems and led to ambiguous solutions.

The output from this first step was then subjected to a (semi)rigid-body refinement procedure implemented in XPLOR-NIH [46]. At this stage, the starting structures were prepared where the relative orientation of the stems was the same as derived in the previous step. The correct orientation of the stems was maintained throughout the simulated annealing procedure by means of artificial angular constraints (in fact, a set of artificial RDC data was used for this purpose). Furthermore, the local structure of the stems was also 'frozen'. In contrast, the junction region was treated as fully flexible. Initially, the coordinates of the junction region were defined by generating a set of random values for relevant torsional angles. The XPLOR procedure was then used to optimize the structure of the junction region in accordance with the pre-set orientation of the stems and a number of conventional restraints (NOEs, J-couplings). Interestingly, this step serves the dual purpose of (i) reconstructing the junction region and refining the structure and (ii) weeding out the 15 false conformations generated in the first step. Of note, all false conformations were easily identified as they showed large number of violations and high energies.

The final structure obtained as a result of this two-step procedure turned out to be dramatically different from the available X-ray structures as illustrated in Fig. 2. For example, the angle made by stems I and II changes between the two structures from 38° to 153°(these findings are supported by data from FRET and other solution-state techniques). The uncertainty in the determined inter-stem angles is estimated to be approximately 10°.

Most recently, McCallum and Pardi [49] investigated the structure of iron-responsive element RNA. This 29-nucleotide motif consists of two helical stems separated by one bulging nucleotide and capped by a six-nucleotide loop. A total of 126 CH, 10 NH, and 74 CC couplings were measured in this system. A starting set of coordinates was obtained from the NOE-based ensemble of structures where the stem regions were defined with a medium precision, 1.2–1.3 Å, while the relative orientation of the two stems was defined very poorly, 3.0 Å (see panel a in Fig. 3).

As a first step, a careful analysis of A_a and R values was undertaken by the authors. The initial estimates for A_a and R in the two stem regions were obtained by direct fitting to the experimental RDC data. The quality of the fits was poor, $R_{\mathrm{dip}} = 30$–45%, betraying the low quality of the underlying NOE-based coordinates. Similar results were obtained using the coordinates

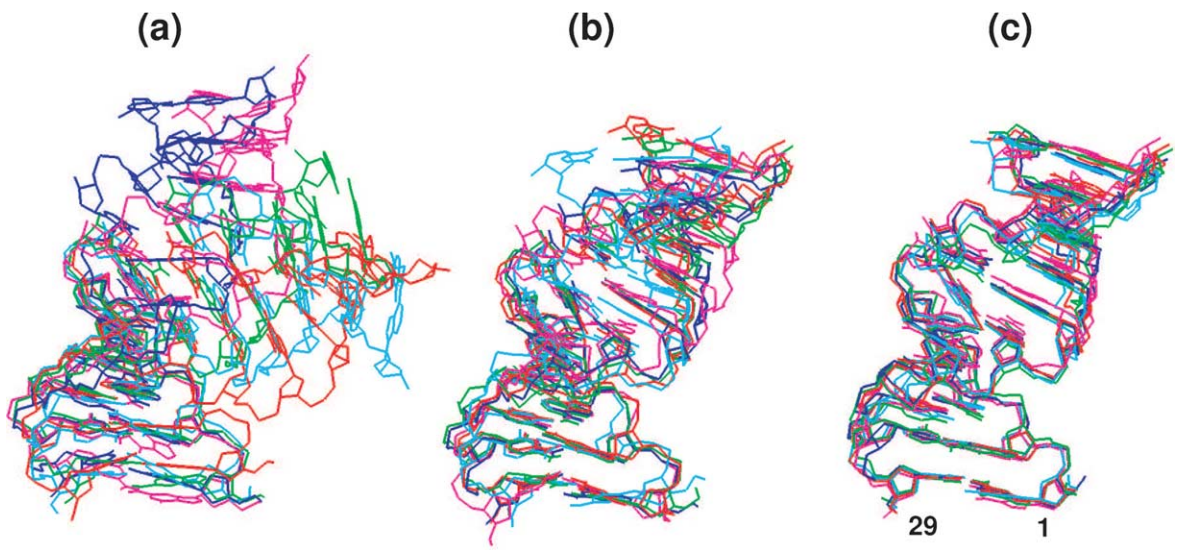

Fig. 3. Structures of the iron-responsive element RNA: (a) NOE-based coordinates; (b) NOE-based coordinates following rigid-body optimization with RDCs; (c) NOE-based coordinates following rigid-body optimization and structure refinement with RDCs. Reprinted with permission from [49]. Copyright 2003, Elsevier.

of canonical A-form helices. Since the presence of 'structural noise' leads to underestimation of A_a, the authors corrected for this effect by rescaling A_a using the factor $[(1/2)\cos\sigma(1+\cos\sigma)]^{-1}$, where σ is the scatter in orientations of dipolar vectors among the members of the NOE ensemble, $\sigma = 18°$. This approach implies that dipolar vectors from the NOE ensemble are randomly distributed within 18° of their true orientations. The corrected A_a values were in good agreement with the estimates obtained from the histogram approach [14]. In this connection, the authors point out that underestimated A_a values can be problematic for structure calculations since they lead to high residual χ^2_{RDC} values. In order to account for the uncertainty in magnitude of alignment the authors expanded the error bounds on the experimental RDC data. Importantly, it has been observed that A_a and R values from the two stems are in good agreement, clearing the way for the rigid-body treatment.

In the next stage, the NOE-based coordinates were processed using a two-step protocol implemented in the program XPLOR: (i) (semi)rigid-body treatment whereby the relative orientation of the stems was corrected; and (ii) refinement of the local geometry within the stems. Both steps made use of the RDC information and the values of A_a and R were held fixed throughout the treatment. In the first step, a single alignment tensor was employed. The local geometry was constrained using synthetic data which consisted of all proton–proton distances in the range from 3 to 11 Å. In the second step, individual alignment tensors were assigned to both stems to prevent distortion of the overall conformation. The alignment parameters A_a and R for the two stems were set equal. As a result of this procedure, the precision of the structure dramatically improved: 0.6–0.7 Å in the stem regions, 1.6 Å overall. The angle between the two helical stems, which was previously very poorly defined, was found to be ca. 35° (see Fig. 3). It is important to note that the results did not change when the two steps were executed in the reverse order: local refinement (ii) followed by a rigid-body treatment (i).

The algorithm ends with gentle overall refinement including RDC data from the loop and bulge regions. This final refinement had practically no effect on the structure and relative orientation of the stems, but revealed the presence of dynamic disorder in the loop region where the experimental RDC constants could not be fitted well. Furthermore, the bulge region was shown to adopt two different conformations, both compatible with the observed RDC values. While this behavior merely reflects the 4-fold degeneracy in RDC-derived orientations, the relaxation data point toward μs-ms motion in the bulge region and hint at the possibility of inter-stem motion.

In several cases RDC data put into evidence the dynamic nature of RNA structures. Al-Hashimi et al. [59] investigated a 27-nucleotide fragment of the RNA from human immunodeficiency virus HIV-1. This motif consists of two helical stems, the bulge which plays the role of the joint, and the capping tetraloop. A sample of ^{15}N, ^{13}C-labeled RNA was aligned in Pf1 bacteriophage media and a total of 18 and 22 carbon–proton and carbon–carbon RDCs were measured in stem I and stem II, respectively. Helical stems were modeled assuming idealized A-form geometry or, alternatively, using the previously reported NOE-based coordinates. Of interest, the authors observed ca. 40% difference in the alignment A_a between the two stems. At the same time, the rhombicity parameters R turned out to be similar (taking error margins into consideration). This situation is compatible with 'diffusion in a cone' model [50] if one assumes that the alignment arises entirely from stem II (through transient interaction with phage particles) and the motion of stem I relative to stem II causes uniform attenuation of the couplings in the

former. In the context of this model, the observed difference in A_a corresponds to stem I moving in the cone with half-angle of opening $\delta_{1/2} = 46^\circ$. By comparison, the amplitude of fluctuations observed in a MD trajectory reaches 50°. Note that in a more realistic model where stem I also has a limited ability to align, the motion is described by $\delta_{1/2} > 46^\circ$.

The problem was further examined in recent work by Zhang, Al-Hashimi and others [60]. The RDC data were recorded in this study using direct field-induced alignment which is not dependent on details of molecular interaction between the RNA and aligned phage particles. Notably, there is no difference in A_a between the two stems. This is expected for helical stems of similar size since the magnitude of field-induced alignment is determined in this case by the number of stacked aromatic bases. However, it turns out that the experimentally determined A_a value is approximately 45% lower than the value predicted from calculations of the magnetic susceptibility tensor [20] using the RDC-optimized structural model. Since the calculations of magnetic susceptibility tensors are known to be reliable and, in this case, correctly predict the orientation of alignment frame and the magnitude of rhombicity, R, the discrepancy should be attributed to interdomain dynamics. To rationalize these findings the authors considered a model where two domains contribute equally to the alignment and the alignment tensors of both domains are axially symmetric. In this situation 'diffusion in a cone' leads to uniform attenuation of all couplings according to the formula $(1/2)\cos(\delta_{1/2}/2)(1 + \cos(\delta_{1/2}/2))$. The amplitude of the motion predicted with this model is very large, $\delta_{1/2} = 97^\circ$. Note that it does not necessarily contradict the previous estimate, $\delta_{1/2} > 46^\circ$.

4.3. Protein domains

Skrynnikov, Kay and co-workers [10] used RDC data to investigate conformation of two-domain 379-residue maltodextrin-binding protein loaded with β-cyclodextrin. The extensive set of RDCs, including 280 NH^N, 262 NC, 276 $C^\alpha C$, 262 CH^N, and 276 $C^\alpha H^N$ couplings, was previously recorded for this protein in Pf1 phage media by Yang, Kay and others [61]. Starting coordinates for the rigid-body reorientation routine were obtained from several X-ray structures where the protein was crystallized either in the apo state or with one of three different ligands, including β-cyclodextrin. Four different crystal-state conformations were parametrized in terms of closure, twist, and bend. The reference structure of free maltodextrin-binding protein was assigned the values of closure, twist, and bend of $\{0^\circ, 0^\circ, 0^\circ\}$. The most closed structure observed in complex with maltose was described by the angles of $\{36^\circ, -4^\circ, -3^\circ\}$. The analyses of the X-ray structures suggest that domain closure in maltodextrin-binding protein occurs as a 'hinge' rotation, with the 'hinge' localized in the linker region.

The RDC data were first analyzed on per-domain basis leading to nearly identical A_a and R values. A rigid-body algorithm was implemented as a simplex search in the 8-parameter space, where three angles, Θ, Φ, and ω, defined the orientation of the N-terminal domain with respect to the C-terminal domain, and the five remaining parameters represented the alignment tensor. The convergence of the search proved to be excellent, and the results were identical to those obtained with the more conventional procedure involving superposition of the alignment frames from the two domains. The resulting conformation turned out to be 11° more closed compared to its X-ray counterpart (false solutions corresponding to large conformational changes were discarded). The difference between solution- and crystal-state conformations was attributed to the effect of crystal-packing forces. An approximate model of the protein, minus the linker region, has been built by assuming the same 'hinge' location as in the X-ray structures. The bound β-cyclodextrin molecule was comfortably accommodated in this model. Furthermore, the results of this work were confirmed by a ^{15}N-relaxation study by Hwang, Skrynnikov and Kay [29] where the degree of closure in isotropic solution was determined to be $12^\circ \pm 4^\circ$.

Depending on the set of starting coordinates, the degree of closure determined in the RDC study varied within $\pm 1^\circ$. The most significant uncertainty, $\pm 3^\circ$, was associated with the twist component. Relatively low precision with respect to twist can be explained by noticing that the twist axis coincides with the long axis of the alignment tensor which happens to be nearly axially symmetric ($R = 0.17$). The errors in conformational analyses were attributed to "structural noise" and this conclusion was confirmed by simulations. In contrast, the measurement errors proved to be a minor source of uncertainty. The authors also considered three hypothetical dynamic scenarios. In the first model it was assumed that the protein exists in dynamic equilibrium between the open and the closed forms. In the second model a distribution of the species was considered (assuming that the alignment tensor is unaffected by small-amplitude opening/closure). The third model addressed a situation where the alignment mainly originates from the C-terminal domain and relative motion of the two domains leads to modulation of RDCs in the N-terminal domain (cf. the discussion of 'differential alignment' in the end of this section). In all three scenarios it was found that the simple rigid-body approach adequately describes the mean structure of maltodextrin-binding protein.

This study was taken further in the work of Evenas, Kay and others [62]. In addition to the complex with β-cyclodextrin, extensive RDC data were recorded for maltotriose-bound and apo forms of the protein. The data were edited to remove the residues with high crystallographic B-factors, resulting in the sets of 725, 818, and 959 couplings for β-cyclodextrin-bound, maltotriose-bound, and apo states, respectively (the latter data set additionally included $C^\alpha C^\beta$ couplings). The (semi)rigid-body reorientation procedure was implemented using the program CNS supplied with the RDC module [12]. During the rMD simulations A_a and R were held fixed according to the best-fit values obtained for individual domains. The internal structure of the two domains was preserved by imposing artificial restraints on H^N–H^N and N–N distances and on ϕ, ψ backbone dihedral

angles. The mobility of the side chains was not restricted. The experimental input consisted of the RDC data from both domains and the linker region and, optionally, from the chemical shift data interpreted by TALOS [63].

The previously observed difference between solution and crystalline states of maltodextrin-binding protein in complex with β-cyclodextrin was confirmed in this study. Importantly, the structure of the entire protein complete with the linker region was obtained by Evenas et al. in this work. The linker region in maltodextrin-binding protein consists of three short backbone strands and does not allow for a significant amount of translational motion (stretch and shear). Therefore, the relative placement of the two domains, including the translational component, is well defined by RDC-based calculations. The quality of the structures generated in this fashion was assessed by the program PROCHECK [64] and found to be at least as good as the quality of the underlying crystallographic coordinates.

The most important result of this work is that the determined solution-state conformations of the free protein (open form) and maltotriose-bound protein (closed form) are in excellent agreement with their crystalline-state counterparts. In the former case, the differences in terms of closure, bend, and twist angles did not exceed 2°, while in the latter case the differences were within 1°. These findings help to put the results of the RDC-based rigid-body method in proper perspective. As a rule it can be expected that RDC-based results should agree with X-ray crystallographic data. However, in many cases – perhaps, more than previously realized – significant variations can be encountered due to the differences between solution and crystalline phases.

Very recently Millet, Hudson and Kay [65] investigated the effect of site-directed mutations on the conformation of maltodextrin-binding protein. A series of mutations involving two residues from the linker region were previously shown to increase binding affinity of the protein for maltose. A total of 5 mutants were constructed: I329C, I329W, I329F, I329C*, and I329W/A96W (C* denotes a non-natural side chain obtained by addition of thiol-reactive compound). Using Pf1 phage media the authors measured from 141 to 181 NH^N couplings per mutant. The (semi)rigid-body structure calculation protocol was identical to that used by Evenas et al. [62]; each calculation was repeated with 12 different sets of starting coordinates corresponding to different crystallographic structures. A series of progressively more closed conformations was found in this manner for the mutants under investigation as indicated by the amplitude of closure relative to the wild-type protein: 5°, 10°, 10°, 21°, and 28° (see Fig. 4). Errors in closure angles were estimated from calculations using multiple starting structures and found to be on the order of 1°–2°. Remarkably, it has been found that the degree of closure very well correlates with the stability of the mutants. This was rationalized by calculating the non-polar solvent-exposed surface area in the obtained five structures, plus the wild-type structure determined by Evenas [62]. It turned out that closure leads to increased exposure of hydrophobic side chains on the opposite side from the binding cleft which, in turn, destabilizes the unligated protein.

The study by Millet et al. [65] proved that the apo-form of the wild-type maltodextrin-binding protein is confined to an energetically favorable open conformation. In contrast, there is ample evidence that the 164-residue two-domain protein bacteriophage T4 lysozyme samples both open and closed conformations in solution. The extensive library of crystallographic structures includes the species that differ by 50° closure along with a substantial amount of bend. The apo-form of T4 lysozyme is typically crystallized in the closed form which can be directly linked to the effect of crystal packing forces (upon crystallization the protein forms 'back-to-back' dimers with concomitant burial of a large hydrophobic surface). The mean conformation of T4 lysozyme in solution was investigated by Goto et al. [38]. The researchers recorded approximately 120 NH^N couplings using a sample of T4 lysozyme aligned in solution of cetylpyridinium chloride with hexanol. In addition, approximately 60 NH^N couplings per sample were collected using three samples aligned in bicelle media with different ionic strengths and lipid concentrations. Starting coordinates for the analyses were obtained from 9 different crystallographic structures including open and closed forms (all structures had a resolution of 2 Å or better). The experimental RDC data were

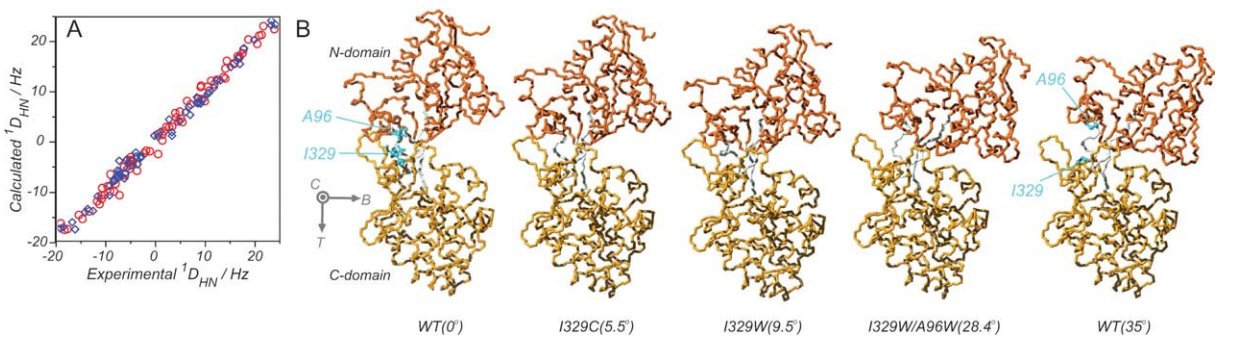

Fig. 4. (A) Correlation between the experimental and best-fit values of NH^N couplings using the optimized structure of maltodextrin-binding protein. (B) Structures of maltodextrin-binding protein obtained with RDC rigid-body procedure: wild-type apo state, I329C, I329W, I329W/A96W, and wild-type from the complex with maltotriose. Closure, twist, and bend axes indicated by the arrows. Reprinted with permission from [65]. Copyright 2003 National Academy of Sciences, USA.

first fitted to atomic coordinates on a per domain basis leading to very similar A_a values between the N- and C-terminal domains (it was demonstrated that the observed differences of less than 10% could be attributed to the effects of 'structural noise').

The previously described rigid-body protocol [10] was employed to interpret the data. The determined solution conformation turned out to be 17° more open than its crystalline counterpart. Notably, the results obtained from different alignment media proved to be consistent as evidenced by the determined closure angles: 18° ± 2°, 15° ± 4°, 19° ± 2°, and 15° ± 4°. This important finding demonstrates that the determined unique conformation does not depend on the type of alignment media and, therefore, is likely to represent the true conformation of protein in aqueous solution. The half-open conformation obtained in this study was found to be similar to the crystallographic conformation of the T4 lysozyme mutant M6I. As a nice consistency check, it was demonstrated that RDC values predicted by the program SSIA [21] using M6I coordinates are in excellent agreement with the experimentally measured couplings, $R_{\rm dip} = 23\%$.

The sizeable errors in the extracted closure angles reflect the influence of 'structural noise' in the situation where relatively few RDCs are available (e.g., 23 couplings in N-terminal domain). As in the case of maltodextrin-binding protein, the twist component was determined with relatively poor precision since the long axis of the alignment tensor was approximately parallel to the twist axis and the rhombicity parameters in all alignment media were small, $R = 0.13$–0.22. The analysis of multiple solutions arising from 4-fold degeneracy in the RDC-derived orientations was carried out according to the recipe of Al-Hashimi et al. [23]. In brief, the interpretation of data from two different alignment media led to two sets of solutions, each consisting of four possible conformations. Only one solution, which corresponded to the true conformation of the protein, was shared between the two sets. Goto demonstrated that this method was applicable even when the principal axes systems of the two alignment tensors were nearly identical (4° difference in the orientation of the z-axes).

As established in the preceding paper [10], even in the presence of interdomain motion RDC analyses lead to a well-defined average conformation. Along these lines, Goto and co-workers considered a more realistic two-state exchange model where the alignment tensors for the open and closed species were computed using the program SSIA [21]. Using a single fitting parameter, $p_{\rm open} = 1 - p_{\rm closed}$, the authors demonstrated that the RDC data point toward $p_{\rm open} \approx p_{\rm closed} \approx 50\%$, consistent with the average half-open conformation described above.

In the second part of the same paper [38], the authors studied a mutated variant of T4 lysozyme where the two domains were additionally linked via a disulphide bond. Measurements using three different alignment media produced from 22 to 32 couplings per sample per domain. In contrast to the wild-type study, the solution conformation derived from the RDC data proved to be by 11° more closed than expected from crystallographic data. Once again, the results from different alignment media were consistent within the error margins. A crude model of the protein structure was built by rotating the N-terminal domain as dictated by the RDC data with position of the hinge chosen according to the X-ray crystallographic coordinates. The resulting structural model, however, was not entirely satisfactory since the predicted C^α–C^α distance across the disulphide bond, 2.9–4.4 Å, was abnormally short. This deficiency can be readily explained by noticing that the location of the hinge associated with the extended linker region can be only approximately estimated from the analyses of two crystallographic structures. Furthermore, hinge model does not account for possible translation of one domain with respect to another along the direction of the hinge axis. In order to correct for these deficiencies, a more advanced (semi)rigid-body protocol [62] has been used. The resulting structure was acceptable as characterized by a reasonable 4.8–5.5 Å C^α–C^α distance across the disulphide bond. With respect to orientational degrees of freedom – closure, twist, and bend – the results of the rigid-body and (semi)rigid-body protocols were in full agreement.

A number of studies employing a RDC-based rigid-body approach produced direct evidence of domain mobility in multidomain proteins. For example, Jacobs et al. [24] observed differential alignment of two domains in peptidyl-prolyl cis/trans isomerase. This 18.4 kDa protein consists of a C-terminal catalytic domain and N-terminal WW domain connected via a very flexible 12-residue linker. The fitting of RDC data to the crystallographic structures of individual domains resulted in A_a values that differed between domains by a factor of 3.5. The difference was observed in both alignment media used in this study: bacteriophage and a mixture of alkyl-poly(ethylene glycol) and hexanol. The latter medium is uncharged so that alignment in this case is expected to be purely steric in nature. This leads the authors to suggest that the alignment mainly stems from the catalytic domain which is three times bigger in size. The smaller WW domain tethered by a long and flexible linker can move around causing partial dynamic averaging of its RDCs (in an exaggerated view, the small domain can be thought of as mobile side chain that does not significantly contribute to the overall alignment). In this highly dynamic situation, the authors do not attempt the calculation of an average conformation.

In addition to the apo form of the protein, Jacobs et al. investigated several peptide-bound forms, including the complex with phosphopeptide WFYpSPR which was previously identified as the optimal model substrate for this protein. It has been found that the protein 'rigidifies' upon binding this peptide as indicated by the convergence of A_a values from the two domains (difference of less than 40% in the ligated state) and confirmed by the results of relaxation studies. This level of differential alignment corresponds to an amplitude of interdomain motion $\delta_{1/2} = 45°$. The authors proceeded to determine the average conformation of the protein in complex with the peptide by means of a standard RDC-based rigid-body approach. The obtained average conformation turned out to be in agreement with existing X-ray crystallographic coordinates.

4.4. Ligand binding

In a massive effort, Koenig et al. [66] investigated the structure of an undecapeptide mimicking the C-terminus of the α-subunit of G protein transducin bound to the photoactivated form of rhodopsin. In this study, rhodopsin is embedded in the oblate-shaped disk membrane. Due to high content of helical rhodopsin, the disk membrane becomes weakly oriented in the external magnetic field via additive susceptibility effects. A large excess of ^{13}C-, ^{15}N-labeled peptide was added to the solution. The peptide binds weakly to rhodopsin establishing a dynamic equilibrium with fast exchange between the free and bound states. Considering the two entities, the oblate disk membrane with embedded rhodopsin and peptide, two respective alignment tensors can be determined. For the disk membrane, the alignment axis is parallel to the long axis of rhodopsin and, therefore, perpendicular to the surface of the disk. For the bound peptide, the orientation of the alignment axis was determined from the experimentally measured RDC data. According to the dogma of the rigid-body approach, the orientation of the peptide relative to the surface of the disk membrane can be obtained by matching their respective alignment axes.

The RDC splittings observed in the spectrum of the peptide represent population-weighted averages of the free and bound phases. The RDCs in the bound phase were extracted from the observed splittings taking into account the difference in relaxation rates between the two states. The resulting data set consisted of 9 NH^N, 9 backbone $C^\alpha H^\alpha$, and 20 side-chain CH couplings measured at two different temperatures. These data, combined with transferred NOEs and some other input information, were used to calculate the structure of the peptide using a standard NIH-XPLOR [46] protocol with some suitable modifications. Note that at this stage RDC data were used to refine the internal geometry of the peptide. The orientation of the long axis of the alignment tensor was optimized concurrently with the structure optimization. The subsequent (implicit) rigid-body step matches the alignment axis of the peptide with the alignment axis of the rhodopsin-bearing disk membrane. At this stage it was established that the peptide binds at the angle of $50^\circ \pm 4^\circ$ to the surface of the membrane. As a consequence of the 4-fold degeneracy in the RDC-derived orientations, it was impossible to determine which terminus of the peptide binds to rhodopsin. This ambiguity was resolved, however, by observing the attenuation of peptide resonances which can be attributed to cross-relaxation between the peptide and rhodopsin and that mainly occurs at the C-terminus of the peptide. These results were further used by Koenig and co-workers to build a tentative model for the complex of the heterotrimeric G protein transducin with the active form of rhodopsin. In doing so, the crystallographic structure of transducin was docked to the rhodopsin-bound peptide by overlaying the backbone atoms of several residues at the N-terminus of the peptide (illustrated in Fig. 5).

4.5. Protein complexes

Clore [67] developed an RDC-based methodology for treating protein complexes and applied it to the 40 kDa complex of the histidine phosphocarrier protein and the N-terminal domain of enzyme I. The structure of this complex was previously solved [68] by NMR using 109 intermolecular NOE restraints and 231 residual NH^N couplings measured in a solution of *fd* bacteriophage. Intermolecular NOE restraints were obtained from several specially prepared samples where one of the proteins was isotopically labeled. It is worth noting that the set of 109 intermolecular NOE restraints represented a small fraction (ca. 3%) of all NOEs used in the structure determination, which illustrates the elusive character of intermolecular NOE contacts.

To incorporate the RDC data in the analyses, the alignment parameters A_a and R were determined from the histogram approach. At the same time, the couplings were fitted to the X-ray coordinates yielding the quality factors $R_{\mathrm{dip}} = 16\%$ and 27% for the two protein structures with a resolution of 1.5 Å and 2.5 Å, respectively. The starting geometry of the complex was generated by positioning the two proteins (represented by the crystallographic structures) at the distance of 38 Å from each other and giving them an a priori incorrect relative orientation. This geometry was subjected to a round of rigid-body optimization where all atomic degrees of freedom were held fixed except 6 coordinates describing relative placement of the proteins and 3 parameters specifying the orientation of the alignment frame. The calculations were carried out in XPLOR equipped with the standard RDC module [12]; the target function included NOE, RDC, and van der Waals terms. In the final stage, side chains at the protein–protein interface (total of 29 side chains) were released and subjected to a round of Powell minimization in order to relieve numerous van der Waals conflicts resulting from the rigid-body procedure.

It is noteworthy that the structure obtained with this protocol is similar to the structure calculated without the RDC data (both fall within 1.2 Å of the original NMR structure for this complex [68]). Hence, the set of 109 intermolecular NOE restraints proved to be sufficient to define the binding topology so that RDC data add little value. However, when the calculations were repeated using a small subset of 8 intermolecular NOEs, the RDC data proved their worth. In this hypothetical situation, the poor quality of the NOE-based structure, 2.6 Å, improved dramatically, to 1.3 Å, upon adding the RDC restraints. This underscores the usefulness of RDC data which can clearly help to avoid much of costly and time-consuming intermolecular NOE measurements.

The RDC-based rigid-body protocol described above was further improved in work by Wang, Clore, and co-workers [41] where the complex of histidine phosphocarrier protein with another enzyme, glucose specific enzyme IIA^{Glc}, was investigated. Fitting 77 (118) experimentally measured NH^N couplings to the high-resolution structure of histidine phosphocarrier protein

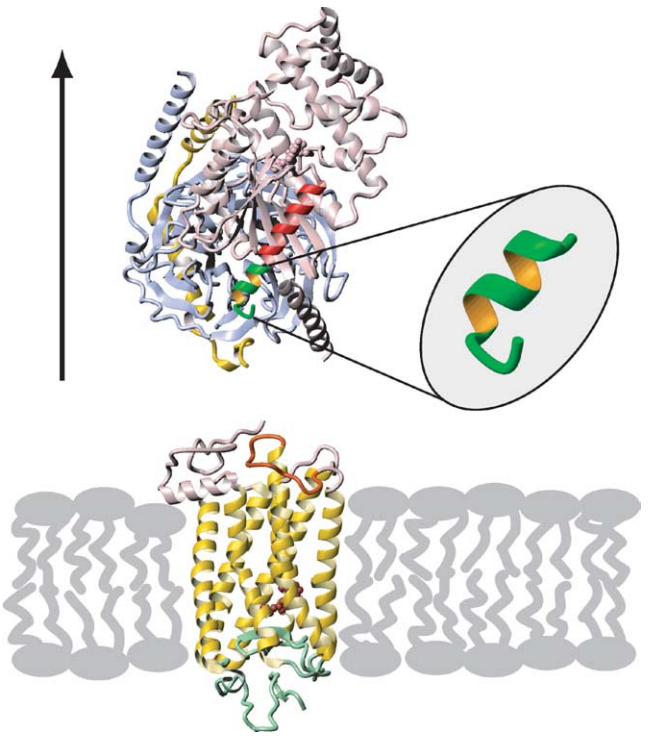

Fig. 5. Model of the α subunit of heterotrimeric G protein transducin binding to light-activated form of rhodopsin. The blow-up image shows the terminal undecapeptide making an angle of 50° to the membrane surface. The entire transducin structure is oriented to satisfy the orientation of the terminal peptide. Reprinted with permission from [71].

(glucose specific enzyme) yielded a quality factor $R_{\rm dip} = 17\%$ (15%). Following a round of rigid-body minimization, it was found that the side chains at protein–protein interface required substantial adjustment which could not be accomplished with a superficial Powell refinement. The need for side chain adjustment was also demonstrated by experimental measurements of three-bond J-couplings. The data indicated that a number of side chains changed their conformations upon transition from the free state to the complex. In order to make these changes possible, an extra step was added to the protocol whereby the interface side chains were released and treated to a round of torsional angle dynamics. During this stage the rest of the protein was maintained in a 'frozen' state using artificial symmetry restraints. Six degrees of freedom describing the relative position of the proteins were allowed to vary during this step. The optimization was controlled by a combination of restraints including selected intramolecular NOEs originating at interfacial side chains, an empirical potential for side-chain torsional angles, radius of gyration, etc. Following this step the structure was subjected to Powell refinement. All calculations were carried out using XPLOR-NIH.

Similar to the previous study, two essentially identical structures of the complex (*rmsd* 0.06 Å) were obtained with and without the RDC data. This indicates that the binding topology was well defined by the set of intermolecular NOEs at hand. Not surprisingly, very similar structures are generated using different RDC data sets. Likewise, cross-validation using independent subsets of the RDC data point toward the high quality of the final structure.

The third in the series of papers on the interactions of histidine phosphocarrier protein [69] describes its complex with the 148-residue cytoplasmic A domain of the mannitol transporter $\mathrm{II}^{\mathrm{Mannitol}}$. Broadly, the same protocol was followed as in the previous work. The total of 185 NH^{N}, 184 $\mathrm{C}^{\alpha}\mathrm{H}^{\alpha}$, and 159 NC^{α} couplings were measured in the two proteins dissolved in poly(ethylen glycol) – hexanol mixture. Very good agreement, $R_{\rm dip} = 19\%$, was obtained by fitting the RDC data to the crystallographic coordinates of the histidine-containing phosphocarrier protein. It can be assumed, therefore, that the structure of this protein does not change upon binding. The situation with the mannitol transporter is somewhat more complicated. The existing X-ray coordinates feature four molecules, A–D, per unit cell. The four structures are generally consistent with each other within 0.3 Å except for several regions where they are 1.5–3 Å apart. Of these structures, D shows a good overall fit with the RDC data, $R_{\rm dip}$ of ca. 20%, whereas the other structures, A–C, show a certain number of deviations that are linked to the regions of structural variability. The structure D was chosen as a template for the subsequent complex assembly procedure.

Structure calculations began with rigid-body minimization performed on two proteins initially placed 30–50 Å apart at random orientations. In the subsequent (semi)rigid-body routine a number of internal degrees of freedom were released. (i) Side-chain torsional angles were completely released for all interfacial side chains. (ii) Backbone dihedral angles were partially released in the regions where the crystal structures of mannitol transporter displayed variability (within the range of angles observed in A–D group). (iii) Side-chain torsional angles were partially released in the regions of structural variability of the mannitol transporter (within the range of angles observed in the A–D group). In addition, backbone dihedral angles were completely released in several residues of the mannitol transporter so that these residues could function as hinge points for concerted displacement of a 9-residue fragment, as observed in the A–D group. This intricate scheme was designed to account for the small degree of local plasticity revealed by the crystallographic coordinates of mannitol transporter. Note that at this stage the optimization retained its essential rigid-body character as 91% of the backbone and majority of side chains remained fixed. The reorientational and translational motion of one protein with respect to the other was permitted during this step. The procedure, implemented in XPLOR-NIH, used the entire array of the input NMR data, including RDCs and NOEs.

The final structure of the complex is in excellent agreement with the RDC data as indicated by R_{dip} factor (ca. 20% for NH^{N} couplings, 20–25% for $\mathrm{C}^{\alpha}\mathrm{H}^{\alpha}$ couplings). These R_{dip} values are similar to the values obtained from crystallographic structures of individual proteins, which is expected of a properly designed rigid-body procedure. The high quality of the final structure was further confirmed using RDC cross-validation. Similar to the previous studies, however, the topology of the complex is uniquely defined by the 107 intermolecular NOEs. Indeed, the structures calculated with and without dipolar couplings turn out to be within ca. 0.5 Å of each other and differ by ca. 5° with regard to relative orientation of the two proteins. The usefulness of the RDC data can only be fully appreciated when the number of observed intermolecular NOE contacts is small. Note, however, that one cannot completely forgo NOE measurements since at least some NOE data are necessary to constrain the translational degrees of freedom in RDC-driven calculations. Several approaches that substitute intermolecular NOEs for other types of data are outlined below.

McCoy and Wyss [70] proposed to use chemical shift mapping in conjunction with the RDC data. In their work, the X-ray structures of the N-terminal domain of enzyme I and the histidine phosphocarrier protein were first oriented according to the RDC data reported by Garrett, Clore and others [68]. The structures were subsequently docked based on the map of chemical shift changes that occur upon protein binding. At each step in the docking procedure, the algorithm evaluated the *rmsd* between the experimentally observed chemical shift perturbations and the predictions from semi-empirical program for calculating chemical shifts. This study demonstrates that the combination of RDCs (reporting on orientational degrees of freedom) and chemical shift data (translational degrees of freedom) can be used to assemble the complex, thus obviating the need for collecting and analyzing intermolecular NOE data. A similar approach was recently proposed by Arumugam and Van Doren [5]. The authors suggested that relaxation enhancement data (obtained with the help of paramagnetic contrast agents) could be used to 'highlight' the solvent-accessible area and, by exclusion, map the protein–protein interface.

The idea of McCoy and Wyss has been further developed in the recent work of Clore and Schwieters [48]. In their approach, the information on chemical shift variations is converted into loose empirical distance restraints. Each individual restraint imposes an upper boundary on the average distance between a given residue from the binding interface of protein A and all residues from the binding interface of protein B (interface residues are identified based on the magnitude of chemical shift perturbations, proximity to the surface, and other criteria). These restraints are designed to reflect the delocalized nature of chemical shift and are intended to replace intermolecular NOE data. The structure calculation protocol used in this study consists of several steps that include: (i) rigid-body minimization controlled by empirical restraints derived from chemical shifts; and (ii) (semi)rigid-body minimization controlled by chemical shift and RDC data where the side chains at the interface are allowed to float.

The protocol was tested for three different complexes of histidine-containing phosphocarrier protein described above [67,41,69]. The chemical shift mapping data were available from the previous studies. Calculations using the new protocol demonstrated that reasonably accurate results can be obtained, although the convergence properties of the algorithm are relatively poor due to amorphous character of the empirical distance restraints. Most of the newly obtained structures are within 1 Å of the previous accurate structures and show similar RDC quality factors, $R_{\mathrm{dip}} = 16$–25%. However, there is a considerable population of structures that are in error by up to 1.8 Å. Furthermore, some completely incorrect solutions, 10.7 Å, were obtained for the complex of histidine-containing phosphocarrier protein with the cytoplasmic A domain of mannitol transporter $\mathrm{II}^{\mathrm{Mannitol}}$. These erroneous structures appear as a result of the four-fold degeneracy in RDC-derived orientations. Since the z-axis of the alignment tensor happens to be approximately perpendicular to the protein–protein interface, a 180^{o} rotation of histidine phosphocarrier protein about this axis does not alter the RDCs and, at the same time, has no significant effect on the loose distance restraints across the binding interface. Clore and Schwieters outline several possible paths for resolving this ambiguity: (i) fully utilize chemical shift information including negative data, i.e. the absence of chemical shift perturbations upon binding; (ii) compute alignment parameters for the two hypothetical structures [21] and compare them with

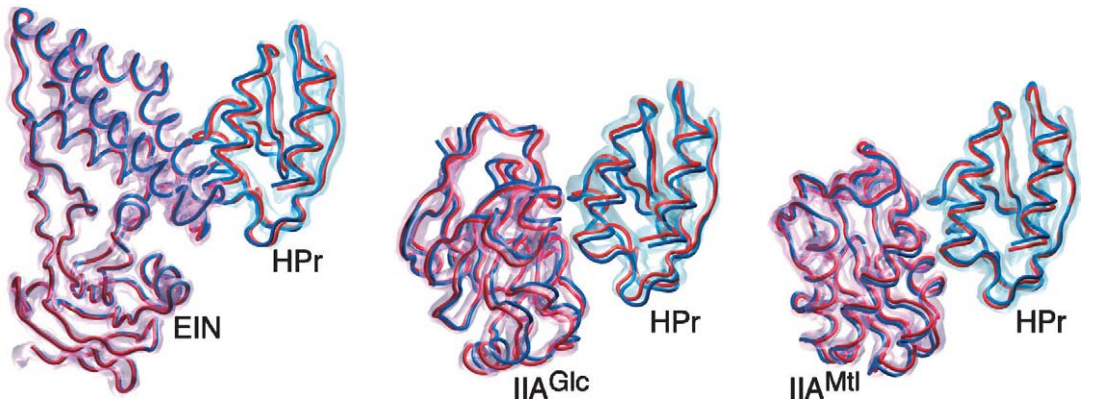

Fig. 6. Complexes of the histidine phosphocarrier protein with the N-terminal domain of enzyme I, glucose specific enzyme IIA^{Glc}, and cytoplasmic A domain of mannitol transporter II^{Mtl}. Structures obtained using RDC and intermolecular NOE data are shown in blue; structures obtained using RDC and chemical shift data are shown in red. Reprinted with permission from [48]. Copyright 2003 American Chemical Society.

the experimental alignment parameters; (iii) use a single histidine–histidine intermolecular distance restraint deduced from the known biological function of the complex; and (iv) record an additional set of dipolar couplings in a different alignment media. The first three options are shown to be fully successful in resolving the ambiguity (structures are shown in Fig. 6). Finally, the authors note that interface side chains are often rearranged as binding takes place. This observation implies that the original method by McCoy and Wyss, which critically depends on the orientation of side-chain aromatic rings at and near the interface, may have a limited scope of application.

5. Concluding remarks

While fairly simple on a conceptual level, the RDC-based rigid-body approach can be enormously useful in practical applications. In fact, it addresses an extremely important class of problems that otherwise cannot be satisfactorily resolved. Specifically, the RDC-based rigid-body approach helps to elucidate the quaternary and quasi-quaternary structure of large biomolecules and molecular complexes at near-physiological conditions (the term ‘quasi-quaternary structure’ is used here to describe spatial arrangement of covalently linked units such as domains). Proteins and nucleic acids typically function as multimodular systems. The spatial arrangement of the modules, such as domains or binding partners, is often controlled by a delicate balance of free energy on the order of no more than several kcal/mol. Unfortunately, in crystallographic studies this finely tuned balance is sometimes compromised by crystal packing forces which leads to non-native quasi-quaternary or quaternary structures. Furthermore, low-temperature crystallographic studies do not sample the free energy landscape in the same way as it is sampled at higher temperature under physiological conditions. As a result, crystallographic coordinates may not adequately represent the mean structure that would exist in solution. Likewise, conventional NMR is usually poorly suited to address these problems. In probing quasi-quaternary/quaternary structure, NMR mainly relies on nuclear Overhauser effect that can be notoriously difficult or, in some cases, practically impossible to measure. Intermolecular NOEs typically constitute a small fraction (several percent) of all available NOE data. Interdomain NOEs in multidomain proteins and nucleic acids are often completely non-existent. In this situation RDC data provide a welcome answer to the problem.

Importantly, the RDC-based rigid-body approach builds on the strengths of X-ray spectroscopy as it utilizes the crystallographic coordinates of individual structural units. While the RDC data do not uniquely define the relative position of structural modules, they go a long way toward this goal. The remaining uncertainty associated with the translation of structural units can be eliminated via limited NOE information or, alternatively, using a variety of sources such as small-angle X-ray scattering, paramagnetic relaxation, chemical shift mapping, etc. The degeneracies intrinsic to RDC data can be resolved by using multiple alignment media or alternative sources of information such as those mentioned above. Importantly, RDC data can be recorded quickly and at relatively low cost which is particularly useful for studies involving variable temperature or pH conditions, a series of molecules with point mutations, or a molecule with multiple binding partners. In summary, the power of the RDC-based rigid-body approach has been convincingly demonstrated over the last four years resulting in more than 50 publications and this method will undoubtedly be very useful in the years to come.

Acknowledgements

I would like to thank Natalie Goto for reading and commenting on the manuscript, Marius Clore, Bernd Koenig, Oscar Millet, and Arthur Pardi for kindly providing the illustrations.

References

[1] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Proc. Natl. Acad. Sci. USA 92 (1995) 9279–9283.
[2] N. Tjandra, A. Bax, Science 278 (1997) 1111–1114.
[3] J.A. Losonczi, M. Andrec, M.W.F. Fischer, J.H. Prestegard, J. Magn. Reson. 138 (1999) 334–342.
[4] M.L. Mattinen, K. Pääkkönen, T. Ikonen, J. Craven, T. Drakenberg, R. Serimaa, J. Waltho, A. Annila, Biophys. J. 83 (2002) 1177–1183.
[5] S. Arumugam, S.R. Van Doren, Biochemistry 42 (2003) 7950–7958.
[6] M.A. Markus, R.B. Gerstner, D.E. Draper, D.A. Torchia, J. Mol. Biol. 292 (1999) 375–387.
[7] G. Cornilescu, J.L. Marquardt, M. Ottiger, A. Bax, J. Am. Chem. Soc. 120 (1998) 6836–6837.
[8] G.M. Clore, D.S. Garrett, J. Am. Chem. Soc. 121 (1999) 9008–9012.
[9] J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, J.H. Prestegard, J. Am. Chem. Soc. 123 (2001) 1416–1424.
[10] N.R. Skrynnikov, N.K. Goto, D.W. Yang, W.Y. Choy, J.R. Tolman, G.A. Mueller, L.E. Kay, J. Mol. Biol. 295 (2000) 1265–1273.
[11] M. Zweckstetter, A. Bax, J. Biomol. NMR 23 (2002) 127–137.
[12] G.M. Clore, A.M. Gronenborn, N. Tjandra, J. Magn. Reson. 131 (1998) 159–162.
[13] G.M. Clore, A.M. Gronenborn, A. Bax, J. Magn. Reson. 133 (1998) 216–221.
[14] J.J. Warren, P.B. Moore, J. Magn. Reson. 149 (2001) 271–275.
[15] J. Meiler, N. Blomberg, M. Nilges, C. Griesinger, J. Biomol. NMR 16 (2000) 245–252.
[16] N.R. Skrynnikov, L.E. Kay, J. Biomol. NMR 18 (2000) 239–252.
[17] W.J. Wedemeyer, C.A. Rohl, H.A. Scheraga, J. Biomol. NMR 22 (2002) 137–151.
[18] N. Sibille, A. Pardi, J.P. Simorre, M. Blackledge, J. Am. Chem. Soc. 123 (2001) 12135–12146.
[19] V. Tsui, L.M. Zhu, T.H. Huang, P.E. Wright, D.A. Case, J. Biomol. NMR 16 (2000) 9–21.
[20] H.M. Al-Hashimi, J.R. Tolman, A. Majumdar, A. Gorin, D.J. Patel, J. Am. Chem. Soc. 123 (2001) 5806–5807.
[21] M. Zweckstetter, A. Bax, J. Am. Chem. Soc. 122 (2000) 3791–3792.
[22] M.X. Fernandes, P. Bernadó, M. Pons, J.G. de la Torre, J. Am. Chem. Soc. 123 (2001) 12037–12047.
[23] H.M. Al-Hashimi, H. Valafar, M. Terrell, E.R. Zartler, M.K. Eidsness, J.H. Prestegard, J. Magn. Reson. 143 (2000) 402–406.
[24] D.M. Jacobs, K. Saxena, M. Vogtherr, P. Bernadó, M. Pons, K.M. Fiebig, J. Biol. Chem. 278 (2003) 26174–26182.
[25] H.M. Al-Hashimi, P.J. Bolon, J.H. Prestegard, J. Magn. Reson. 142 (2000) 153–158.
[26] G.S. Thompson, H. Shimizu, S.W. Homans, A. Donohue-Rolfe, Biochemistry 39 (2000) 13153–13156.
[27] A. Bax, Protein Sci. 12 (2003) 1–16.
[28] R. Brüschweiler, X.B. Liao, P.E. Wright, Science 268 (1995) 886–889.
[29] P.M. Hwang, N.R. Skrynnikov, L.E. Kay, J. Biomol. NMR 20 (2001) 83–88.
[30] J. Kikuchi, J. Iwahara, T. Kigawa, Y. Murakami, T. Okazaki, S. Yokoyama, J. Biomol. NMR 22 (2002) 333–347.
[31] R. Varadan, O. Walker, C. Pickart, D. Fushman, J. Mol. Biol. 324 (2002) 637–647.
[32] C.A. Fowler, F. Tian, H.M. Al-Hashimi, J.H. Prestegard, J. Mol. Biol. 304 (2000) 447–460.
[33] J.H. Prestegard, Nat. Struct. Biol. 5 (1998) 517–522.
[34] P.J. Bolon, H.M. Al-Hashimi, J.H. Prestegard, J. Mol. Biol. 293 (1999) 107–115.
[35] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Nat. Struct. Biol. 4 (1997) 292–297.
[36] M.W.F. Fischer, J.A. Losonczi, J.L. Weaver, J.H. Prestegard, Biochemistry 38 (1999) 9013–9022.
[37] J.J. Chou, S.P. Li, C.B. Klee, A. Bax, Nat. Struct. Biol. 8 (2001) 990–997.
[38] N.K. Goto, N.R. Skrynnikov, F.W. Dahlquist, L.E. Kay, J. Mol. Biol. 308 (2001) 745–764.
[39] K. Bondensgaard, E.T. Mollova, A. Pardi, Biochemistry 41 (2002) 11532–11542.
[40] C.A.E.M. Spronk, J.P. Linge, C.W. Hilbers, G.W. Vuister, J. Biomol. NMR 22 (2002) 281–289.
[41] G.S. Wang, J.M. Louis, M. Sondej, Y.J. Seok, A. Peterkofsky, G.M. Clore, EMBO J. 19 (2000) 5635–5649.
[42] D. Fushman, R. Ghose, D. Cowburn, J. Am. Chem. Soc. 122 (2000) 10640–10649.
[43] A.T. Brunger, XPLOR: A System for X-ray Crystallography and NMR, Yale University Press, New Haven, CT, 1993.
[44] Molecular Simulations Inc., San-Diego.
[45] P. Güntert, C. Mumenthaler, K. Wüthrich, J. Mol. Biol. 273 (1997) 283–298.
[46] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, J. Magn. Reson. 160 (2003) 65–73.
[47] G.M. Clore, C.A. Bewley, J. Magn. Reson. 154 (2002) 329–335.
[48] G.M. Clore, C.D. Schwieters, J. Am. Chem. Soc. 125 (2003) 2902–2912.
[49] S.A. McCallum, A. Pardi, J. Mol. Biol. 326 (2003) 1037–1050.
[50] G. Lipari, A. Szabo, Biophys. J. 30 (1980) 489–506.
[51] G.A. Mueller, W.Y. Choy, D.W. Yang, J.D. Forman-Kay, R.A. Venters, L.E. Kay, J. Mol. Biol. 300 (2000) 197–212.
[52] J.C. Hus, D. Marion, M. Blackledge, J. Am. Chem. Soc. 123 (2001) 1541–1542.
[53] L. Trantírek, M. Urbášek, R. Štefl, J. Feigon, V. Sklenář, J. Am. Chem. Soc. 122 (2000) 10454–10455.

[54] F. Delaglio, G. Kontaxis, A. Bax, J. Am. Chem. Soc. 122 (2000) 2142–2143.
[55] M. Andrec, P.C. Du, R.M. Levy, J. Biomol. NMR 21 (2001) 335–347.
[56] C.A. Rohl, D. Baker, J. Am. Chem. Soc. 124 (2002) 2723–2729.
[57] L.C. Wang, Y.X. Pang, T. Holder, J.R. Brender, A.V. Kurochkin, E.R.P. Zuiderweg, Proc. Natl. Acad. Sci. USA 98 (2001) 7684–7689.
[58] E.T. Mollova, M.R. Hansen, A. Pardi, J. Am. Chem. Soc. 122 (2000) 11561–11562.
[59] H.M. Al-Hashimi, Y. Gosser, A. Gorin, W.D. Hu, A. Majumdar, D.J. Patel, J. Mol. Biol. 315 (2002) 95–102.
[60] Q. Zhang, R. Throolin, S.W. Pitt, A. Serganov, H.M. Al-Hashimi, J. Am. Chem. Soc. 125 (2003) 10530–10531.
[61] D.W. Yang, R.A. Venters, G.A. Mueller, W.Y. Choy, L.E. Kay, J. Biomol. NMR 14 (1999) 333–343.
[62] J. Evenas, V. Tugarinov, N.R. Skrynnikov, N.K. Goto, R. Muhandiram, L.E. Kay, J. Mol. Biol. 309 (2001) 961–974.
[63] G. Cornilescu, F. Delaglio, A. Bax, J. Biomol. NMR 13 (1999) 289–302.
[64] R.A. Laskowski, J.A.C. Rullmann, M.W. MacArthur, R. Kaptein, J.M. Thornton, J. Biomol. NMR 8 (1996) 477–486.
[65] O. Millet, R.P. Hudson, L.E. Kay, Proc. Natl. Acad. Sci. USA 100 (2003) 12700–12705.
[66] B.W. Koenig, G. Kontaxis, D.C. Mitchell, J.M. Louis, B.J. Litman, A. Bax, J. Mol. Biol. 322 (2002) 441–461.
[67] G.M. Clore, Proc. Natl. Acad. Sci. USA 97 (2000) 9021–9025.
[68] D.S. Garrett, Y.J. Seok, A. Peterkofsky, A.M. Gronenborn, G.M. Clore, Nat. Struct. Biol. 6 (1999) 166–173.
[69] G. Cornilescu, B.R. Lee, C.C. Cornilescu, G.S. Wang, A. Peterkofsky, G.M. Clore, J. Biol. Chem. 277 (2002) 42289–42298.
[70] M.A. McCoy, D.F. Wyss, J. Am. Chem. Soc. 124 (2002) 11758–11763.
[71] B.W. Koenig, ChemBioChem 3 (2002) 975–980.