van Dijk, A.D., Fushman, D. & Bonvin, A.M. Various strategies of using residual dipolar couplings in NMR-driven protein docking: application to Lys48-linked di-ubiquitin and valida...

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · AUGUST 2005

Impact Factor: 2.63 · DOI: 10.1002/prot.20476 · Source: PubMed

3 AUTHORS, INCLUDING:

Aalt - Jan Van Dijk
Wageningen University
65 PUBLICATIONS **1,691** CITATIONS
SEE PROFILE

Alexandre M J J Bonvin
Utrecht University
224 PUBLICATIONS **8,948** CITATIONS
SEE PROFILE

Various Strategies of Using Residual Dipolar Couplings in NMR-Driven Protein Docking: Application to Lys48-Linked Di-Ubiquitin and Validation Against ^{15}N-Relaxation Data

Aalt D.J. van Dijk,[1] David Fushman,[2] and Alexandre M.J.J. Bonvin[1]*

[1]*Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
[2]*Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, Maryland*

ABSTRACT **When classical, Nuclear Overhauser Effect (NOE)-based approaches fail, it is possible, given high-resolution structures of the free molecules, to model the structure of a complex in solution based solely on chemical shift perturbation (CSP) data in combination with orientational restraints from residual dipolar couplings (RDCs) when available. RDCs can be incorporated into the docking following various strategies: as direct restraints and/or as intermolecular intervector projection angle restraints (Meiler et al., J Biomol NMR 2000;16:245–252). The advantage of the latter for docking is that they directly define the relative orientation of the molecules. A combined protocol in which RDCs are first introduced as intervector projection angle restraints and at a later stage as direct restraints is shown here to give the best performance. This approach, implemented in our information-driven docking approach HADDOCK (Dominguez et al., J Am Chem Soc 2003;125:1731–1737), is used to determine the solution structure of the Lys48-linked di-ubiquitin, for which chemical shift mapping, RDCs, and ^{15}N-relaxation data have been previously obtained (Varadan et al., J Mol Biol 2002;324:637–647). The resulting structures, derived from CSP and RDC data, are cross-validated using ^{15}N-relaxation data. The solution structure differs from the crystal structure by a 20° rotation of the two ubiquitin units relative to each other. Proteins 2005;60:367–381.** © 2005 Wiley-Liss, Inc.

Key words: protein complexes; chemical shift perturbation; residual dipolar couplings; docking; diffusion anisotropy; di-ubiquitin

INTRODUCTION

In this post-genomic era, biochemical research focuses more and more on proteomics. Recent advances in high-throughput methods have provided a first glimpse of the overall structure of protein–protein interaction networks in biological systems (for a review, see ref. 1 and references therein) and promise to contribute increasingly to our understanding of how proteins cooperate.

Most proteins achieve their function by interacting with other proteins and forming an active complex. The structure determination of such complexes by X-ray and/or nuclear magnetic resonance (NMR) still presents a challenging task. Because of that, *ab initio* protein–protein docking is becoming increasingly popular (for a review, see ref. 2). Recently, the CAPRI (Critical Assessment of PRedicted Interactions) experiment[3] provided a blind test of the performance of different docking approaches (see http://capri.ebi.ac.uk). Although it is clear that considerable progress has been made in recent years, it is also obvious that docking based only on shape complementarity and/or a general force field, without any auxiliary information, is still very challenging. Reasonable solutions can be obtained, however, when a small amount of experimental information is included. We recently developed an information-driven method called HADDOCK (High Ambiguity Driven protein–protein DOCKing) to dock proteins using biochemical and/or biophysical data, such as mutagenesis data or the easily available chemical shift perturbation (CSP) data from NMR titration experiments.[4] Such information allows one to map the residues at the interface of the constituents of the complex and can be used to define so-called ambiguous interaction restraints (AIRs) to drive the docking. Here the ambiguity refers to the fact that residues can be identified to be located at the interface, but that there is no explicit knowledge on the specific pairwise interactions that residues from one component form with residues from the other component. In defining the ambiguous interaction restraints, one distinguishes between two types of residues: active residues that have been experimentally identified and passive residues that correspond to surface neighbors of active residues. AIRs are defined between each active residue of one component and all passive and all active residues of the other component. The use of passive residues, which introduces some fuzziness

---

The Supplementary Materials referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat

Grant sponsor: National Institutes of Health; Grant number: GM65334 to D. Fushman.

*Correspondence to: Alexandre M.J.J. Bonvin, Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, 3584 CH, Utrecht, The Netherlands. E-mail: a.m.j.j.bonvin@chem.uu.nl

Received 27 October 2004; Revised 22 December 2004; Accepted 14 January 2005

Published online 6 June 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20476

in the definition of the interface, is required to account for the fact that not all interface residues can typically be experimentally identified. For example, not all interacting residues in a complex show significant CSPs in NMR titration experiments. The reverse is also true: sometimes residues that are not part of the interface do show significant CSPs as a result of propagated effects, such as a weakening of hydrogen bond strength by a slight β-sheet opening or small conformational changes. The use of a solvent accessibility criterion allows one to filter out some of these "false positive" residues.

Classically, NMR-structure determination is completed in a lengthy process of collecting and assigning Nuclear Overhauser Effects (NOEs), representing short-range interproton distances. During the last few years, residual dipolar couplings (RDCs) and ^{15}N-relaxation have been shown to be very useful sources of long-range structural information.[5–10] In contrast to NOEs, RDCs and relaxation rates provide orientational information, which is of a global character. For complexes, one way to use this orientational information is to extract from the measured RDCs or relaxation rates for each component the alignment or diffusion tensor of the whole molecule. When the eigenvalues of the tensor of interest are equal for all components of the complex, this strongly indicates that the individual proteins orient or tumble together as a single rigid entity. Then, by rotating the individual components such that their alignment or diffusion tensors become collinear, the orientation of the individual molecules with respect to each other can be obtained.[11–14] Some ambiguity, however, typically remains because of the intrinsic degeneracy of RDC and relaxation data and the lack of information about the relative positioning of the molecules. When, in addition, a limited number of intermolecular NOE-derived distances can be used, the structure of the complex can be solved. This approach has been demonstrated for a variety of systems such as RNA,[15] protein–RNA,[16] and protein–protein complexes.[17]

More recently, approaches have emerged that do not use NOEs but rely on chemical shift mapping to determine the interface, with the advantage that weak and transient complexes can be studied in this way. Our HADDOCK approach, which we described above, makes use of such information. CSP information can also be combined with RDCs to refine the relative orientation of the components. It has been previously shown that by combining restraints from CSPs with orientational restraints from RDCs, the structure of a complex can be determined quite accurately;[18] the CSP restraints were taken into account by minimizing the difference between experimental and simulated CSPs using the program SHIFTS.[19] In another study, the chemical shift mapping data were transformed into a set of highly ambiguous intermolecular distance restraints, as proposed in HADDOCK, and combined with orientational restraints from RDCs.[20] RDC and CSP data have also been used *a posteriori* to filter solutions obtained from *ab initio* docking[21] with the program FTDOCK.[22]

RDCs are typically used directly as restraints in structure calculation methods by defining an external alignment tensor and back-calculating the RDC values from the angle between the internuclear vector of interest and the alignment tensor. In the case of complexes, the orientation of each component is optimized with respect to a single alignment tensor. In this way, the restraints only indirectly define the relative orientation of the components. It has been shown that, by combining RDCs in a pairwise manner, it is possible to define intervector projection angle restraints that become independent of an external alignment tensor. The latter is thus no longer needed in structure calculations.[23] These angular restraints have the advantage that they can be introduced in the initial steps of structure calculation starting from random conformations. Such an approach has, until now, never been used for docking. From a computational point of view, it should have some definite advantages. By combining RDCs from the individual components, intermolecular projection angle restraints can be generated that directly define the relative orientation of the components and can be distinguished from the intramolecular components. The use of such restraints should provide an easy way to use data from different alignment media together or to combine RDC restraints with other orientational information, such as relaxation anisotropy restraints.[24]

In this article we show that the structure of complexes can be efficiently determined from a combination of CSP data and RDCs using a combined approach in which the latter are first introduced as intermolecular intervector projection angle restraints and, at a later refinement stage, as direct restraints. The resulting models are independently validated using ^{15}N-relaxation data. This allows one to assess the possible influence of conformational averaging on the calculated structures, since the timescales for the relaxation data differ from that for the RDCs. This approach is applied in order to determine the solution structure of the Gly76-Lys48 linked di-ubiquitin chain (Ub_2) for which NMR CSPs, RDCs and relaxation data have been obtained previously.[25] The ubiquitin proteasomal pathway is the major mechanism of protein degradation in eukaryotic cells, and Lys48-linked tetra-ubiquitin (Ub_4) is the minimum signal required for efficient targeting of proteins to proteasomes.[26–28] Knowledge of the structure of polyubiquitin could provide useful insights into the mechanisms of its interaction with various chain recognition factors of the proteasome. At present, there is one crystal structure of a Lys48-linked Ub_2 (accession number: 1AAR)[29] and two crystal structures of Lys48-Ub_4 (accession numbers: 1F9J[30] and 1TBE[31]) in the Protein Data Bank,[32] all obtained at acidic pH (≈4.5). These structures differ from each other in the relative orientation and contacts between Ub units. In order to characterize the solution conformation of polyubiquitin, Lys48-linked Ub_2 and Ub_4 and Lys63-linked Ub_2 were recently investigated using a combination of NMR techniques, including chemical shift mapping, ^{15}N-relaxation and RDC measurements.[25,33] It was found that Lys48-linked chains exhibit a pH-dependent conformational switch from an open (at pH 4.5) to a closed form (pH 6.8 and higher). In the closed conformation observed under near-physiological conditions in solution, the relative orientation of the two monomers with respect to each

other was found to be distinct from that in the crystal structure. However, no interdomain NOEs could be reliably detected, probably due to conformational exchange broadening of the amide signals from the interface residues. Therefore, no solution structure of Ub_2 was generated, because the relaxation- and RDC-derived orientational restraints alone do not provide information on the relative positioning of the two domains. Here we show, using the previously collected NMR data[25] at pH 6.8, that the combination of intermolecular intervector projection angle restraints derived from RDC data and AIRs derived from CSP data can be used to determine the first solution structure of Lys48-linked Ub_2. The ^{15}N-relaxation data are used for cross-validation of our structure. Finally, the solution structure of Ub_2 generated in this way is compared to the crystal structure obtained at pH 4.5.

MATERIAL AND METHODS

Docking Protocol

The original HADDOCK protocol is described in ref. 4. It uses AIRs defined from CSP obtained in NMR titration experiments. The protocol consists of three consecutive stages (for details, see ref. 4):

(i) Randomization of orientations followed by rigid body energy minimization (EM);
(ii) Semi-flexible simulated annealing in torsion angle space (TAD-SA), which consists of (ii-a) a rigid body molecular dynamics search and first simulated annealing, (ii-b) a second semi-flexible simulated annealing during which side chains at the interface are free to move, and (ii-c) a third semi-flexible simulated annealing during which both side chains and backbone at the interface are free to move; and
(iii) Final refinement in Cartesian space with explicit solvent.

The docking was performed starting from the 10 conformations in the NMR ensemble [Protein Data Bank (PDB) entry 1D3Z] and from the two Ub monomers taken from the crystal structure of Ub_2 (PDB entry 1AAR) following the standard HADDOCK protocol. These 12 starting conformations resulted in 144 pairwise combinations for docking. In the initial rigid-body docking phase, 1440 structures were generated (each combination being used 10 times), and the best 200 in terms of total intermolecular energy were further submitted to the semi-flexible simulated annealing and final water refinement. After water refinement, the resulting structures were clustered based on root mean square deviation (RMSD), and the lowest energy structures from the lowest energy cluster were selected; clustering was completed with a 1.5 Å cutoff using the backbone RMSD of both proteins calculated after positional least square fitting on the first protein only.

Use of RDC Restraints for Docking

The measured RDCs are given by

$$D^i(\beta^i, \alpha^i) = 0.5D_0\Big[A_a(3\cos^2\beta^i - 1) + \frac{3}{2}A_r(\cos 2\alpha^i \sin^2\beta^i)\Big] \quad (1)$$

Here A_a is the axially symmetric part of the alignment tensor, equal to $[A_{zz} - 1/2(A_{xx} + A_{yy})]$ and A_r is the rhombic component of the alignment tensor, equal to $(A_{xx} - A_{yy})$, where A_{xx}, A_{yy} and A_{zz} are the x, y and z-components of the alignment tensor, respectively; α^i and β^i are the azimuthal and polar angles of the vector for which the RDC is reported, in the frame of the alignment tensor; $D_0 = -\left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_i\gamma_j h}{2\pi^2 r_{NH}^3} = 21.7$ kHz is the strength of the (static) dipolar coupling in the NH-pair; r_{NH} is the length of the NH-vector; μ_o is the magnetic permeability of vacuum; γ_i is the gyromagnetic ratio of spin i; and h is Planck's constant.

In order to use RDCs for structure calculation, the alignment tensor components need to be determined. This was completed by back-calculating RDCs based on the known structure of the single domains (10 NMR structures and two monomers from the crystal structure) and optimizing the tensor parameters, using the software Pales[34] with the option –bestFit (which uses singular value decomposition). This fitting procedure was performed separately for the RDC data measured for the distal and proximal domains, leading to 24 sets of tensor parameters. Only RDCs from residues located in secondary structure elements were used (50 in total). We then selected the tensor parameters from the structure and RDC data of the domain with the lowest Q-factor.

Experimental RDCs were introduced in the standard HADDOCK protocol in three different ways: (a) directly, using a floating alignment tensor (SANI[35] energy term in CNS[36]); (b) as intervector (NH-NH) projection angle restraints (VEAN[23] energy term in CNS); and (c) as a combination of (a) and (b). In the last case, VEAN was used in stages (i) and (ii) of the docking protocol (rigid-body docking and semi-flexible SA), and SANI was used in stage (iii), the final water refinement. The values of the force constants in the protocol are given in Table I. The use of RDCs as direct restraints in structure calculations is described in ref. 35. The use of RDCs as intermolecular intervector projection angle restraints is described in ref. 23; however, since it is less commonly used, we will describe it again in the following section.

RDCs as Intermolecular Intervector Projection Angle Restraints

The energy function[23] in this case is given by

$$E = \begin{cases} k_{border}(\phi^{ij} - \phi^{ij}_{ext1})^2 & 0 \le \phi^{ij} \le \phi^{ij}_{ext1} \\ 0 & \phi^{ij}_{ext1} \le \phi^{ij} \le \phi^{ij}_{ext2} \\ k_{centre}\cos^2\left[\pi\left(\dfrac{\phi^{ij} - \phi^{ij}_{ext2}}{\phi^{ij}_{ext3} - \phi^{ij}_{ext2}} - \dfrac{1}{2}\right)\right] & \phi^{ij}_{ext2} \le \phi^{ij} \le \phi^{ij}_{ext3} \\ 0 & \phi^{ij}_{ext3} \le \phi^{ij} \le \phi^{ij}_{ext4} \\ k_{border}(\phi^{ij} - \phi^{ij}_{ext4})^2 & \phi^{ij}_{ext4} \le \phi^{ij} < 180° \end{cases} \quad (2)$$

TABLE I. Force Constants Used During Different Stages of Docking Protocol

Stage[a]		k_{sani} (kcal mol^{-1} Hz^{-2})	$k_{vean,border}$ (kcal mol^{-1})	$k_{vean,centre}$ (kcal mol^{-1})	k_{air} (kcal mol^{-1} Å^{-2})
i)	Rigid body EM	0.0–0.04	0.0–4.0	0.0–1.0	1–10
ii-a)	SA	0.01–0.02	1.0–10	0.25–2.5	10–50
ii-b)	SA	0.02–0.2	10–40	2.5–10	50
ii-c)	SA	0.2	40	10	50
iii)	Water refinement	0.2	40	10	50

a) i, ii, and iii refer to the rigid body, simulated annealing (SA) and water refinement stages of the protocol, respectively. k_{sani} is the force constant for direct RDC restraints. $k_{vean,border}$ and $k_{vean,centre}$ are the force constants for RDC restraints expressed as intervector projection angles (VEAN) for the border and central parts of the VEAN potential function, respectively [see eq. (2)] k_{air} is the force constant for Ambiguous Interaction Restraints. During the SANI run, $k_{vean,border}$ and $k_{vean,center}$ were set to 0 and during the VEAN run, k_{sani} was set to 0; during the VEAN-SANI run, k_{sani} was set to 0 during stage i and ii, and $k_{vean,border}$ and $k_{vean,center}$ were set to 0 during stage iii.

Here ϕ^{ij} denotes the value of the projection angle between the two vectors to which the restraint is applied, in our case, NH-vectors; its boundary values are given by the following (note that the original versions of these equations in ref. 23 are not correct; the corrected versions are given in ref. 37):

$$\cos\phi^{ij}_{ext} = \sqrt{\frac{2(3D^i - 2D_{zz} + D_{xx} + D_{yy})}{3\left[\frac{3}{2}(D_{xx} - D_{yy})\cos\alpha^i - 2D_{zz} + D_{xx} + D_{yy}\right]}} \sqrt{\frac{2(3D^j - 2D_{zz} + D_{xx} + D_{yy})}{3\left[\frac{3}{2}(D_{xx} - D_{yy})\cos\alpha^j - 2D_{zz} + D_{xx} + D_{yy}\right]}} \cos(\alpha^j \pm \alpha^i) \pm \sqrt{1 - \frac{2(3D^i - 2D_{zz} + D_{xx} + D_{yy})}{3\left[\frac{3}{2}(D_{xx} - D_{yy})\cos\alpha^i - 2D_{zz} + D_{xx} + D_{yy}\right]}} \sqrt{1 - \frac{2(3D^j - 2D_{zz} + D_{xx} + D_{yy})}{3\left[\frac{3}{2}(D_{xx} - D_{yy})\cos\alpha^j - 2D_{zz} + D_{xx} + D_{yy}\right]}} \quad (3)$$

with

$$\alpha^i_{min} = \frac{1}{2}\arccos\left(\frac{6D^i + 2D_{zz} - D_{xx} - D_{yy}}{3(D_{xx} - D_{yy})}\right)$$

$$\alpha^i_{max} = \pi - \frac{1}{2}\arccos\left(\frac{6D^i + 2D_{zz} - D_{xx} - D_{yy}}{3(D_{xx} - D_{yy})}\right)$$

if

$$\left|\frac{6D^i + 2D_{zz} - D_{xx} - D_{yy}}{3(D_{xx} - D_{yy})}\right| \leq 1$$

and otherwise

$$\alpha^i{}_{min} = 0$$
$$\alpha^i{}_{max} = \pi$$

Here, D^i and D^j are the values of the respective RDCs of the two NH-vectors; D_{xx}, D_{yy} and D_{zz} are given by $D_0 \cdot A_{xx}$, $D_0 \cdot A_{yy}$ and $D_0 \cdot A_{zz}$, respectively. Generally, two minima are allowed, due to the well-known degeneracy of the dependency of the RDC on α^i and β^i, and correspondingly, there are four boundary values for ϕ^{ij}, denoted $\phi^{ij}{}_{ext1}$ to $\phi^{ij}{}_{ext4}$ in eq. (2). Two force constants must be defined, one for the border potential function (k_{border}) and one for the central part between the two minima (k_{centre}).

The generation of intermolecular intervector projection angle restraints was implemented in a slightly modified version of a python script (dipolar_segid.py) kindly provided by Drs. Helen Mott and Wayne Boucher (Cambridge University). This script is distributed with the HADDOCK package (see www.nmr.chem.uu.nl/haddock).

NMR Data

The NMR data used in this study include RDC values for backbone amides, ^{15}N-relaxation rates R_1 and R_2 and steady-state ^{1}H-^{15}N NOEs. Ub_2 chains were assembled from unlabeled and U-^{15}N labeled Ub units, such that only one of the two Ub domains was ^{15}N labeled in each NMR sample. The NMR studies were conducted at pH 6.8. The details of chain assembly and experimental NMR procedures have been presented elsewhere.[25] The overall average levels of R_1 in the distal Ub appeared to be slightly elevated (6.5%) compared to those in the proximal domain, while the R_2 values showed the opposite tendency. This indicates a slight difference in the apparent overall rotational correlation time that could be attributed to interdomain mobility in Ub_2. It is also possible that small variations in protein concentration between NMR samples, in which the distal or the proximal domain in Ub_2 was isotope labeled, contribute (due to aggregation effects) to the observed differences in the relaxation rates between the domains. Ub_2 showed some tendency to aggregate at NMR concentrations (≈1 mM), as inferred from the observed concentration dependence of spin relaxation rates. Note that varying protein concentration can affect the viscosity of the sample (hence the relaxation rates); however, at these low ubiquitin concentrations, this effect is expected to be negligible. Although the relaxation measurements for these studies were performed at Ub_2 concentrations as low as 250–500 μM, some amount of aggregation could still be present. Note that the sub-nanosecond local backbone dynamics in both Ub domains are very similar to

each other and to those in the monomeric Ub.[38] In order to be able to fit the experimental data for both domains simultaneously, relaxation rates for the distal domain were uniformly rescaled as follows: $R_2 \rightarrow R_2 \cdot 1.065$, $R_1 \rightarrow R_1/1.065$. This scaling factor represents an average value of the ratio of R_2 values between the proximal and the distal domains and of the inverse ratio of R_1 values for these domains. Analysis of the diffusion tensor data showed that this procedure mostly affects the principal values but not the orientation of the diffusion tensor. This is further supported by the agreement between the orientation of the diffusion tensors derived from these data and calculated from hydrodynamic properties of the Ub_2 structures (Supplementary Material).

The analysis of the experimental data using ROTDIF[38,39] was performed including core residues, identified as those belonging to Ub secondary structure, excluding six NH groups (residues 7,23,25,48,70,71) in the distal and seven (residues 23,25,48,56,68,70,71) in the proximal Ub that show conformational exchange broadening or could not be reliably evaluated from the spectra due to signal overlap. We also performed a similar analysis including most of the NH vectors except for those mentioned above and those in the β1/β2 loop and in the C-terminus. The results of these analyses were very similar in terms of the derived diffusion tensor characteristics. Only the "core"-residue results are presented here. The ROTDIF program determines the overall rotational diffusion tensor of a molecule from the ratio $\rho = \left(\frac{2R_2'}{R_1'} - 1\right)^{-1}$ of the rates of ^{15}N longitudinal and transverse relaxation, R_1' and R_2'; the prime indicates that these rates are modified by subtracting the contributions from high-frequency components of the spectral density.[40,41] Note that this ratio, ρ, is independent, to the first approximation, of the site-specific values of the ^{15}N CSA tensor and of the backbone order parameters. The tensor derivation is based on the minimization of the target function

$$\chi^2 = \sum_i \left(\frac{\rho_i^{exp} - \rho_i^{calc}}{\sigma_i} \right)^2 \quad (4)$$

where the superscripts 'exp' and 'calc' indicate the experimentally measured and the calculated values of ρ for the i-th NH vector, σ_i is the experimental error in ρ_i^{exp}, and ρ_i^{calc} is calculated for a given model of the overall tumbling as described in detail in ref. 38. All data were treated using an axially symmetric and a fully anisotropic model for the rotational diffusion tensor.

Structural Coordinates

The structural coordinates for the ubiquitin monomer used as the starting structure in the docking were taken from the Brookhaven Protein Data Bank[32] entry 1D3Z,[42] an NMR structure of the monomer, of which the 10 models were used, and entry 1AAR,[29] the crystal structure of Ub_2, which was split into two parts, both being used as starting structures for a Ub monomer.

The ensemble of the 10 solution structures of Ub_2 described in this study have been deposited in the PDB under accession code 2BGF, together with the various NMR restraints used for the docking.

Analysis of Intermolecular Contacts

Intermolecular contacts (hydrogen bonds and non-bonded contacts) were analyzed with DIMPLOT, part of the LIGPLOT software,[43] using the default settings (3.9 Å heavy-atom distance cutoff for non-bonded contacts; 2.7 and 3.35 Å proton-acceptor and donor-acceptor distance cutoffs respectively with minimum 90° angles (D-H-A, H-A-AA, D-A-AA) for hydrogen bonds. A contact is defined to be present in the solution structure if it is found in at least four of the 10 best structures.

RESULTS AND DISCUSSION

To compare the performance of using RDCs as direct restraints or as intermolecular intervector projection angle restraints, the NMR solution structure of the Gly76-Lys48 linked Ub_2 was first calculated in three different manners with HADDOCK:[4] (i) using solely CSP data, (ii) using CSP data in combination with direct RDC restraints, and (iii) using CSP data in combination with intermolecular intervector projection angle restraints derived from RDC data (see Material and Methods). These three runs will be denoted in the following as (i) CSP, (ii) SANI, and (iii) VEAN, respectively (SANI and VEAN refer to the energy terms in CNS[36]). Analysis of the results led us to define a protocol termed VEAN-SANI, in which both approaches are combined, using first the intervector projection angle restraints to drive the docking and then the direct RDC restraints for the final explicit solvent refinement. In cases where ambiguity is present in the experimental data, different orientations can be obtained by docking, which should then be distinguished based on energetic considerations, as shown previously.[4] To assess the performance and convergence of the various protocols presented here, we therefore compare the populations of the lowest energy clusters.

Restraints Definition

The restraints that were used are listed in Table II. The ubiquitin monomers, which are linked to one another via a Gly76—Lys48 isopeptide bond, are designated distal (D: containing Gly76) and proximal (P: containing Lys48 and a free C-terminus), respectively. For both distal and proximal domains, the active and passive residues used in the docking were defined using the procedure described previously:[4] the residues with combined ^1H and ^{15}N CSP above average (0.033 ppm) whose backbone or side chain showed more than 50% relative solvent-accessible surface area were defined as active residues. Their solvent-accessible surface neighbors were defined as passive residues. From the list of active and passive residues for both monomers, 21 AIRs were defined with an upper distance bound of 2 Å.

Two sets of each 46 NH-RDCs (for the distal and proximal domains) were available. As starting structures, we used the 10 models from the NMR structure of ubiquitin (1D3Z) and the two units of the Ub_2 X-ray structure

TABLE II. Data Used in Docking

Distal Domain[a]	
Active residues[b]	L8, T9, A46, G47, K48, Q49, E51, H68, R72, L73
Passive residues[b]	K6, G10, K11, T12, D39, D52, G53, R54, L71, R74, G75, G76
RDCs	46 NH RDCs
Flexible segments[c]	F4-T14; P37-L56; T66-L71
Fully flexible segments[d]	R72-G76
Proximal Domain[a]	
Active residues[b]	L8, T9, G47, K48, E51, H68, V70, R72, L73, R74, G76
Passive residues[b]	K6, G10, K11, T12, D39, A46, Q49, D52, G53, R54, L71, G75
RDCs	46 NH RDCs
Flexible segments[c]	F4-T14; P37-Q41; F45-L56; T66-L71
Fully flexible segments[d]	K48; R72-G76
Intervector Projection Angle Restraints[e]	**Number of Restraints**
Intermolecular	981
Intermolecular	972
Isopeptide Bond (G76-K48)[f]	**Unambiguous Restraint Distance (Å)**
O-NZ	2.25 ± 0.05
C-NZ	1.35 ± 0.05
C-CE	2.45 ± 0.05
CA-NZ	2.45 ± 0.05

[a]The ubiquitin monomers, which are linked to one another via a Gly76–Lys48 isopeptide bond, are designated distal (D: containing Gly76) and proximal (P: containing Lys48 and a free C- terminus), respectively.
[b]Our docking protocol uses active residues [interface residues defined based on experimental CSP data (see text)] and passive residues (surface neighbors of active residues). Ambiguous interaction restraints, AIRs, are defined from each active residue of one domain to all active and passive residues of the other domain, using a 2 Å upper distance bound.
[c]Flexible segments are the parts of the molecules that are free to move during the SA stage ii-b and ii-c (see text). These are typically defined as all active and passive residues plus two sequential neighbors.
[d]Fully flexible segments are the parts of the molecules that are free to move during all SA stages (see text).
[e]One docking run was performed using only Chemical Shift Perturbation data (CSP). RDCs were used either as direct restraints (SANI) or as intervector projection angle restraints (VEAN), in both cases combined with CSP. The final protocol consists of a combination of VEAN and SANI together with CSP (VEAN-SANI); see text for details.
[f]The distal and proximal domains are connected via an isopeptide bond between Gly76 carbonyl C and Lys48 NZ; in the docking this is represented as a set of distance restraints based on typical distances for this bond in the crystallographic structure.

(1AAR). For each of those 12 starting structures, the principal components of the alignment tensor were calculated using Pales[34] for each RDC set independently. For this, only residues located in secondary structure elements were considered (50 for the two units together). The goodness of fit was assessed with the RDC Q-factor.[42] Model 8 from the NMR ensemble, fitted to the RDC data for the distal domain, gave the lowest Q-factor ($Q = 0.21$; $R = 0.96$); the resulting principal component values ($A_{xx} = 32.3 \times 10^{-5}, A_{yy} = 38.3 \times 10^{-5}, A_{zz} = -70.6 \times 10^{-5}; A_a = -105.9 \times 10^{-5}$, and $A_r = -6.02 \times 10^{-5}$) were used subsequently in the structure calculations. The RDC data were introduced either as direct restraints (92 in total) with explicit inclusion of an alignment tensor (SANI) or as intervector projection angle restraints (VEAN). The latter were defined by taking pairwise combinations of the total set of 92 RDCs and selecting only the intervector projection angle restraints restricting at least 25% of the conformational space. In this way, 981 inter-domain and 972 intra-domain restraints were defined. Furthermore, four unambiguous restraints were introduced to represent the covalent isopeptide bond from Gly76 of the distal domain to Lys48 of the proximal domain. This approach was chosen instead of defining a covalent bond to allow for randomization of the starting orientations and to facilitate the docking.

Docking From CSP

Figure 1 shows the intermolecular energies of the water-refined solutions as a function of the RMSD from the lowest energy structure of the respective runs. The 200 solutions after final refinement in explicit water were clustered using a 1.5 Å pairwise RMSD cutoff, and clusters were ranked using the average intermolecular energies calculated for the 10 lowest-energy structures. In this way, different cluster sizes do not affect the ranking. After clustering, the lowest energy structure(s) of the lowest energy cluster were taken as the best docking solution. The lowest energy cluster for the CSP docking has 39 members. As shown in Table III, the Q-factor for this run is relatively high (0.64 ± 0.07), indicating that those structures do not satisfy the RDC restraints.

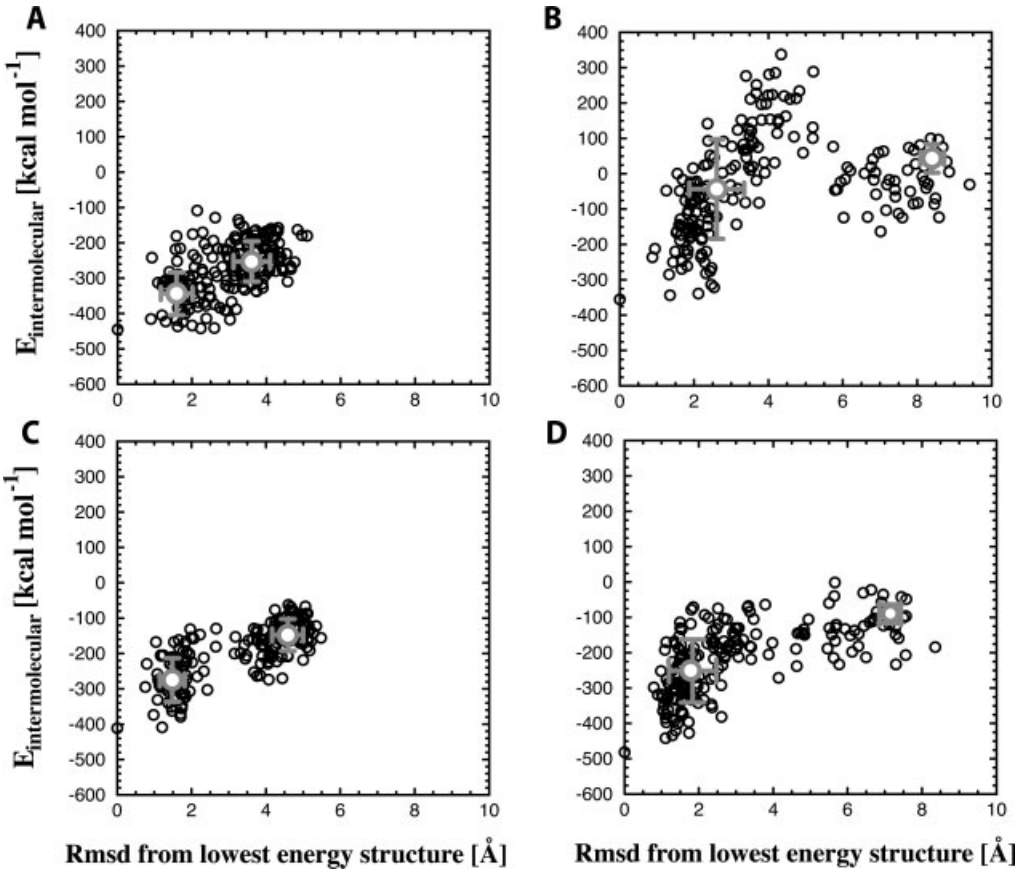

Fig. 1. Intermolecular energies ($E_{vdw} + E_{elec} + E_{restraints}$) versus backbone RMSD from the lowest-energy structure. Values for single structures are indicated by black dots, and cluster averages and standard deviations are shown in gray: (a) CSP, (b) VEAN, (c) SANI, and (d) final protocol, VEAN-SANI. For a description of the different docking runs, see text and footnote e of Table II.

TABLE III. Consistency of Docked Ub_2 Structures With RDC Data

	Crystal (1AAR)	Solution Structures[a]			
		CSP	SANI	VEAN	VEAN-SANI
Q-factor[b]	0.44	0.64 (0.07)	0.14 (0.01)	0.35 (0.03)	0.15 (0.01)
RDC viol.[c] > 1 Hz	78	77 (3)	38 (5)	63 (3)	41 (4)
RDC viol.[c] > 3 Hz	48	51 (6)	4 (1)	31 (3)	6 (2)
VEAN inter viol.[d] > 10^0	148	154 (13)	62 (5)	29 (10)	61 (6)
VEAN intra viol.[d] > 10^0	71	123 (12)	56 (4)	18 (8)	54 (3)

[a]For description of different docking runs, see footnote e of Table II. The statistics are for the 10 lowest energy structures of the lowest energy cluster. See also Figure 2 for a histogram of deviations from experimental RDC values.
[b]Q-factor = RMS($D^{calc} - D^{obs}$)/RMS(D^{obs}),[36] where D^{calc} and D^{obs} are calculated and observed values of the RDCs, respectively.
[c]RDC viol.: number of violations of direct RDC restraints (using cutoffs of 1 and 3 Hz, respectively); total number of direct RDC restraints is 92.
[d]VEAN inter viol. and intra viol.: number of violations of intermolecular and intramolecular VEAN-restraints, respectively (using cutoff of 10°). The VEAN restraints consist of 981 inter- and 972 intra-molecular restraints.

The CSP-structures are quite similar to the crystal structure, with an average backbone RMSD from it of 1.0 Å (Table IV). The crystal structure itself has a Q-factor of 0.44, which indicates that it is also not consistent with the experimental RDCs.

Docking From CSP and Intermolecular Intervector Projection Angle Restraints (VEAN)

As is clear from Figure 1(b), the VEAN run has a very good convergence; the lowest energy cluster is highly populated with 107 members. Note that the energies in Figure 1 contain the restraint energies (AIR, VEAN, SANI), which explains why they differ between the various runs. The corresponding VEAN structures have a better Q-factor (0.35 ± 0.03) than the CSP-only structures and the crystal structure. As can be seen in Table III, the number of direct RDC (which were not used) or intervector projection angle restraint violations is lower than for the crystal or the CSP-only docked structure. The structures differ from the CSP structures

TABLE IV. Average Pairwise RMSD (Å) Between Crystal Structure and Various Docked Structures[a]

	Crystal	CSP	VEAN	SANI	VEAN-SANI
Crystal	—	1.0 (0.2)	2.4 (0.5)	1.6 (0.3)	1.7 (0.3)
CSP	1.4 (0.2)	0.8 (0.2)/0.9 (0.2)	1.5 (0.3)	1.4 (0.3)	1.5 (0.3)
VEAN	2.7 (0.5)	1.8 (0.4)	0.8 (0.3)/0.9 (0.3)	1.0 (0.2)	1.0 (0.3)
SANI	2.0 (0.3)	1.6 (0.4)	1.3 (0.2)	0.7 (0.2)/0.9 (0.2)	0.9 (0.2)
VEAN-SANI	2.0 (0.4)	1.8 (0.3)	1.2 (0.2)	1.2 (0.2)	0.7 (0.2)/0.8 (0.3)

[a]Upper right: backbone RMSD; lower left: all-heavy atom RMSD. The RMSD values are calculated for the respective ensembles of structures as the averages from the respective average structure. Standard deviations are indicated in brackets. Residues 1 to 71 were used to calculate RMSD for both distal and proximal domains (excluding residues 72–76, which are a flexible tail). For a description of the different docking runs, see footnote e of Table II.

(backbone RMSD from average CSP structure of 1.5 ± 0.3 Å) and from the crystal structure (backbone RMSD of 2.4 ± 0.5 Å). These differences are larger than the spread within the ensemble of 10 structures (see Table IV).

Docking From CSP and Direct RDC Restraints (SANI)

Figure 1(c) shows that the SANI run does not converge as well as the VEAN run; there are only 65 structures in the lowest energy cluster. The Q-factor is low, 0.14 ± 0.01, which indicates excellent agreement with the experimental RDC data. Table III shows that the SANI structures have less direct RDC violations than the VEAN structures but more intervector projection angle restraint violations (although still less than the CSP-only docked structures or the crystal structure). This indicates that the two kinds of RDC restraints do not define or restrict the orientation of the N-H vectors in a fully similar way. The average backbone RMSD of the SANI structures from the crystal structure is 1.6 ± 0.3 Å and from the average CSP structure 1.4 ± 0.3 Å.

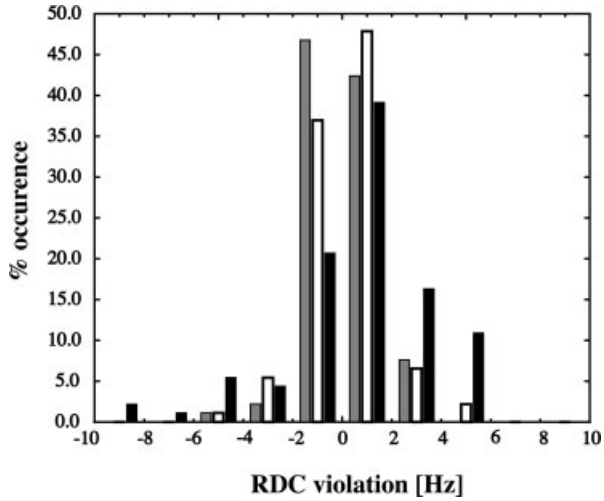

Fig. 2. Histogram of differences between experimental and calculated (best-fit using the program Pales[34]) RDC values for the structures from the different docking runs using RDCs (open black bar: SANI; filled black bar: VEAN; filled gray bar: VEAN-SANI; see footnote e of Table II). The bin size used was 2 Hz.

Comparison of Direct RDC (SANI) and Intervector Projection Angle (VEAN) Run

The VEAN structures and the SANI structures are both more consistent with the experimental RDC data than the crystal structure or the CSP structures. Still, there are differences between those structures, reflected in an average backbone RMSD from the mean of VEAN from SANI of 1.0 ± 0.2 Å. To compare the consistency of both sets of structures with the RDC data in more detail, a histogram of the differences between calculated and observed RDCs is presented in Figure 2. Both runs show a Gaussian-like distribution around a mean value of 0, but the distribution for SANI is tighter than the one for VEAN. The maximum violation in SANI is 4.2 Hz, which, considering the 50-Hz range of the measured RDCs, is below 10% of the experimental values. Note also that we are using a flat bottom harmonic potential for the RDCs that allows larger deviations from experimental values than the often used purely harmonic potential.

Final Protocol: Docking With CSP and A Combination of VEAN and SANI

From the results described above it is clear that VEAN has the advantage of better convergence (107 versus 65 structures for SANI in the lowest energy cluster), but SANI shows better consistency with the experimental RDC data. To combine these advantages, we defined a protocol in which we used intervector projection angle restraints (VEAN) in the first two stages of our docking protocol (rigid body EM and semi-flexible SA) and direct RDC restraints (SANI) in the final water refinement (for details, see Material and Methods); we refer to this as VEAN-SANI. In Figure 1(d), it can be seen that the convergence of this protocol is comparable to that of VEAN; the lowest energy cluster is the most highly populated with 129 members. Moreover, the structures satisfy the RDC restraints as well as the SANI structures: the Q-factor for the best 10 structures after water refinement is 0.15 ± 0.01; see also Table III and Figure 2. The structures have an average backbone RMSD from the average structure of 1.0 ± 0.3 Å to VEAN and 0.9 ± 0.2 Å to SANI, respectively. These differences are within the precision of the respective ensembles, indicating that the various structures calculated using RDCs (SANI, VEAN, and VEAN-SANI) are to a large extent similar.

Figure 3 shows a plot of the experimental versus calculated RDC values for the crystal and the VEAN-SANI structures, together with both structures color-coded ac-

cording to the differences between experimental and calculated RDCs. From these data, it is clear that our solution structure has a much better consistency with the RDC data. It can also be seen that the deviations in the crystal structure are not only present in loop regions but also in secondary structure elements, indicating that it is unlikely that they could originate from intramolecular flexibility only.

Validation Using ^{15}N-Relaxation Data

As an independent validation of the Ub_2 structures, we determined how well they agree with the ^{15}N-relaxation data. The crystal structure of Ub_2 and the ten final Ub_2 structures of each run were used as input for the computer program ROTDIF[25,38,39] that fits relaxation data to a fully anisotropic rotational diffusion model. The experimental data were also fitted for each domain separately, as controls. The results are summarized in Table V; details are presented in the Supplementary Material.

The orientation of the principal axes of the diffusion tensor with respect to the molecule is very similar in the SANI, VEAN, and VEAN-SANI ensembles (see Fig. 4). The CSP structures and the crystal structure resulted in a somewhat different orientation of the diffusion tensor, consistent with the different interdomain orientation for these two structures. It is worth mentioning that the Ub_2-diffusion tensor orientation derived from ^{15}N-relaxation data is in reasonable agreement with the tensor orientation predicted solely from hydrodynamics properties of our solution structures using HYDRONMR[44] (Supplementary Material).

We have also compared the orientation of the diffusion tensor axes of Ub_2 obtained from fitting both domains simultaneously with those derived from fitting each individual domain alone. For a proper alignment of the domains in a molecule, these three sets of axes should agree with each other. There is excellent agreement between the "both-domains"-axes and those for the proximal domain in all NMR-derived structures (Table V). For the distal domain, the agreement is very good for the VEAN-SANI ensemble, while there is a somewhat greater variation in the axes' orientation for the other structures, with CSP showing the least agreement. Note that the diffusion tensor for the distal domain appears more axially symmetric,[25] and therefore the orientation of its x- and y-axes (given by the angle γ) is less well defined, resulting in a greater variation in the γ angle among the structures. To better assess the agreement between the orientations of the diffusion tensors, we also determined the angles between the corresponding axes for every structure in each ensemble (Supplementary Material). The average values of the intervening angle between the z-axes of the diffusion tensors ("both-domains" vs. "distal-only") were $6 \pm 2°$ (VEAN-SANI), $8 \pm 1°$ (SANI), $8 \pm 3°$ (VEAN), and $14 \pm 5°$ (CSP), while the corresponding angles between the "both-domains" and the "proximal-only" diffusion tensors were $5 \pm 1°$ (VEAN-SANI), $4 \pm 1°$ (SANI), $7 \pm 5°$ (VEAN), and $13 \pm 6°$ (CSP). A comparison between the axes of the "distal-only" and the "proximal-only" diffusion tensors gives for the same intervening angle $9 \pm 3°$ (VEAN-SANI), $10 \pm 1°$ (SANI), $13 \pm 6°$ (VEAN), and $24 \pm 9°$ (CSP). For the crystal structure, the z-z angles were 14, 16, and 28° respectively.

The structures were then compared in terms of the residuals of fit, represented by the value of the target function, χ^2 (Table V). The average χ^2 value per degree of freedom, χ^2/df, for the core residues was 4.78 ± 0.88 (SD) for CSP, 4.78 ± 0.68 for VEAN, 3.37 ± 0.40 for SANI, and 3.34 ± 0.19 for VEAN-SANI. These values should be compared with 4.77 for the crystal structure. In order to assess how well the relative orientation of the two domains in the derived structures agrees with the ^{15}N-relaxation data, we also compared the residuals of fit for Ub_2 with the sum of χ^2 values obtained when fitting the same experimental data for the two Ub domains separately. Such a comparison is necessary because the total χ^2 could be affected by greater residuals of fit for the individual domains (as e.g. in the case of the distal domain). The ratios of the corresponding χ^2 values (Table V) indicate that all docked structures are in better agreement with the experimental relaxation data than the crystal structure. A statistical F-test[45] comparing the corresponding variances for CSP, VEAN, SANI, and VEAN-SANI, results in F-test values (Table IV, last column) below the critical value of 1.50755 at the 95% confidence level. This suggests that an increase in χ^2 when fitting both domains according to their relative orientation in a given structure versus fitting them separately is not statistically significant for the docked structures. In comparison, the F-test value for the crystal structure suggests that the observed increase in χ^2/df could be significant: the probability for such an increase to occur by chance is 0.04. The VEAN-SANI ensemble, which has the lowest χ^2/df values, agrees best with the experimental relaxation data. The SANI ensemble is of comparable quality (although it has a higher standard deviation), while the VEAN ensemble is in slightly lesser agreement with the ^{15}N-relaxation data, although it is still somewhat better than the crystal structure and the CSP-only structures.

Structural Analysis of Ub_2 Solution Structure: Quality of Structure

Based on the good convergence, the consistency with the RDC data, and the independent validation with the relaxation data, we conclude that the VEAN-SANI structures are our best representation of the solution structure of Ub_2. The final ensemble of the 10 lowest-energy structures is shown in Figure 5. In Table VI, we analyze the solution structure and compare it with the crystal structure (1AAR[29]). The solution structure has only a small number of violations of AIR restraints and RDC data. This is not true for the crystal structure, which has 10 violations of ambiguous interaction restraints and has a higher Q-factor (see Table III). To make sure that the solutions resulting from the docking with RDC-restraints, while in better agreement with the RDC and relaxation data, are not artificial in the sense that their geometry is worse, we

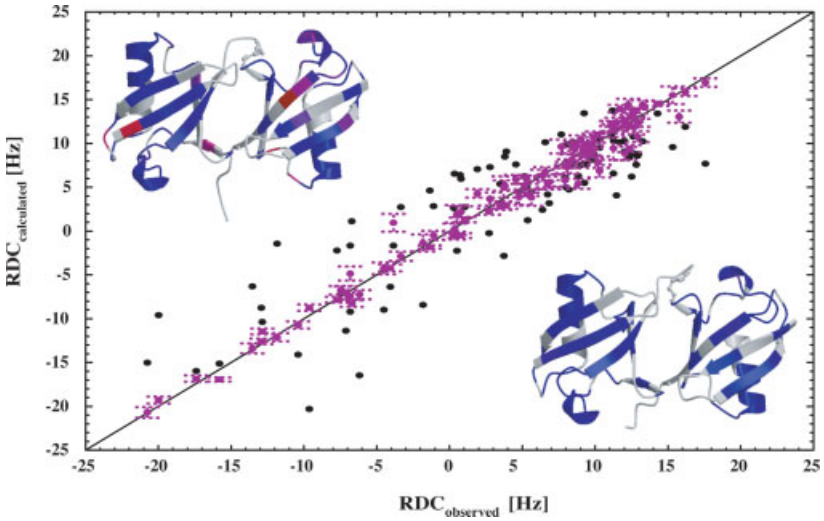

Fig. 3. Comparison of the experimentally measured dipolar couplings $RDC_{observed}$ with the back-calculated values using the program Pales ($RDC_{calculated}$) for the ensemble of the best 10 solution structures (purple, error bars are indicated) and for the crystal structure (black). Upper left corner: mapping on the crystal structure of the relative difference between experimental and calculated value of RDC ($abs(D^{obs} - D^{calc})/max[1.0,abs(D^{obs})]$. Residues are color-coded on a scale from blue (relative difference 0.0) to red (relative difference 6.1). White indicates that no RDCs were measured. The isopeptide bond is indicated in ball-and-stick. Lower right corner: idem, but for the solution structure (closest to the mean).

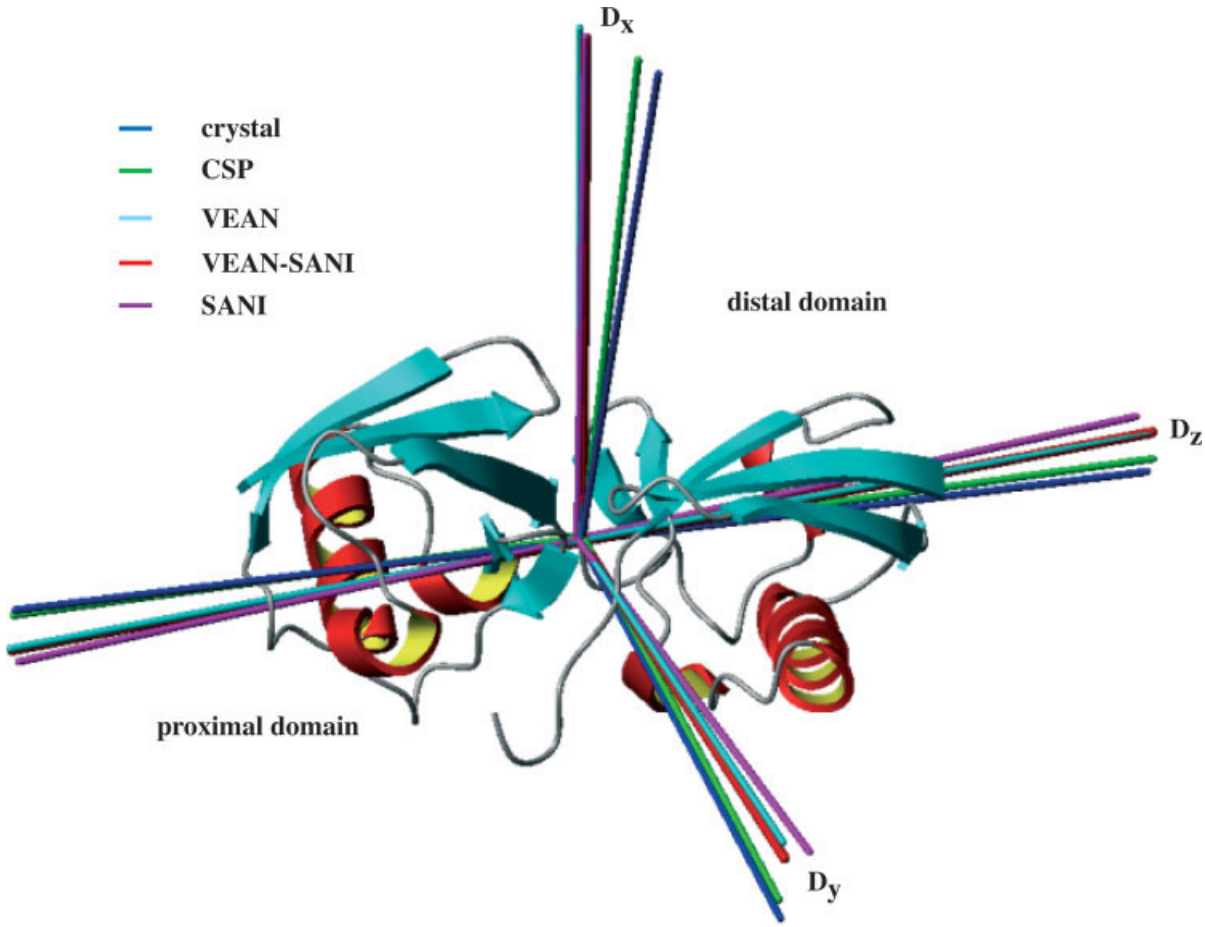

Fig. 4.

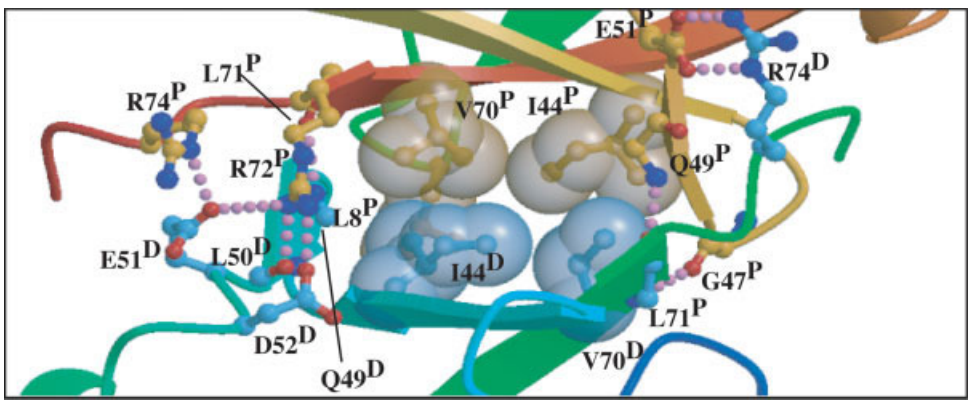

Fig. 7.

TABLE V. Summary of Results of ROTDIF Analysis of ^{15}N-Relaxation Data for Ub_2, Applied to Both Domains Simultaneously and to Each Ub Domain Separately

	Both Domains					Distal					Proximal						
Str	α^a	β^a	γ^a	χ^{2b}	$\chi^{2/df c}$	α^a	β^a	γ^a	χ^{2b}	$\chi^{2/df c}$	α^a	β^a	γ^a	χ^{2b}	$\chi^{2/df c}$	$\chi^2_{D\&P}/[\chi^2_{D_d}+\chi^2_P]$	F^e
CSP	−16	90	3	330.1	4.8	−23	103	−49	145.9	4.6	−4	90	8	90.2	2.9	1.39	1.27
	(3)	(3)	(4)	(60.7)	(0.9)	(3)	(6)	(27)	(19.6)	(0.6)	(8)	(2)	(3)	(19.3)	(0.6)	(0.16)	(0.14)
VEAN	15	85	182	330.0	4.8	14	77	160	134.2	4.2	13	87	181	127.1	4.1	1.27	1.16
	(4)	(2)	(10)	(46.9)	(0.7)	(4)	(3)	(29)	(26.6)	(0.8)	(8)	(4)	(10)	(40.3)	(1.3)	(0.10)	(0.09)
SANI	18	87	178	232.2	3.4	18	79	158	115.5	3.6	14	88	178	61.2	2.0	1.31	1.20
	(2)	(3)	(5)	(27.6)	(0.4)	(2)	(3)	(28)	(8.5)	(0.3)	(3)	(3)	(3)	(6.0)	(0.2)	(0.08)	(0.08)
VEAN-SANI	97	90	120	230.2	3.3	103	87	109	109.1	3.4	94	87	120	81.3	2.6	1.21	1.11
	(2)	(2)	(3)	(13.4)	(0.2)	(2)	(3)	(9)	(8.8)	(0.3)	(2)	(2)	(3)	(10.2)	(0.3)	(0.07)	(0.06)
CRYST	101	87	163	329.2	4.8	113	79	130	105.4	3.3	85	84	173	88.7	2.9	1.70	1.55

Shown are average values of the parameters over the corresponding ensemble of 10 selected structures (except for the crystal structure); the numbers in parentheses represent the standard deviations within each ensemble.

[a]Euler angles {α,β,γ}, in degrees, determine the orientation of the diffusion tensor frame with respect to the molecular coordinate frame of the specified Ub_2 structure.

[b]The values of the target function representing the residuals of fit using ROTDIF, see eq. (4).

[c]The value of the target function per degree of freedom.

[d]The ratio of χ^2 value from the simultaneous fit of both Ub domains, $\chi^2_{D\&P}$, to the sum of the residuals of fit when treating the two Ub domains separately (χ^2_D and χ^2_P). A value of this ratio close to 1 indicates that the relative orientation of the two domains in the analyzed structure agrees well with the relaxation data.

[e]The F-test value determined here as a ratio of the variance from the simultaneous fit of both domains, $var(D\&P) = \chi^2_{D\&P}/df_{D\&P}$, to the combined variance from fitting the individual domains, $var(D + P) = [\chi^D_2 + \chi^2_P]/[df_D + df_P]$, where df represents the number of degrees of freedom. At a 95% confidence level, the critical F-test value is 1.50755, for $df_{D\&P} = 69$, $df_D + df_P = 63$.

[f]Note that the difference in angles of the VEAN-SANI structures with respect to the SANI and VEAN structures is only the result of a different orientation of the input coordinates and that the diffusion tensors for the three sets of structures are nicely collinear (see Fig. 4).

analyzed the RMSD from idealized covalent geometry and the Ramachandran plot statistics with Procheck.[46] As shown in Table VI, the results are satisfactory. Although the crystal structure has a somewhat higher percentage of backbone dihedrals in the most favored region of the Ramachandran plot, the solution structure is still very good, with 100% of the residues having their backbone dihedrals in the most favored or additionally allowed regions of the Ramachandran plot. The RMSDs from idealized geometry are also very good, indicating that the additional orientational restraints do not result in local distortions of the covalent geometry. We also investigated whether the improvement in Q-factor for the structures docked using RDC restraints was due to local distortions of N-H bonds; this was completed by removing the hydrogens and adding them again in CNS without using experimental restraints. The Q-factor of the structures generated in this way was on average 0.19, compared to 0.15 for the refined structure, indicating again that the use of RDCs does not artificially distort the geometry of the structure (otherwise it would be expected that adding the N-H protons according to standard geometry would give much higher Q-factors).

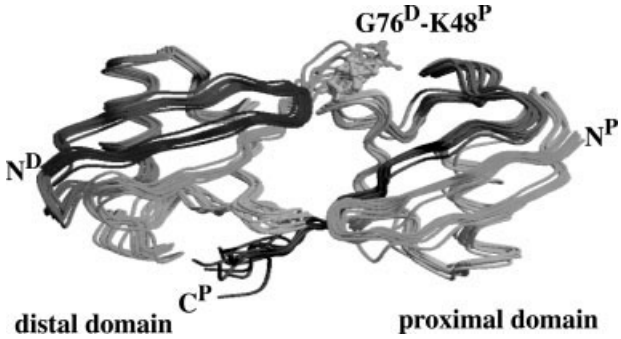

Fig. 5. Ensemble of 10 lowest-energy structures from the lowest-energy cluster for the solution structure (calculated using the final VEAN-SANI protocol) of Ub_2. The structures are fitted on the interface residues (see Table II); their orientation is the same as in Figure 3. The isopeptide bond and the N- and C-terminal residues of the distal (D) and proximal (P) domains are indicated. This figure was generated using MolScript[56] and Raster3D.[57]

Fig. 4. Orientation of the principal axes of the rotational diffusion tensor with respect to the molecular frame of Ub_2 for the various structural models discussed here. The diffusion tensors were derived from ^{15}N-relaxation data using the computer program ROTDIF.[25,38,39] The rods represent the average orientation of the tensor axes (X, Y, and Z indicated here as D_x, D_y, and D_z) for each ensemble of the docked structures and for the crystal structure. The rods coloring is as follows: red (VEAN-SANI), magenta (SANI), cyan (VEAN), green (CSP), blue (crystal). The Ub_2 structures from these ensembles were superimposed onto the representative VEAN-SANI structure (shown here as a ribbon cartoon) using the backbone atoms of the residues belonging to the elements of secondary structure. The figure was generated using MolMol.[55]

Fig. 7. Detailed view of the interface of the Ub_2 solution structure. Residues involved in hydrophobic non-bonded contacts (ball-and-stick and transparent CPK representation) or in inter-domain hydrogen bonds or salt-bridges (ball-and-stick representation) are shown (see also Supporting Table 3S; note that for a better visualization not all contacts are shown). Dotted lines represent hydrogen bonds. The residues are labeled with one-letter residue code and residue number, followed by D or P to indicate the distal or proximal domain, respectively. This figure was generated using MolScript[56] and Raster3D.[57]

Structural Analysis of Ub_2 Solution Structure: Orientation of the Two Domains

A rotation of the two domains (distal and proximal) relative to each other is observed when comparing solution

TABLE VI. Structural Statistics of Ub_2 Solution Structure and Comparison With Crystal Structure (PDB entry 1AAR[29])[a]

	Crystal	Solution Structure
Docking Statistics		
E_{vdw}[b] (kcal/mol)	—	−61 (9)
E_{elec}[b] (kcal/mol)		−425 (35)
Cluster-population[c]	—	129
AIR-energy (kcal/mol)	—	0.4 (0.9)
AIR-violations > 0.5 Å	10	0.3 (0.6)
AIR RMS (Å)	—	0.08 (0.1)
Structural Statistics		
RMSD backbone[d] (Å)	—	0.7 (0.2)
BSA[e] (Å^2)	1528	1749 (54)
Helix-angle[f] (°)	128	154 (6)
Deviations From Ideal Geometry		
RMSD-angles (°)	3.2	0.43 (0.01)
RMSD-bonds (Å)	0.018	0.003 (0.0001)
RMSD-impropers (°)	1.64	0.47 (0.02)
Ramachandran Analysis		
Most favored (%)	93.9	87.0
Additional allowed (%)	6.1	13.0
Generously allowed (%)	0.0	0.0
Disallowed (%)	0.0	0.0

[a]The reported values for the solution structure are averages over the 10 final structures with standard deviation indicated in brackets.
[b]The non-bonded energies E_{vdw} and E_{elec} were calculated with an 8.5 Å distance cutoff using the OPLS nonbonded parameters[53] from the parallhdg5.3.pro parameter file.[54]
[c]Number of structures in the lowest energy cluster (out of a total of 200 structures).
[d]Average RMSD from the average structure.
[e]Buried surface area (calculated with NACCESS[48]).
[f]The helix-angle is defined as the angle between the helices of the distal and proximal domains, calculated from the helix axis of residues 24–33 of each domain.

structure and crystal structure. This is confirmed by analysis with Dyndom,[47] which allows identification and characterization of domain motions. The rotation axis is 40% ± 10% a twist axis (normal to the interface) and 60 ± 10% a closure axis (parallel to the interface); the rotation around this axis is 20° (± 3°). Figure 6 gives a graphical representation of this rotation; the rotation axis as determined by Dyndom for the representative structure (closest to the mean) of the ensemble is shown, together with this structure and the crystal structure. We also calculated the angle between the helix axes in the distal and proximal domain: this angle is 128° in the crystal structure and 154° (±6°) in the solution structure. This difference in domain orientation between solution and crystal structure is also reflected in the RMSD value of 1.7 ± 0.3 Å between the two.

Structural Analysis of Ub_2 Solution Structure: The Interface

It was previously shown based on NMR data[25,38] that the functionally important residues Leu8, Ile44, and Val70 at the ubiquitin Ub_2 interface are not rigidly buried but that some interdomain dynamics are retained that could be sufficient to expose these groups and allow them to interact with other molecules. It was also suggested that the interface in the closed conformation (at pH 6.8, for which we determined the structure here) is somewhat more open than that in the crystal structure.

Our solution structure supports this latter conclusion. Specifically, we looked at the percentage of decrease in solvent accessibility (calculated using NACCESS[48]) observed upon going from the ubiquitin monomer to the Ub_2 structure. For Val70 (both distal and proximal domains) there is not much difference between the solution and the crystal structure; for both, the decrease in accessibility is between 95 and 100%. For Ile44 (both distal and proximal domains), in the crystal structure the accessibility decreases by more than 99%, whereas in the solution structure, Ile44 remains a little bit more accessible (decrease in accessibility of 94 ± 7%). The accessibility of Leu8 in the crystal structure is decreased by 65 and 70% for the distal and proximal domains, respectively, compared with only 52 ± 11% and 58 ± 7% for these domains in the solution structure. This shows that in the solution structures these hydrophobic residues are slightly more accessible than in the crystal structure. Note that these data also suggest that the interactions between the functionally important hydrophobic residues (Leu8, Ile44, and Val70) in Ub_2 and various recognition factors will involve significant interdomain motions, such as a transition between the closed

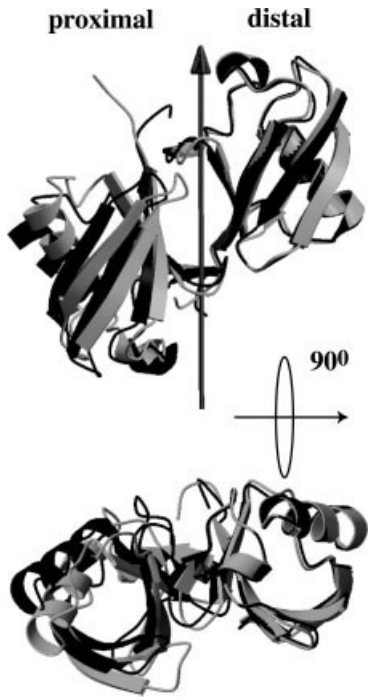

Fig. 6. Result of Dyndom[47] analysis, showing the rotation of the proximal domain with respect to the distal domain when comparing the representative solution structure (black) with the crystal structure (gray). The structures are fitted on the distal domain, and secondary structure elements are indicated. Two orthogonal views are shown, corresponding to a 90° rotation around a horizontal axis in the plane of the paper. The rotation axis as determined by Dyndom is indicated in red. This figure was generated using MolScript[56] and Raster3D.[57]

conformation (which we study here) and the open conformations of Ub_2.

We also analyzed the intermolecular contacts using the program DIMPLOT.[43] A number of non-bonded contacts were detected that are similar in the solution and the crystal structures. However, there are additional contacts that are specific for the solution structure (see Supplementary Material Table 3S). Figure 7 shows a detailed view of the interface of the solution structure.

Discussion

The structures obtained by docking using only CSP data satisfied neither RDC nor relaxation data. To some extent, this indicates that applying only CSP data to drive protein–protein docking should be done with care. However, it is still reassuring that the resulting CSP structures are quite close to the crystal structure (RMSD of 1.0 Å). In our opinion, this shows that the (easily available) CSP data are very valuable to get a fair representation of the solution structure of a protein–protein complex.

When introducing RDCs in the docking protocol, both SANI and VEAN gave good and comparable results for Lys48-linked Ub_2. In order to get an unbiased account of the way the structures fulfill the experimental RDC data, cross-validation was performed: 10% of the RDC data were excluded from the structure calculations; these data were subsequently used for independent validation. After cross-validation, the values of the Q-factor were 0.48 for SANI and 0.58 for VEAN. This shows that the direct RDC structure and the intermolecular intervector projection angle structure are of comparable quality. However, proper assessment would require full cross-validation. This was not done because it is computationally very expensive. We also checked whether the higher Q-factors of the structures calculated with intervector projection angle restraints compared to those obtained with direct RDC restraints were due to the force constants for the VEAN energy term being too small: we repeated the docking with a value two times larger for the VEAN force constants. This did not result in a better agreement with the experimental data (results not shown), indicating that looser restraining might well be an intrinsic property of this type of restraint: the accumulation of errors when taking pairs of RDCs could lead to less restrictive restraints.

The two approaches (SANI and VEAN) were combined, using intervector projection angle restraints in the first stages of the calculations to benefit from better convergence, and direct RDC restraints in the final explicit solvent refinement to fine-tune the structures. The structures calculated with this final protocol better fit the experimental RDC restraints compared to the structures calculated using only CSPs or the crystal structure. This is to be expected since those structures have been optimized against these RDC data. The anisotropic relaxation data provide, however, an independent validation; the structures calculated using the RDCs also better fit those data. It was estimated previously,[25] on the basis of RDC and relaxation data, that the angle between the α-helices in the two domains should be between 161° (relaxation data) and 145° (RDC data) for the structure in solution, compared to 128° in the crystal. In agreement with this, the structures obtained here have an average angle between the α-helices of 154°.

As previously discussed,[25,38] the interface in Lys48-linked Ub_2 is not rigidly locked, which allows ligands to compete with the Ub–Ub interactions.[49] A model based on chemical shift titration suggests that at pH 6.8 Lys48-linked Ub_2 is in dynamic equilibrium between the predominantly populated (>85%) "closed" state and a minor, "open" state. This equilibrium to some extent represents the on/off events in protein–protein interactions. It is important to note that the so-called "open" state of Lys48-linked Ub_2 at pH 6.8 is a compact conformation, probably characterized by somewhat weaker interdomain contacts compared to the "closed" state but not very different structurally. This is supported by recent small-angle X-ray scattering data[50] that indicate a mono-dispersed compact globular conformation of Lys48-linked Ub_2 under similar conditions in solution.

From the RDC data alone, it is difficult to determine whether this solution structure at pH 6.8 represents a single static conformation or results from motion averaging of RDCs; a similar problem has been noted in the case of other proteins.[14,51] It is not clear to what extent the measured RDCs are affected by the dynamic equilibrium in Ub_2. Several interface residues in Ub_2 exhibit line-broadening indicative of conformational exchange in a microsecond–millisecond time range,[25,38] although it is not possible to determine the amplitudes of these motions from the relaxation data. An estimate of the effect of motions on RDC measurements can be obtained by calculating the 'internal generalized degree of order' as proposed by Tolman et al.;[52] its value is correlated with possible structural distortions due to dynamics. In our case, for the distal and proximal domains, the values ranged from 0.96 to 1.0, indicating only very small structural distortions due to motion, if any. In addition, independent validation using ^{15}N-relaxation data, which are probing dynamics in a much narrower time range compared to RDCs, clearly indicates that both RDC and relaxation data are affected by motions in Ub_2 in approximately the same way. This suggests that whatever conformational averaging is present, it happens primarily on the time scale of the overall rotational diffusion or faster. If dynamic events on a slower time scale would influence the structure calculated here, one would expect a statistically significant increase in the value of χ^2 when comparing the fits of relaxation data for the distal or proximal domain with di-ubiquitin. However, this is not the case. All this indicates (1) that the conformational equilibrium between open and closed conformations on a slow time scale does not significantly affect the measured RDCs and thus our proposed model and (2) that the value of ≈15% for the population of the open state is most likely an overestimate. It is not yet possible to rule out the possibility that our structure reflects conformational averaging on a ≈10 ns or faster time scale.

CONCLUSIONS

Our results show that a docking method based on CSP data, HADDOCK,[4] can successfully be combined with RDCs following a combined strategy in which RDCs are first introduced as intervector projection angle restraints and at the later refinement stage as direct restraints. The use of intervector projection angles has some definite advantages. First, it leads to a better convergence of the calculations, which in turn gives more confidence in the results, and also means that structural characteristics (like interdomain orientation in our case) can be calculated more precisely because relevant parts of the configuration space are sampled more extensively. The better convergence of the docking run using intervector projection angle restraints could be due to the fact that it is possible to distinguish between intermolecular and intramolecular restraints; by using only the intermolecular restraints in the rigid-body docking phase, it is possible to define directly the relative orientation of the two components. Finally, by avoiding the use of an external tensor in structure calculations, intervector projection angles offer a convenient means to combine RDC data obtained from different alignment media or simultaneously refine structures against both RDCs and relaxation data.

ACKNOWLEDGMENTS

This work was supported by NIH grant GM65334 to D.F., and a "Jonge Chemici" grant from the Netherlands Organization for Scientific Research (N.W.O.) to A.B. We thank Charles Geraghty (University of Maryland) for performing hydrodynamic calculations of the rotational diffusion tensor, and Ranjani Varadan (University of Maryland) and Cyril Dominguez (Utrecht University) for data and helpful discussions.

REFERENCES

1. Alm E, Arkin AP. Biological networks. Curr Opin Struct Biol 2003;13(2):193–202.
2. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: An overview of search algorithms and a guide to scoring functions. Proteins 2002;47(4):409–443.
3. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ,. CAPRI: a Critical Assessment of PRedicted Interactions. Proteins 2003;52(1):2–9.
4. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 2003;125(7):1731–1737.
5. Prestegard JH, al_Hashimi HM, Tolman JR. NMR structures of biomolecules using field oriented media and residual dipolar couplings. Q Rev Biophys 2000;33(4):371–424.
6. de Alba E, Tjandra N. NMR dipolar couplings for the structure determination of biopolymers in solution. Prog NMR Spectrosc 2002;40(2):175–197.
7. Bax A. Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 2003;12(1):1–16.
8. Tolman JR. Dipolar couplings as a probe of molecular dynamics and structure in solution. Curr Opin Struct Biol 2001;11(5):532–539.
9. Bax A, Kontaxis G, Tjandra N. Dipolar couplings in macromolecular structure determination. Methods Enzymol 2001;339:127–174.
10. Fushman D, Cowburn D. Characterization of inter-domain orientations in solution using the NMR relaxation approach. In: Krishna NR, editors. Protein NMR for the millennium. Kluwer; 2002. p 53-78.
11. Zuiderweg ER. Mapping protein-protein interactions in solution by NMR spectroscopy. Biochemistry 2002;41(1):1–7.
12. Fischer MWF, Losonczi JA, Weaver JL, Prestegard JH. Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. Biochemistry 1999;38(28):9013–9022.
13. Fushman D, Xu R, Cowburn D. Direct determination of changes of interdomain orientation on ligation: Use of the orientational dependence of N-15 NMR relaxation in Abl SH(32). Biochemistry 1999;38(32):10225–10230.
14. Skrynnikov NR, Goto NK, Yang DW, Choy WY, Tolman JR, Mueller GA, Kay LE. Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: Differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. J Mol Biol 2000;295(5):1265–1273.
15. Mollova ET, Hansen MR, Pardi A. Global structure of RNA determined with residual dipolar couplings. J Am Chem Soc 2000;122(46):11561-11562.
16. Bayer P, Varani L, Varani G. Refinement of the structure of protein-RNA complexes by residual dipolar coupling analysis. J Biomol NMR 1999;14(2):149–155.
17. Clore GM. Accurate and rapid docking of protein-protein complexes on the basis of intermolecular nuclear overhauser enhancement data and dipolar couplings by rigid body minimization. Proc Natl Acad Sci USA 2000;97(16):9021–9025.
18. McCoy MA, Wyss DF. Structures of protein-protein complexes are docked using only NMR restraints from residual dipolar coupling and chemical shift perturbations. J Am Chem Soc 2002;124(10): 2104-2105.
19. Xu XP, Case DA. Automated prediction of N-15, C-13(alpha), C-13(beta) and C-13′ chemical shifts in proteins using a density functional database. J Biomol NMR 2001;21(4):321–333.
20. Clore GM, Schwieters CD. Docking of protein-protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from 1H/15N chemical shift mapping and backbone 15N–1H residual dipolar couplings using conjoined rigid body/torsion angle dynamics. J Am Chem Soc 2003;125(10):2902–2912.
21. Dobrodumov A, Gronenborn AM. Filtering and selection of structural models: combining docking and NMR. Proteins 2003;53(1): 18–32.
22. Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 1997;272(1):106–120.
23. Meiler J, Blomberg N, Nilges M, Griesinger C. A new approach for applying residual dipolar couplings as restraints in structure elucidation. J Biomol NMR 2000;16(3):245–252.
24. Tjandra N, Garrett DS, Gronenborn AM, Bax A, Clore GM. Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. Nat Struct Biol 1997;4(6):443–449.
25. Varadan R, Walker O, Pickart C, Fushman D. Structural properties of polyubiquitin chains in solution. J Mol Biol 2002;324(4):637–647.
26. Weissman AM. Themes and variations on ubiquitylation. Nat Rev Mol Cell Bio 2001;2(3):169–178.
27. Glickman MH, Ciechanover A. The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. Physiol Rev 2002;82(2):373–428.
28. Pickart CM. Mechanisms underlying ubiquitination. Annu Rev Biochem 2001;70:503–533.
29. Cook WJ, Jeffrey LC, Carson M, Chen Z, Pickart CM. Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). J Biol Chem 1992;267(23):16467–16471.
30. Phillips CL, Thrower J, Pickart CM, Hill CP. Structure of a new crystal form of tetraubiquitin. Acta Crystallogr D Biol Crystallogr 2001;57(Pt 2):341–344.
31. Cook WJ, Jeffrey LC, Kasperek E, Pickart CM. Structure of tetraubiquitin shows how multiubiquitin chains can be formed. J Mol Biol 1994;236(2):601–609.
32. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28(1):235–242.
33. Varadan R, Assfalg M, Haririnia A, Raasi S, Pickart C, Fushman D. Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. J Biol Chem 2004;279(8):7055–7063.
34. Zweckstetter M, Bax A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. J Am Chem Soc 2000;122(15):3791-3792.

35. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. Use of dipolar 1H–15N and 1H–13C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Biol 1997;4(9):732–738.
36. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse_Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998; 54(Pt 5):905–921.
37. Meiler J, Blomberg N, Nilges M, Griesinger C. A new approach for applying residual dipolar couplings as restraints in structure elucidation. J Biomol NMR 2000;17(2):185.
38. Fushman D, Varadan R, Assfalg M, Walker O. Determining domain orientation in macromolecules by using spin-relaxation and residual dipolar coupling measurements. Prog NMR Spectrosc 2004;44:189–214.
39. Walker O, Varadan R, Fushman D. Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from 15N relaxation data using computer program ROTDIF. J Magn Reson 2004;168:336–345.
40. Fushman D, Tjandra N, Cowburn D. Direct measurement of N-15 chemical shift anisotropy in solution. J Am Chem Soc 1998;120(42): 10947–10952.
41. Fushman D, Tjandra N, Cowburn D. An approach to direct determination of protein dynamics from N-15 NMR relaxation at multiple fields, independent of variable N-15 chemical shift anisotropy and chemical exchange contributions. J Am Chem Soc 1999;121(37):8577–8582.
42. Cornilescu G, Marquardt JL, Ottiger M, Bax A. J Am Chem Soc 1998;120(27):6836-6837.
43. Wallace AC, Laskowski RA, Thornton JM. Ligplot - a program to generate schematic diagrams of protein ligand interactions. Protein Eng 1995;8(2):127–134.
44. de la Torre JG, Huertas ML, Carrasco B. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. J Magn Reson 2000;147(1): 138–146.
45. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in C. New York: Cambridge UP; 1992. p 994.
46. Laskowski R, MacArthur M, Moss D, Thornton J. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–291.
47. Hayward S, Berendsen HJC. Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. Proteins 1998;30(2):144–154.
48. Hubbard SJ, Thornton JM. NACCESS. London: Department of Biochemistry and Molecular Biology, University College; 1993.
49. Verma R, Peters NR, D'Onofrio M, Tochtrop GP, Sakamoto KM, Varadan R, Zhang MS, Coffino P, Fushman D, Deshaies RJ, King RW. Ubistatins inhibit proteasome-dependent degradation by binding the ubiquitin chain. Science 2004;306(5693):117–120.
50. Tenno T, Fujiwara K, Tochio H, Iwai K, Morita EH, Hayashi H, Murata S, Hiroaki H, Sato M, Tanaka K, Shirakawa M. Structural basis for distinct roles of Lys63- and Lys48-linked polyubiquitin chains. Genes Cells 2004;9(10):865–875.
51. Lukin JA, Kontaxis G, Simplaceanu V, Yuan Y, Bax A, Ho C. Quaternary structure of hemoglobin in solution. Proc Natl Acad Sci USA 2003;100(2):517–520.
52. Tolman JR, Al-Hashimi HM, Kay LE, Prestegard JH. Structural and dynamic analysis of residual dipolar coupling data for proteins. J Am Chem Soc 2001;123(7):1416–1424.
53. Jorgensen WL, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 1998;110:1657–1666.
54. Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M. Refinement of protein structures in explicit solvent. Proteins 2003;50(3):496–506.
55. Koradi R, Billeter M, Wuthrich K. MOLMOL: A program for display and analysis of macromolecular structures. J Mol Graphics 1996;14(1):51.
56. Kraulis PJ. MOLSCRIPT: A program to produce both detailled and schematic plots of protein structures. J Appl Cryst 1991;24: 946–950.
57. Merrit EA, Murphy MEP. Raster3D version 2.0: A program for photorealistic molecular graphics. Acta Cryst 1994;D50:869–873.