J|A|C|S
ARTICLES
Published on Web 10/25/2002

Direct Structure Determination Using Residual Dipolar Couplings: Reaction-Site Conformation of Methionine Sulfoxide Reductase in Solution

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

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

Received May 13, 2002

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

Classical NMR structure determination requires extensive assignment of backbone and side-chain resonances followed by unambiguous identification of nuclear Overhauser effect (nOe) correlations between these assigned frequencies, before even low-resolution structural models can be derived.[1] Complete structural analysis can thus be particularly time-consuming for large proteins. In contrast, residual dipolar couplings (RDC) in macromolecules aligned in dilute liquid crystalline media[2−4] can be measured routinely from pairs of nuclei distributed throughout the molecule immediately following assignment of the backbone resonances. RDC provide coherent, long-range structural data, allowing entirely new approaches to structural biology in solution.[5−8] This coherence makes RDC particularly appropriate for comparison with available structural homologues, enabling rapid validation of initial molecular models or identification of folds from database structures[9−11] and subsequent refinement of these initial models.[12] Although this approach requires the presence of a structural homologue in accessible databases, the identification of even a low-resolution molecular fold from primary data is a major breakthrough for solution-state NMR.

* Corresponding author: telephone (33) 4 38 78 95 54; fax (33) 4 38 78 54 94; e-mail martin@rmn.ibs.fr.
[†] Institut de Biologie Structurale, Jean-Pierre Ebel, CNRS-CEA.
[‡] Laboratoire de Chimie Bacterienne, CNRS, Institut Biologie Structurale et Microbiologie.

(1) Wüthrich, K. *NMR of proteins and nucleic acids*; Wiley: New York, 1986.
(2) Gayathri, C.; Bothner-by, A. A.; van Zijl, P. C. M.; Maclean, C. *Chem. Phys. Lett.* **1982**, *87*, 192−196.
(3) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279−9283.
(4) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.
(5) Prestegard, J. H. *Nat. Struct. Biol.* **1998**, *5*, 517−522.
(6) Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5891−5898.
(7) Tjandra, N. *Structure* **1999**, *7*, R205−R211.
(8) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127−174.
(9) Annila, A.; Aitio, H.; Thulin, E.; Drakenberg, T. *J. Biomol. NMR* **1999**, *14*, 223−230.
(10) Meiler, J.; Peti, W.; Griesinger, C. *J. Biomol. NMR* **2000**, *17*, 283−294.
(11) Andrec, M.; Du, P.; Levy, R. M. *J. Am. Chem. Soc.* **2001**, *123*, 1222−1229.
(12) (a) Chou, J. J.; Li, S.; Bax, A. *J. Biomol. NMR* **2000**, *18*, 217−227. (b) Champier, L.; Sibille, N.; Bersch, B.; Brutscher, B.; Blackledge, M.; Coves, J. *Biochemistry* **2002**, *41*, 3770−3780.

In cases where sufficient measurements can be made from throughout the peptide chain in the presence of two alignment media, protein structure can be determined from RDC alone, either in comparison with short fragments from structural databases[13] or by sequentially building the peptide chain from unambiguously oriented peptide units.[14] The requirement of continuous data from peptide planes and/or tetrahedral junctions has so far limited the number of applications of these *de novo* methods to particularly well-behaved systems. In the absence of continuous RDC data, locally well-defined substructures may be isolated in the primary sequence and cannot easily be placed with respect to the rest of the molecule unless additional structural information, for example long-range nOes, is also available.[15]

Methionine sulfoxide reductase (MsrA) catalyzes the reduction of free and protein-bound oxidized methionine residues and is consequently required for the repair of important enzymes.[16−18] The enzyme, found in nearly all living organisms, is present at high concentration in tissues and cells susceptible to oxidative stress damage[19] and has thus been linked to aging processes, lifespan determination,[20] and diverse degenerative pathologies such as Alzheimer's disease.[21] Two crystal structures of MsrA have recently been elucidated: MsrA from *Escherichia coli*[22] (MsrA^{Ecoli}) and *Bos taurus*[23] (MsrA^{bov}). Although very similar, these structures differ significantly in the C-terminal strand involved in freeing the catalytic site for further function. To further understand the mechanism by which MsrA provides protection against oxidative stress, we have studied the solution structure of reduced MsrA from *Erwinia chrysanthemi* (MsrA^{Echmi}, 221 amino acids), a plant pathogen that requires the enzyme for full virulence,[24] using RDC measured in partially aligned liquid crystal media.

To characterize the catalytically important C-terminal loop domain of MsrA^{Echmi}, we have developed a novel combination of the two RDC-based approaches presented above: The tertiary fold is initially validated by RDC, providing low-resolution structural information and positioning the molecule in the common alignment frames. We then focus on the region of interest, which is determined from RDC alone by the *meccano* approach. This region can be unambiguously positioned relative to the remaining molecular scaffold by use of the graphic modeling tool Module, a program specifically developed for RDC analysis to facilitate the manipulation of oriented molecular domains relative to a common alignment frame.[25]

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

Methods

Samples of ^{15}N-labeled and ^{13}C,^{15}N,^{2}D-labeled MsrA^{Echmi} were prepared as previously described.[26] Both samples were studied under reducing conditions with 10 mM dithiothreitol.

Residual Dipolar Coupling Measurement. All NMR experiments were performed on a Varian Inova 600 spectrometer, equipped with a triple-resonance (^{1}H, ^{15}N, ^{13}C) probe and shielded *z*-gradients. Sample temperature was set to 298 K for all studies. RDC were collected on a 1.0 mM uniformly ^{13}C-, ^{15}N-, and 82% ^{2}H-labeled sample suspended in a liquid crystalline medium consisting of 20 mg/mL of the filamentous phage Pf1 (Asla Ltd., Riga, Latvia) in 80 mM potassium phosphate buffer, pH 7.5. RDC were also collected on an identical sample suspended in a liquid crystalline medium consisting of 5% C12E6/hexanol[27] in 25 mM potassium phosphate buffer, pH 7.5. Four different types of dipolar couplings were measured as previously described[28,29] at 600 MHz, in both media: $^{1}D_{NH}$, $^{1}D_{C'C\alpha}$ and $^{2}D_{C'HN}$ couplings were obtained from 3D HNCO-type experiments, and $^{1}D_{C\alpha C\beta}$ were measured in a 3D HNcoCA-type experiment. Data processing and peak-picking were performed with FELIX version 2000 (Accelrys Inc.).

Relaxation Experiments and Data Analysis. The ^{15}N R_1 and $R_{1\rho}$ relaxation and ^{1}H−^{15}N nOe measurements were performed on the ^{15}N-labeled sample (1.4 mM concentration in 25 mM potassium phosphate buffer) at ^{1}H frequency of 600 MHz and temperature of 298 K with the classical ^{1}H-detected pulse sequence based on established methods[30] and described in detail elsewhere.[31] Data were analyzed by the model-free Lipari−Szabo approach with the program TENSOR2.[32]

Fold Validation. The crystal structure of MsrA^{Ecoli} (molecule A in PDB file 1FF3) was used to analyze and validate the solution structure of MsrA^{Echmi}. Secondary structure was identified from experimental ^{13}C chemical shifts from random-coil values (chemical shift index, CSI[33]), and the intensity distribution of short- and medium range nOes involving H^N protons.

The program Module[25] was used to evaluate the accordance between the MsrA^{Ecoli} structure and the experimental RDC data. Residual dipolar couplings can be expressed in terms of the orientation $\{\theta, \phi\}$ of the internuclear vector relative to a common alignment tensor for the molecule:

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

where A_a and A_r are the axial and rhombic components of the alignment tensor, r_{ij} is the internuclear distance, and S is the order parameter.

(13) Delaglio, F.; Kontaxis, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 2142−2143.
(14) Hus, J. C.; Marion, D.; Blackledge, M. *J. Am. Chem. Soc.* **2001**, *123*, 1541−1542.
(15) Fowler, C. A.; Tian, F.; Al-Hashimi, H. A.; Prestegard, J. H. *J. Mol. Biol.* **2000**, *304*, 447−460.
(16) Davis, D. A.; Newcomb, F. M.; Moskovitz, J.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J.; Fales, H. M.; Levine, R. L.; Yarchoan; R. *Biochem. J.* **2000**, *346*, 305−311.
(17) Sigalov, A. B.; Stern, L. J. *FEBS Lett.* **1998**, *433*, 196−200.
(18) Sun, H.; Gao, J.; Ferrington, D. A.; Biesiada, H.; Williams, T. D.; Squier, T. C. *Biochemistry* **1999**, *38*, 105−112.
(19) Brot, N.; Weissbach, H. *Biopolymers* **2000**, *55*, 288−296.
(20) Moskovitz, J.; Bar-Noy, S.; Williams, W. M.; Requena, J.; Berlett, B. S.; Stadtman, E. R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12920−12925.
(21) Gabbita, S. P.; Aksenov, M. Y.; Lovell, M. A.; Markesbery, W. R. *J. Neurochem.* **1999**, *73*, 1660−1666.
(22) Tete-Favier, F.; Cobessi, D.; Boschi-Muller, S.; Azza, S.; Branlant, G.; Aubry, A. *Structure* **2000**, *8*, 1167−1178.
(23) Lowther, W. T.; Brot, N.; Weissbach, H.; Matthews, B. W. *Biochemistry* **2000**, *39*, 13307−13312.
(24) Hassouni, M. E.; Chambost, J. P.; Expert, D.; Van Gijsegem, F.; Barras, F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 887−892.
(25) Dosset, P.; Hus, J. C.; Marion, D.; Blackledge, M. *J. Biomol. NMR* **2001**, *20*, 223−231.
(26) Béraud, S.; Chambost, J. P.; Bersch, B.; Gans, P.; Barras, F.; Marion, D. *J. Biomol. NMR* **2001**, *20*, 97−98.
(27) Rückert, M.; Otting, G. *J. Am. Chem. Soc.* **2000**, *122*, 7793−7797.
(28) Yang, D.; Venters, R. A.; Mueller, G. A.; Choy, W. Y.; Kay, L. E. *J. Biomol. NMR* **1999**, *14*, 333−343.
(29) Brutscher, B. *J. Magn. Reson.* **2001**, *151*, 332−338.
(30) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T., Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984−6003.
(31) Tsan, P.; Hus, J.-C.; Caffrey, M.; Marion, M.; Blackledge, M. *J. Am. Chem. Soc.* **2000**, *122*, 5603−5612.
(32) Dosset, P.; Hus, J.-C.; Blackledge, M.; Marion, D. *J. Biomol. NMR* **2000**, *16*, 23−28.
(33) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **2001**, *239*, 363−392.

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

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

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

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

De Novo Structure Calculation. The recently published *meccano* approach[14] was slightly modified. The molecular dynamics (MD) program Sculptor[34] was incorporated into the sequential positioning algorithm designed to place the peptide units. This provides randomized sampling of conformational space, which facilitates inspection of the conformational precision in the final ensemble. The use of a molecular dynamics force field also allows the straightforward introduction of repulsive nonbonding interactions. In this case a specific force field was written to include only experimental RDC, reinforced backbone covalent terms, and simple repulsive nonbonding terms. Only (C′, C^α, C^β, H^α, H^N, N) atoms were used in the calculation.

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

Docking the Meccano Peptide to the Protein Scaffold. The $MsrA^{Echmi}$ P196-L202 peptide was positioned relative to the crystal coordinates of $MsrA^{Ecoli}$ with the molecular modeling tools available in the program Module. This program allows positioning of different oriented modules using only the equivalent orientations and three-dimensional translational freedom available in a common alignment frame. The crystal structure and peptide were aligned relative to their common frame and the peptide was positioned laterally such that the expected covalence was best satisfied. Module works in a single-alignment tensor subspace, so the procedure was repeated for both datasets to verify the common solution.

All structure comparisons of the aligned peptide were performed with the rms difference between the coordinates in the common alignment frame rather than superimposing the coordinates to a best fit, as is the standard procedure, to ensure that the orientational information is preserved. The crystal structure of $MsrA^{bov}$ (molecule B; PDB code 1FVB) was compared to the *meccano*-constructed $MsrA^{Echmi}$ P196-L202 peptide.

The chimeric structure, containing the *meccano* peptide with all atoms present and the backbone conformation of the $MsrA^{Ecoli}$ crystal structure, was refined by low-temperature (300 K) RDC-restrained MD with Sculptor to regularize covalence. During this calculation, carried out with the force field AMBER4, the backbone atoms of the $MsrA^{Ecoli}$ were tethered to their crystal structure conformation by use of the additional force-field potential:

$$E_{\text{teth}} = k_{\text{teth}} \sum_i \sqrt{(x_i - x_i^0)^2/N} \quad (3)$$

with a k_{teth} value of 20 kcal·mol^{-1}·Å^{-2}. All measured RDCs were used as standard restraints in the presence of the two independent tensors.

Results and Discussion

We have assigned the backbone resonances[26] and measured extensive dipolar couplings ($N-H^N$, $C'-H^N$, $C^\alpha-C'$, and $C^\alpha-C^\beta$) from the reduced form of ^{13}C-, ^{15}N-, and ^2H-labeled $MsrA^{Echmi}$ in two liquid crystalline solutions. These couplings were initially compared to expected values from the recently published crystal structure of the 75% identical primary sequence homologue $MsrA^{Ecoli}$ by use of the alignment tensor optimization and molecular modeling program Module. The reduced form of $MsrA^{Ecoli}$ consists of an α/β roll core structure, comprising 80/123 residues in secondary structural elements, and predominantly coil N- and C-terminal regions wrapped around this core. The 45 amino acid C-terminal region is of particular interest as two of the three cysteines, Cys200 and Cys208, present in the proposed catalytic cycle[36,37] are found in this strand ($MsrA^{Echmi}$ numbering is used unless otherwise stated). This cycle involves initial nucleophilic attack of Cys53 on Met-SO, followed by a disulfide cascade implicating Cys53−Cys200 and then Cys200−Cys208 disulfide bridges, the latter step freeing the catalytic sulfur (Cys53) for further function.

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

(34) Hus, J. C.; Marion, D.; Blackledge, M. *J. Mol. Biol.* **2000**, *298*, 927−936.
(35) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 4.0*, University of California, San Francisco, CA, 1991.
(36) Boschi-Muller, S.; Azza, S.; Sanglier-Cianferani, S.; Talfournier, F.; Van Dorsselear, A.; Branlant, G. *J. Biol. Chem.* **2000**, *275*, 35908−35913.
(37) Lowther, W. T.; Brot, N.; Weissbach, H.; Honek, J. F.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6463−6468.

Table 1. Alignment Tensor Characteristics in the Two Liquid Crystal Media[a]

medium		N[b]	A_a (10^{-4})	A_r (10^{-4})	α (deg)	β (deg)	γ (deg)	χ^2	χ^2/N
phages[c]	α-helices[d]	144	13.8 ± 0.3	6.0 ± 0.3	110 ± 5	157 ± 2	126 ± 5	139	0.96
	β-sheets	110	14.8 ± 0.4	6.4 ± 0.5	108 ± 6	156 ± 3	128 ± 4	228	2.07
	C-loop[e]	204	13.2 ± 0.2	6.2 ± 0.3	109 ± 4	157 ± 2	130 ± 3	291	1.43
C12E6/hexanol[f]	α-helices	150	11.5 ± 0.3	7.3 ± 0.3	140 ± 2	87 ± 2	161 ± 2	140	0.93
	β-sheets	106	11.8 ± 0.3	6.9 ± 0.3	139 ± 5	88 ± 2	162 ± 3	197	0.96
	C-loop	210	10.9 ± 0.2	6.3 ± 0.2	140 ± 2	87 ± 2	162 ± 2	355	1.69

[a] $^1D_{NH}$, $^1D_{C'C\alpha}$, $^2D_{C'HN}$ and $^1D_{C\alpha C\beta}$ RDC from MsrA^{Echmi} were fitted to the conformation of MsrA^{Ecoli} by use of the program Module. [b] Number of couplings present in the structural motif. [c] All secondary structural elements were also fitted together: in this case $A_a = (14.0 \pm 0.3) \times 10^{-4}$, $A_r = (6.4 \pm 0.3) \times 10^{-4}$, and $\chi^2 = 380$ (254 couplings). [d] α-helices (56−62, 85−89, 110−119, and 145−164) and β-sheets (45−50, 68−75, 95−104, 140−142, 175−177, and 183−185) using the numbering of the MsrA^{Echmi} sequence. [e] C-loop region contains the C-terminal helix and the coil region until C208, but without the P196-L202 RDC data. [f] All secondary structural elements were also fitted together: in this case $A_a = 11.62 \times 10^{-4}$, $A_{r=} = 6.89 \times 10^{-4}$, and $\chi^2 = 356$ (256 couplings considered).

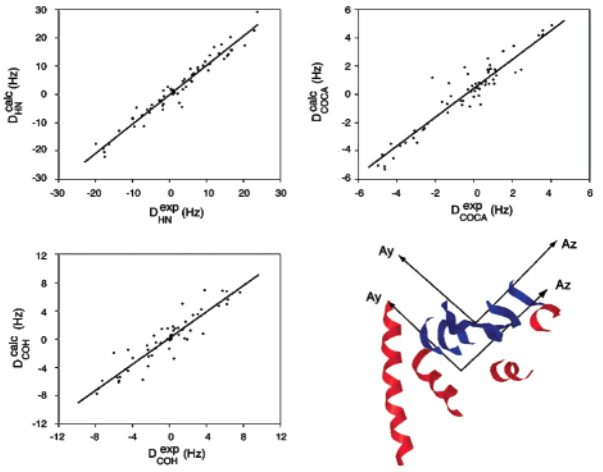

Figure 1. Comparison of measured RDC from MsrA^{Echmi} with expected values for the crystal structure coordinates from MsrA^{Ecoli}. (a, bottom right) Relative orientation of the alignment tensors for the different secondary structure elements in MsrA^{Ecoli} for the RDC from the C12E6/hexanol aligned sample. In this figure the alignment tensors of the central β-sheet (46−50, 68−75, 95−103, 139−141, 174−176, and 182−184), and α-helices (55−62, 84−88, 109−118, and 145−164) are compared relative to the crystal structure. The axial and rhombic components and χ^2 of the five secondary structure elements are given in Table 1. The different secondary structure elements were also fitted as a single domain using all couplings simultaneously as shown in panels b−d. (b, top left) Correlation between D_{NH}^{exp} and D_{NH}^{calc} in C12E6/hexanol for all core secondary structural elements shown in panel a. (c, bottom left) Correlation between $^2D_{C'HN}^{exp}$ and $^2D_{C'HN}^{calc}$ in C12E6/hexanol for all core secondary structural elements. (d, top right) Correlation between $^1D_{C'C\alpha}^{exp}$ and $^1D_{C'C\alpha}^{calc}$ in C12E6/hexanol for all core secondary structural elements.

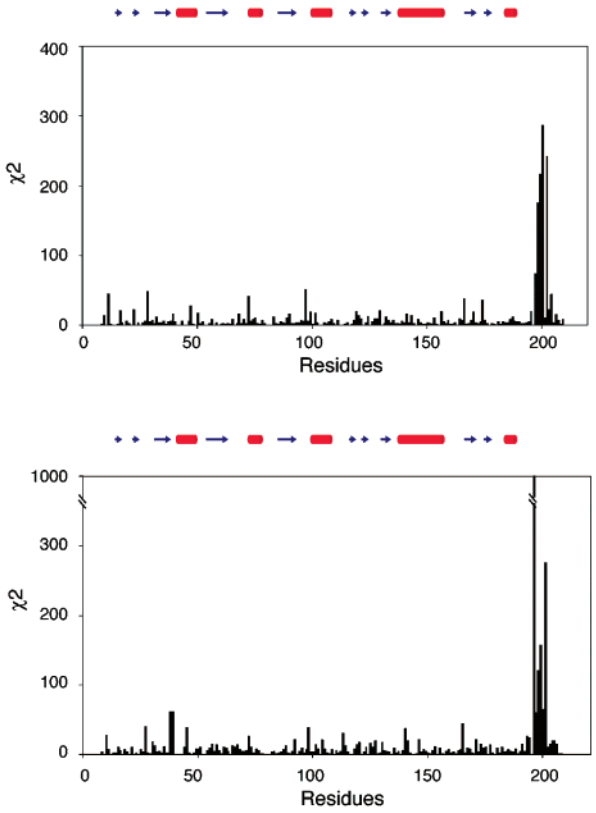

Figure 2. Local quality factor of the fit of RDC data from MsrA^{Echmi} to the crystal structure of MsrA^{Ecoli}. (a, top) Total χ^2 for the four different couplings measured in phage alignment medium with respect to the primary sequence. Residues showing significant rapid backbone dynamics were excluded from the plot ($^1H-^{15}N$ heteronuclear nOe < 0.7). (b, bottom) Total χ^2 for the four different couplings measured in 5% C12E6/hexanol alignment medium with respect to the primary sequence. Residues showing significant rapid backbone dynamics were excluded from the plot ($^1H-^{15}N$ heteronuclear nOe < 0.7). Secondary structure elements are represented along the sequence at the top of each graphic.

mental RDC throughout the primary sequence, excepting the segment from Pro196 to Leu202, where strong disagreement is found (Figure 2). This local inconsistency is systematically found for each individual coupling type, in both alignment media. Removal of these data from the fit reproduces very similar alignment tensor values for the coil region alone (Table 1) as were found for the core region and verifies the relative orientation of these regions in the solution state.

The observed systematic disagreement could be explained by differential dynamics in this region, producing time-averaged RDC values that are in disagreement with a single conformational model, or by a different local conformation in MsrA^{Echmi} compared to the MsrA^{Ecoli} crystal structure. ^{15}N relaxation data measured at 600 MHz ^1H frequency present no evidence for large amplitude motion on the rapid (pico- to nanosecond) or intermediate (micro- to millisecond) dynamic time scale in this region. By use of the Lipari−Szabo approach, the lowest order parameter in this region is 0.75 and the average is 0.84 (Figure S1, Supporting Information). None of the residues in the strand requires an exchange contribution to R_2. It should be noted that this analysis does not exclude the possibility of differential dynamics occurring in this region on a time scale that is not easily detectable by ^{15}N relaxation measurements. RDC quenching due to broad sampling of orientational space appears to be unlikely in view of the presence of large measured RDCs throughout the strand relative to the available range from the known tensors (Table S2, Supporting Information). In the absence of direct evidence for mobility, we have therefore attempted to calculate the local conformation of this peptide segment using only RDC, to determine whether a single conformation can be found that satisfies the available data and expected covalent geometry.

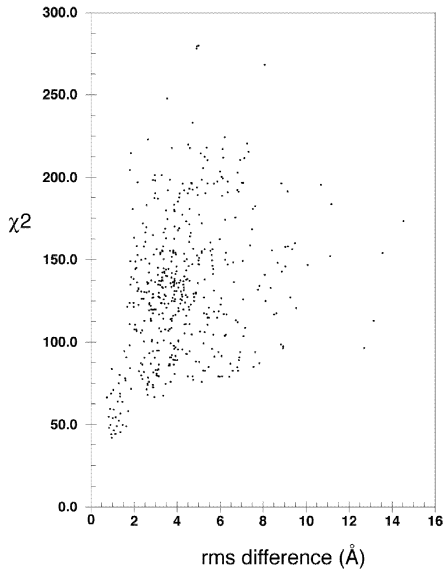

Figure 3. Conformational sampling of the structure calculation algorithm. The root-mean-square difference of the structures in the ensemble compared to the lowest χ^2 structure ($\chi^2 = 40$) is plotted with respect to each individual χ^2. Only conformers that do not violate any expected covalent angle by more than 10° are shown in the figure. Structures whose χ^2 falls below 62 form the ensemble shown in Figure 4.

De Novo Structure Determination with *meccano*. This analysis was performed via the *meccano* approach, previously demonstrated for the *de novo* determination of the backbone conformation of the protein ubiquitin.[14] In the case of MsrA^{Echmi}, the remainder of the protein structure is assumed to be identical to MsrA^{Ecoli}, as suggested from the RDC analysis in the core region. Only the structure of the P196−L202 peptide is determined *de novo*. The *meccano* method requires that the eigenvalues and relative orientation of the alignment tensors are known, or can be determined, and that sufficient RDC are available from two differently orienting media to sequentially build the backbone conformation in the molecular calculation frame. In the case of MsrA^{Echmi} the two tensors are determined from the analysis of the secondary structural core of the molecule, as described above, in comparison with the coordinates of MsrA^{Ecoli}. Despite the incompleteness of the RDC dataset over the whole sequence (only 90% of all possible RDC could be measured), nearly all potential RDC (47 from 48 potential couplings) were available from the peptide region, comprising six peptide planes and seven C^α junctions (Table S2, Supporting Information).

Conformational sampling of the algorithm is illustrated in Figure 3, where the root-mean-square difference of the structural ensemble compared to the lowest χ^2 structure is plotted with respect to each χ^2 (for simplicity only those structures with valid covalent angles are included in the figure). The 20 lowest target function (χ^2) conformers, selected on the basis of the total χ^2 only, are shown in Figure 4 (average pairwise rms difference in this alignment frame is (1.2 ± 0.3) Å over all atoms). The highest target function conformer in the final ensemble has $\chi^2 = 62$, and the best-fitting structure has a χ^2 of 40, close both to the expected experimental error and to the average from the core region ($\chi^2/N = 0.85$ compared to $\chi^2/N < 0.95$ for the helical regions of the crystal structure). While it is difficult to quantify the degree of confidence in the proposed model, the data appear to define a unique conformation as all structures with $\chi^2 < 70$ have the same fold as the minimum χ^2 conformer (all-atom rms difference of the aligned conformer is less than 1.9 Å). Nevertheless over the whole 2000-conformer ensemble, χ^2 ranges to a maximum of 640, and conformational space is sampled very broadly (pairwise backbone rmsd of aligned conformers 10.5 ± 4.2 Å), illustrating that algorithm efficiency is low.

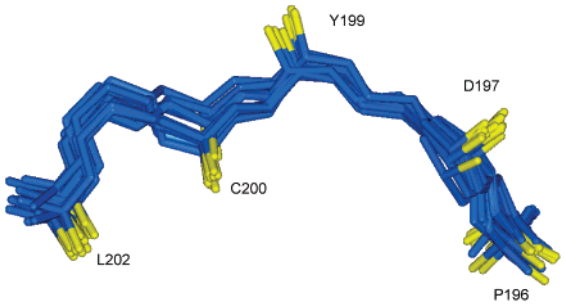

Figure 4. Ensemble of 20 lowest target function structures representing the backbone conformation of peptide P196−L202, determined by use of only RDC with the algorithm *meccano*. The conformers were placed relative to the common alignment frame, rather than superimposing their coordinates, because their relative orientation is known a priori. The average pairwise rmsd in this alignment frame is (1.2 ± 0.3) Å over the (C′, C^α, C^β, H^α, H^N, N) atoms. For clarity only the (C′, C^α, C^β, N) atoms are shown, and the C^β atoms are colored in yellow.

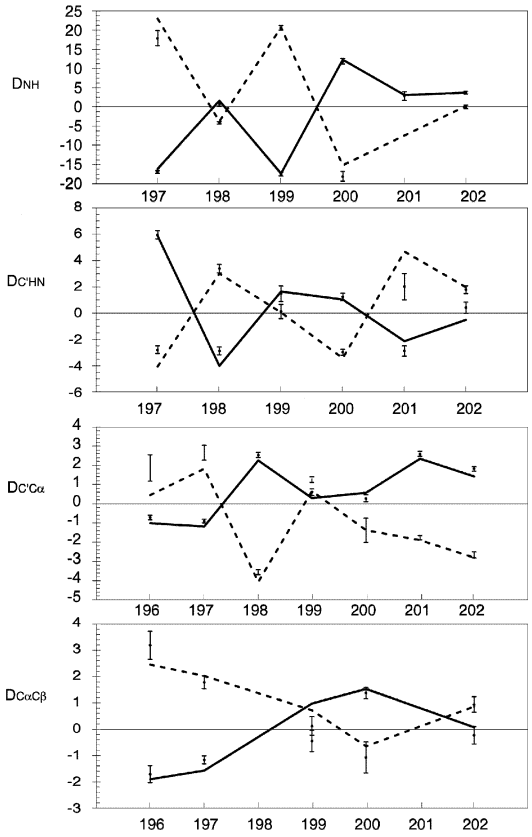

Figure 5. RDC values for each residue in the lowest energy *meccano*-derived conformation. Lines represent the values calculated from the model structure; the error bars are centered on the experimental values. Dashed lines correspond to the sample aligned in the alcohol medium, and solid lines, to the phage-aligned sample. These values are given in the Supporting Information.

The experimental and calculated values from the lowest target function model are shown in Figure 5. The covalent distortion present in the molecule due to possible overrestrained vector orientations appears to be minimal in the final ensemble as illustrated from the distances and dihedral and covalent angles

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

atoms	distance (Å)
$N-H^N$	1.02 ± 0.01
$C_\alpha-C_\beta$	1.53 ± 0.01
$C'-C_\alpha$	1.52 ± 0.02
$C_\alpha-N$	1.450 ± 0.003
$C'-N$	1.335 ± 0.002
atoms	**angles[a] (deg)**
$C'-C_\alpha-C_\beta$	109.2 ± 0.4
$N-C_\alpha-C_\beta$	109.3 ± 0.3
$C'-C_\alpha-N$	109.6 ± 0.1
$C'-N-C_\alpha$	121.2 ± 0.2
$C_\alpha-C'-N$	118.7 ± 0.2
$C'-N-H^N$	118.4 ± 0.2
$C_\alpha-N-H^N$	121.2 ± 0.2
$C_\alpha-C'-N-C_\alpha$	179.4 ± 0.3

[a] No deviations of angles greater than 2.6° from the force field values were found.

shown in Table 2. Total RDC violation (χ^2) and molecular strain (as measured by the residual energy in the covalent terms of the molecular force field) are correlated over all structures ($r = 0.89$, data not shown), illustrating that the experimental data are coherent with conformers of expected peptide chain geometry and in disagreement with conformers of incorrect geometry. All dihedral angles are in most-favored or additionally-allowed regions of the Ramachandran plot.

Insertion of the *meccano* Peptide into the MsrA^{Ecoli} Scaffold. Importantly, the length between the fragment termini P196-C^α and L202-C^α over the final ensemble (14.7 ± 0.4 Å) closely reproduces the equivalent length in MsrA^{Ecoli} (14.9 Å). To dock the *meccano* peptide, the equivalent peptide in MsrA^{Ecoli} was removed from the X-ray model. The procedure for positioning the fragment must then respect the stringent orientational degrees of freedom available for the two structural domains relative to the common alignment tensor frames. Under these conditions a single relative alignment of the two domains exists,[38] imposing an unambiguous orientation of the peptide fragment relative to the crystal structure. The translational docking must then be performed so that this relative alignment is respected, while best satisfying the expected covalence. This procedure is illustrated with respect to one of the alignment tensors by use of the program Module (Figure 6). We find that the fragment can be directly accommodated into the crystal conformation with no serious violation of known covalence. The final ensemble falls into two χ^2-equivalent subfamilies that differ at the Pro ψ backbone dihedral angle, and these subfamilies can be docked with covalent distances of $\{d(^{N195}C,^{P196}N) = 1.4 \pm 0.1$ Å and $d(^{L202}C,^{G203}N) = 1.35 \pm 0.07$ Å$\}$ and $\{d(^{N195}C,^{P196}N) = 2.0 \pm 0.1$ Å and $d(^{L202}C,^{G203}N) = 2.1 \pm 0.1$ Å$)\}$. The expected peptide bond distance between C^i and N^{i+1} is approximately 1.33 Å. The fact that the minimum χ^2 conformation of this peptide reproduces all measured RDC and is compatible with the MsrA^{Ecoli} structural scaffold provide a qualitative measure of our confidence in this approach to local structure determination.

Focused *de Novo* Structure Determination Compared to Restrained MD Refinement. Structure refinement of imprecise structural models with RDC has already been described, by use of restrained molecular dynamics in the presence of dipolar

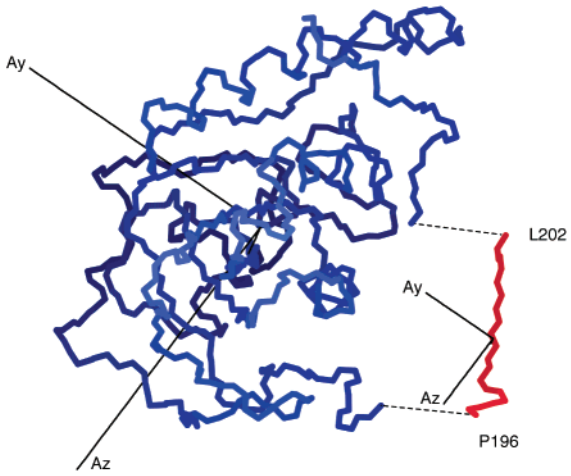

Figure 6. Representation of the positioning of the P196−L202 *meccano*-peptide relative to the crystal structure of MsrA^{Ecoli} by use of the program Module to place the fragment with respect to the alignment tensor in phage. Only transverse degrees of freedom are available in the common coordinate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

couplings measured in a single alignment medium.[12] It may be instructive to underline the differences between our analysis and this kind of study: In contrast to restrained structure refinement, which is subject to stringent constraints, for example on the terminal positions of the peptide of interest, our approach is de novo, so that the conformation of the peptide is determined only by RDC and local covalence. Once the local structure has been determined, the ability to replace the oriented peptide in the molecular scaffold then provides an independent measure of the probability that the conformation is realistic. This *de novo* approach is of course only made possible because the available data is sufficient to unambiguously define the molecular conformation (eight RDC per peptide unit).[14] This is not true if RDC from a single alignment medium are available, where multiple conformations can exist for the same measured data. In such cases molecular dynamics-based refinement of an available structure becomes the most appropriate method. The different conformational sampling characteristics of the two methods are evident.

Comparison with the Available Crystal Structures. The differences between the *meccano* conformation of MsrA^{Echmi} and the X-ray crystal MsrA^{Ecoli} structure in this region provide potentially important information concerning the proposed functional cycle of MsrA. While the Leu202 $C^\alpha-C^\beta$ vector direction is the same in both conformers, implying that the hydrophobic interactions for this side chain are conserved, the peptide chain around Cys200 has a very different conformation. Most importantly, the $C^\alpha-C^\beta$ vector is pointing away from the core in the X-ray structure and points into the core in the solution model (Figure 7, top). The Cys200 and Cys53 C^β atoms are therefore 3.2 Å closer together in the chimeric model ($d_{C\beta-C\beta} = 7.9$ Å) of MsrA^{Echmi}, and the disulfide-forming sulfurs are concomitantly closer, apparently obviating the need for previously suggested large-scale conformational rearrangement before the formation of this bond.[22] Note that no information concerning the exact position of the S^γ is available, as the χ^1 angle is undefined in our study.

It may be relevant that the crystal lattice contains three monomers, of which only one monomer (A) contains coordi-

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

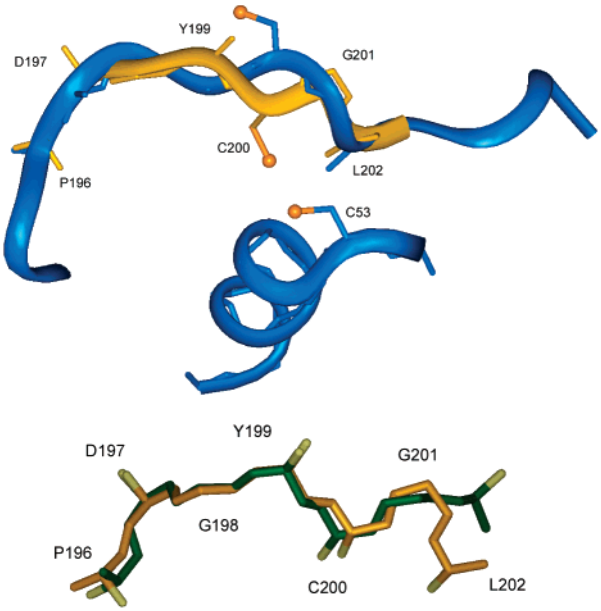

Figure 7. *De novo* RDC derived and crystal structures of the P196−L202 peptide. (a, top) Comparison of the native MsrA^{Ecoli} conformation (blue) and the lowest target function MsrA^{Echmi} model (yellow) after positioning of the *meccano* peptide onto the MsrA^{Ecoli} structure. The backbone dihedral angles are $\{(\phi,\psi) = (-129°, 1°)_{D197}, (-43°, 163°)_{G198}, (-72°, 170°)_{Y199}, (-150°, 98°)_{C200}, (-86°, 18°)_{G201}, (-34°,\text{undefined})_{L202}\}$ for the MsrA^{Echmi} model, compared to $\{\phi,\psi = (-78°, 166°)_{Y197}, (-71°, 133°)_{G198}, (-137°, 120°)_{Y199}, (-131°, 46°)_{C200}, (-116°, 17°)_{G201}, (-120°,/)_{I202}\}$ for the MsrA^{Ecoli} crystal structure. Note that the MsrA^{Echmi} numbering is used here for both molecules. The principal implication of the different backbone conformations involves the direction of the Cys200 $C^{\alpha}-C^{\beta}$ vector, and the consequent inter-S^{γ} distances between Cys53 (shown in the MsrA^{Ecoli} crystal configuration) and Cys200. The S^{γ} position is unknown from the RDC-defined structure and is shown here placed on the available cone with the assumption of a $\chi^1 = -60°$ conformation. (b, bottom) Comparison of the Pro196−Leu202 peptide backbone conformation in MsrA^{Echmi} determined from the direct structure calculation with RDC (yellow) and the equivalent conformation in MsrA^{bov} (green). For clarity only the (C', C^{α}, C^{β}, N) atoms are shown.

nates for the C-terminal region beyond Lys194. Superposition of the common main chain of monomer A on monomer B reveals that the Cys200 S^{γ} atoms from molecules A and B would be 5.5 Å apart if the two C-terminal conformations were identical. This suggests that the A-monomer structure may indeed be influenced by the close proximity of the Cys-containing strands, as is already the case for the intermonomer disulfide bridge observed between ^{A}Cys208 and ^{B}Cys88.[22] Interestingly, the more distantly related MsrA^{bov} (51% identity with several insertions) differs most significantly from the MsrA^{Ecoli} conformation in precisely this region, where the Cys218, equivalent to Cys200 in MsrA^{Echmi}, points toward the core of the molecule, as in our model of MsrA^{Echmi}. Indeed, the backbone of the RDC-defined peptide closely reproduces the MsrA^{bov} conformation between residues Pro196 and Cys200 (0.73 Å backbone rmsd over this strand; Figure 7, bottom). The inter-C^{β} distance between Cys200 and Cys53 ($d_{C\beta-C\beta} = 7.3$ Å and 8.0 Å) from the two crystalline conformations in MsrA^{bov} is also close to that predicted from the MsrA^{Echmi} model. Note that the Leu202 conformation differs significantly in MsrA^{bov}, resulting in a distinct L202 $C^{\alpha}-C^{\beta}$ bond orientation compared to MsrA^{Ecoli} and MsrA^{Echmi}.

The MsrA^{Ecoli} crystal contains high levels of disorder in the P196−L202 region with a maximal amplitude at Cys200. It has therefore been suggested that the C-terminal strand is inherently flexible and that this flexibility is related to the formation of disulfide intermediates involving Cys53, -200, and -208. Using NMR relaxation measurements we find no evidence of large-amplitude mobility on the ps-pico- to nanosecond time scale in this strand. Our data suggest that in MsrA^{Echmi} the conformation of this peptide is well-defined and sufficiently close to Cys53 to allow the Cys53−Cys200 disulfide bridge to occur without conformational rearrangement. This hypothesis is supported by the observation that, in contrast to MsrA^{Ecoli}, the equivalent peptide strand in MsrA^{bov} shows no differential disorder in the crystal compared to the rest of the peptide chain. Note that the *meccano* model is most similar to the MsrA^{bov} conformation, although it contains elements common to both crystal structures. Of course the second step of the proposed catalytic cycle, involving the Cys200−Cys208 bridge, still requires backbone reorganization as these Cys C^{β} atoms are on the order of 20 Å apart. This conformational change may be induced by destabilization of the packing of the C-terminal loop region onto the core following formation of the initial disulfide bridge.

Conclusions

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

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

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

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

JA0268783