JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Improved Accuracy from Joint X-ray and NMR Refinement of a Protein–RNA Complex Structure

Azzurra Carlon,[†] Enrico Ravera,[†] Janosch Hennig,[‡,§,∥] Giacomo Parigi,[†] Michael Sattler,*[,‡,§] and Claudio Luchinat*[,†]

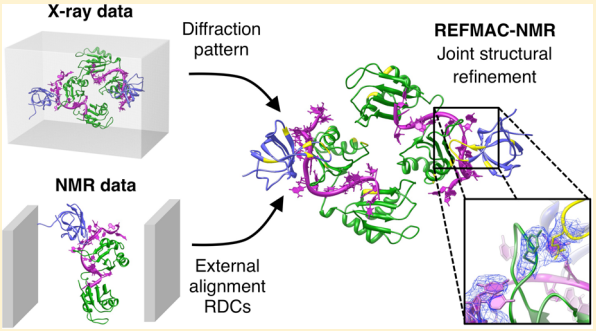
[†]Magnetic Resonance Center "CERM" and Department of Chemistry "Ugo Schiff", University of Florence and Magnetic Resonance Consortium (CIRMMP), Via L. Sacconi 6, 50019 Sesto Fiorentino, Firenze, Italy

[‡]Center for Integrated Protein Science Munich (CIPSM) at Department Chemie, Technische Universität München, 85747 Garching, Germany

[§]Institute of Structural Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

Supporting Information

ABSTRACT: Integrated experimental approaches play an increasingly important role in structural biology, taking advantage of the complementary information provided by different techniques. In particular, the combination of NMR data with X-ray diffraction patterns may provide accurate and precise information about local conformations not available from average-resolution X-ray structures alone. Here, we refined the structure of a ternary protein–protein–RNA complex comprising three domains, Sxl and Unr, bound to a single-stranded region derived in the *msl2* mRNA. The joint X-ray and NMR refinement reveals that—despite the poor quality of the fit found for the original structural model—the NMR data can be largely accommodated within the uncertainty in the atom positioning (structural noise) from the primary X-ray data and that the overall domain arrangements and binding interfaces are preserved on passing from the crystalline state to the solution. The refinement highlights local conformational differences, which provide additional information on specific features of the structure. For example, conformational dynamics and heterogeneity observed at the interface between the CSD1 and the Sxl protein components in the ternary complex are revealed by the combination of NMR and crystallographic data. The joint refinement protocol offers unique opportunities to detect structural differences arising from various experimental conditions and reveals static or dynamic differences in the conformation of the biomolecule between the solution and the crystals.



INTRODUCTION

X-ray crystallography and NMR spectroscopy are the most popular techniques able to retrieve information at the atomic resolution level. The structural knowledge provided by these two techniques is complementary, since X-ray diffraction patterns are mainly derived from heavy atom contributions, whereas NMR structural restraints mostly involve hydrogen nuclei. Moreover, the crystalline and solution states are two distinct physical environments, which may influence the structural arrangement of macromolecular systems. Indeed, a number of studies document the presence of differences between solution and X-ray structures, where the crystalline state reports on structural snapshots or minor conformations that in some cases are not expected to exist in solution.[1−14] It is not hard to imagine that the presence of a crystal lattice may add additional constraints (e.g., crystal packing forces), which can induce changes in the intra- and intermolecular conformations and/or in the dynamic features of the system.

Obtaining a comprehensive data set for complete structural characterization by NMR spectroscopy is usually difficult and very time consuming, especially in the case of high molecular weight systems where extensive isotope labeling schemes need to be applied to enhance the spectral quality.[15−17] On the other hand, residual dipolar coupling (RDC) data can be collected, with relative ease, also for large systems and can be used to detect potential inconsistencies between solution and crystal states.[8,10,18] In case no detectable inconsistencies are found, a joint structural refinement using both NMR and X-ray data provides a method to obtain a more reliable structural model, which may disclose additional relevant information on its functional mechanisms.

Uncertainty related to the experimental measurements is an important issue that needs to be carefully analyzed to assess the significance of the inconsistencies found when NMR data are used in conjunction with an X-ray structural model. Extensive, long-standing, and controversial discussions have built on such inconsistencies.[11,19−25] In this regard, it is important to realize

Received: November 5, 2015
Published: January 13, 2016

DOI: 10.1021/jacs.5b11598
J. Am. Chem. Soc. 2016, 138, 1601−1610

that besides the uncertainty related to NMR data measurements also the atomic coordinates in X-ray models may exhibit a non-negligible level of inaccuracy. Such inaccuracies, which may affect the positioning of different atomic moieties, mainly depend on the resolution of the X-ray reflections and on the structural refinement protocol employed.[11,26] Therefore, this so-called "structural noise" should be actively taken into account during the evaluation of inconsistencies, if any, between solution and crystal information.[26−28]

In the recent literature a number of approaches have been reported for refining X-ray structures with NMR data.[2,8,11−13,18,29−31] Most common refinement protocols consist in starting from an X-ray-derived structure and morphing the latter to achieve an acceptable agreement with the NMR data. This approach strongly relies on molecular libraries where the correct bonding geometry has to be kept (almost) completely rigid, because NMR measurements generally do not provide sufficient information to constrain the atom coordinates. Some of us have thus recently developed an approach (REFMAC-NMR) based on the simultaneous refinement of structural models against X-ray and NMR experimental data.[32] This allows for the joint use of the information about heavy atom positions, which often dominate X-ray reflections, together with the information about bond orientations for different nuclear pairs derived, in this particular case, from RDC data. Here, we show that REFMAC-NMR[32] refinement can be used to assess whether experimental NMR data can be explained by a structural model derived from X-ray crystallography within the accuracy of its diffraction pattern, and beyond the validation of the structure, RDCs can be used to provide a structure that complies with more than one data set and is thus more reliable. Moreover, we demonstrate that local conformational differences between the crystalline and solution states can be detected and exploited as useful hints on the functional mechanism of the system.

To provide a proof-of-principle of this workflow on a biologically relevant system, we assessed the recently reported crystal structure of the ternary Sxl−Unr-*msl2*-mRNA complex,[33] which consists of the two RNA recognition motifs (RRMs) of Sxl, the first of five cold shock domains of Unr (CSD1), and an 18-mer single-stranded RNA derived from *msl2*-mRNA. Assembly of this complex is vital for female viability in fruit flies, as translational repression of *msl2*-mRNA by Sex-lethal (Sxl) and Upstream-of-N-Ras (Unr) prevents the formation of the dosage compensation complex resulting in normal transcription of X-linked genes. The structure has unique protein−RNA and protein−protein interfaces that demonstrate how specificity and affinity for the cognate RNA is achieved by cooperative action of two distinct RNA binding proteins. The structure of this ternary complex constitutes an ideal test case for our purpose, as complementary NMR data are also available.[33,34] In the present manuscript we focus on the use of a set of residual dipolar couplings obtained for Pf1 phages alignment medium together with the available X-ray data at 2.8 Å resolution (PDB code 4QQB).

MATERIALS AND METHODS

Structural Refinement. Structural refinements were performed by the simultaneous use of the X-ray diffraction pattern and RDC data employing the recently developed program REFMAC-NMR.[32] The general approach consists of (i) a first minimization against the X-ray data alone, with an automatic setting of the *weight matrix* value (i.e., of the relative weights of geometry violations and X-ray violations), possibly followed by manual adjusting of the weight matrix to reduce the calculated rmsd of bond lengths, bond angles, and chiral volumes, if too large, and (ii) a second minimization performed including RDC restraints in order to decrease their *Q* factor. RDCs were included in the calculation only for residues which are not dynamic, most of which are part of secondary structure elements. In particular, the NMR restraints contribution (t) to the total minimized function is

$$t = k_{\mathrm{RDC}} \sum_i w_i[\max(|RDC_i^{\mathrm{calc}} - \mathrm{RDC}_i^{\mathrm{obs}}| - T_i, 0)]^2 \quad (1)$$

where T_i is the tolerance on each RDC value, w_i is its weight, and k_{RDC} is the overall weighting factor for RDCs. In tables the products of the k_{RDC} and w_i values will be indicated as RDC weight. The second minimization, besides optimization of the weight matrix value, requires optimization of the weights of the NMR data and of additional torsion angle restraints. Three further torsion angles were in fact introduced in the REFMAC library to restrain the planarity of the $O_i-C_i-N_{i+1}-C^{\alpha}_{i+1}$, the $C_{i-1}-N_i-C^{\alpha}_i-H_i$ (out of plane bending of H^N-N bonds), and the $C^{\alpha}_i-C_i-N_{i+1}-C^{\alpha}_{i+1}$ dihedral angles (*pep1*, *pep2* and ω, respectively; force constants and tolerances used in the calculations are reported in Table S1). This was needed to avoid worsening of the deviations of geometric parameters from ideality by the inclusion of the NMR data in the refinement. Furthermore, overall weighting parameters over ideal geometries of all atoms involved (*weight refined_atoms*) or not involved (*weight other_atoms*) in the calculation of gradients and of the second derivatives corresponding to X-ray reflections were also introduced. Of note, bond distances of hydrogens in X-ray libraries are different from those in NMR libraries, because the hydrogen electron is not centered on the position of the nucleus but closer to the atom to which it is attached. Therefore, the coordinates of the hydrogens used for back-calculating the NMR restraints were recalculated by increasing the distances between the hydrogens and their binding nuclei to the values used in the AMBER[35,36] library (H^N-N distance of 1.020 Å). This correction for the evaluation of the NMR restraints does not affect the geometric restraints in the usual X-ray refinement, which considers hydrogen positions according to the standard crystallographic library.

Alignment Tensor Calculation. The alignment tensors and the agreement between experimental and back-calculated RDCs were computed using the FANTEN web application,[37] available in the WeNMR portal.[38] From the fit of the experimental RDCs to eq 2

$$\mathrm{RDC} = -\frac{3\mu_0 S_{\mathrm{LS}}}{4\pi^2}\frac{\gamma_A\gamma_B\hbar}{r_{\mathrm{AB}}^3}\left[D_{zz}\frac{2z_{\mathrm{AB}}^2 - x_{\mathrm{AB}}^2 - y_{\mathrm{AB}}^2}{2r_{\mathrm{AB}}^2} + (D_{xx} - D_{yy})\frac{x_{\mathrm{AB}}^2 - y_{\mathrm{AB}}^2}{2r_{\mathrm{AB}}^2} + D_{xy}\frac{2x_{\mathrm{AB}}y_{\mathrm{AB}}}{r_{\mathrm{AB}}^2} + D_{xz}\frac{2x_{\mathrm{AB}}z_{\mathrm{AB}}}{r_{\mathrm{AB}}^2} + D_{yz}\frac{2y_{\mathrm{AB}}z_{\mathrm{AB}}}{r_{\mathrm{AB}}^2}\right] \quad (2)$$

the program provides the five independent elements of the alignment tensor ($D_{xx} - D_{yy}$, D_{zz}, D_{xy}, D_{xz}, D_{yz}), from which the axial component of the tensor and its rhombicity, *A* and *R*, respectively, and the Euler angles defining the principal directions of the tensor can be derived (x_{AB}, y_{AB}, and z_{AB} are the components of the distance between the two coupled nuclei A and B). The similarity between tensors calculated from the best fit against different structures was assessed according to the following indicators: (1) the ratio of the axial components of the tensors; (2) the ratio of the tensor sizes, taking into account also their rhombicity; (3) the normalized dot product between the five independent elements of the alignment tensor. The first two criteria report on the similarity of the size of the tensors; the third criterion encodes information on their shape and orientation. In all cases, values close to 1 indicate good similarity between tensors.[19,39−41]

The experimental RDC data were taken from Hennig et al.;[33] X-ray data were taken from PDB accession code 4QQB.

^{15}N NMR Relaxation Measurements. ^{15}N R_1 and R_2 relaxation rates for Sxl at two different concentrations (0.25 and 1 mM) were measured at 800 MHz proton Larmor frequencies at 298 K.[42] $\{^1H\}-^{15}N$ heteronuclear NOE data for CSD1 were acquired at 0.3

DOI: 10.1021/jacs.5b11598
J. Am. Chem. Soc. 2016, 138, 1601−1610

mM protein concentration. For Sxl at 0.25 mM, R_1 relaxation rates were derived from measuring 14 different relaxation delays, including 3 duplicates for error estimation (21.6 (2×), 43.2, 86.4, 162, 345.6, 518.4 (2×), 669.6, 885.6 (2×), 1080, 1274.4, 1555.2, 1944, and 2376 ms). For $R_{1\rho}$ 12 different relaxation delays with two duplicates were measured (5 (2×), 10, 15, 25, 40, 55, 70, 80 (2×), 90, 110, and 140 ms). At 1 mM concentration of Sxl, R_1 was derived from 10 different relaxation delays recorded with two duplicates (21.6 (2×), 86.4, 162, 345.6, 518.4 (2×), 669.6, 885.6, 1274.4, 1728, and 2376 ms). R_2 was derived from measuring eight relaxation delays with two duplicates (10.88 (2×), 21.76, 32.64, 43.52 (2×), 54.4, 65.28, 76.16, and 87.04 ms). Data were fitted and analyzed using the software PINT.[43]

RESULTS

REFMAC-NMR Refinement. The crystal structure of the Sxl–Unr-*msl2*-mRNA complex was used to evaluate (a) the conformation and (b) the relative arrangement of the two Sxl RRM domains and CSD1 domain in solution against H^N–N and C–N RDC data.[33] From the best fit against a single alignment tensor, the agreement of the RDCs with the available X-ray structure (PDB code 4QQB) provides a *Q* factor of 0.440 (Tables 1 and S2). Despite the quality of the fit appearing to be rather modest, the result is in line with what can be expected from the X-ray data resolution for this system (2.8 Å).[26] The quality of the fit for the two complexes found in the asymmetric unit (chains A–P–X and B–C–Y) is slightly different, although the derived alignment tensor parameters are rather similar (Table 1).

Table 1. REFMAC-NMR Refinement Calculations Performed as for the Original Structure and Without (−NMR) and With (+NMR) the Inclusion of NMR Restraints[a]

PDB code 4QQB; resolution 2.80 Å					
		three tensors Sxl (Nter), Sxl (Cter), and CSD1	two tensors Sxl and CSD1	single-tensor Sxl–CSD1 complex	
parameters	original structure	−NMR	+NMR	+NMR	+NMR
R	0.198	0.198	0.201	0.201	0.201
R^{free}	0.236	0.234	0.236	0.236	0.235
RMSD bond length	0.006	0.006	0.009	0.009	0.009
RMSD bond angles	1.113	1.260	1.592	1.591	1.595
RMSD chiral volume	0.074	0.099	0.097	0.097	0.097
Q factor RDC	0.440		0.124	0.131	0.144

[a]Simultaneous refinement of X-ray and RDC data was performed using independent tensors for two domains of Sxl (RRM1, RRM2) and for CSD1 (three tensors) using independent tensor for full-length Sxl and CSD1 (two tensors) and using a unique tensor for the overall Sxl–CSD1 complex (single tensor).

To evaluate the presence of any intradomain conformational differences, RDC data were used to refine the conformations of the individual structural units of the protein components of the ternary complex (both RRM domains of Sxl and the CSD1 domain). REFMAC-NMR was employed for performing the structural refinement using the protocol previously described.[32] The peculiarity of this approach consists of taking into account the experimental uncertainty and coordinate precision of the X-ray data when RDC data are included as structural restraints (see Table 1). The joint refinement against both X-ray and RDC data allows for small but relevant changes of the atomic coordinates in order to satisfy the RDC data still being in agreement with the X-ray data. The joint refinement leads to an overall drop of the *Q* factor from 0.440 to 0.124, without any significant increase in the *R* or R^{free} values or in violations of geometrical constraints (rmsd for bond lengths, bond angles, and chiral volumes; see Table 1). These results suggest that the poor agreement of the RDC data initially observed for the original X-ray structure was mainly due to the presence of inaccuracy in atom positions and that no significant (i.e., outside the experimental error) structural differences in the conformation of the single domains constituting the complex exist between the crystalline and the solution states. Moreover, such improvement in the agreement of RDCs reveals that the additional information provided by these restraints assists in better defining the structure upon joint refinement, whereas the fact that the crystallographic R^{free} does not significantly increase with respect to *R* indicates the absence of over-refinement. Actually, the electron density map is even slightly changed after inclusion of NMR restraints (see Figure 4).

In order to test the presence of interdomain rearrangements, the tensors calculated for the individual units were compared with one another in terms of magnitude, alignment, and shape. These parameters turned out to be very similar for the two domains of Sxl (Table 2), pointing out that both domains could be refined by using the same tensor without any significant worsening in the agreement with RDC data. As expected, the refinement results obtained by imposing a single tensor for the two domains are satisfactory (with only a small increase in the *Q* factor from 0.124 to 0.131, see Table 1). This indicates that the Sxl domains in solution maintain the same relative rearrangement as observed in the crystal and that the presence of significant interdomain motion can be reasonably excluded. On the contrary, a notable difference in the magnitude of the alignment tensor was observed for CSD1. However, the shape and orientation of the tensor is almost indistinguishable from the tensor determined for Sxl (Table 2). The most likely explanation for this is a difference in the experimental conditions, e.g., a slight variation in alignment medium concentrations in the different samples used for the RDC

Table 2. Comparison between the Alignment Tensors Calculated Independently for the Two Domains of Sxl (RRM1, RRM2), and for Sxl and CSD1

comparison between tensors calculated for Sxl domains			
	magnitude of the axial component $A_{Sxl(RRM1)}/A_{Sxl(RRM2)}$	magnitude of the axial and rhombic components $(D_{zz}-D_{xx})_{Sxl(RRM1)}/(D_{zz}-D_{xx})_{Sxl(RRM2)}$	orientation and shape $(D_{Sxl(RRM1)}\cdot D_{Sxl(RRM2)})/(\|D_{Sxl(RRM1)}\|\|D_{Sxl(RRM2)}\|)$
chain A	0.98	1.02	0.95
chain B	1.02	1.07	0.98
comparison between tensors calculated for Sxl and CSD1			
	magnitude of the axial component A_{Sxl}/A_{CSD1}	magnitude of the axial and rhombic components $(D_{zz}-D_{xx})_{Sxl}/(D_{zz}-D_{xx})_{CSD1}$	orientation and shape $(D_{Sxl}\cdot D_{CSD1})/(\|D_{Sxl}\|\|D_{CSD1}\|)$
chains A, X	0.80	0.81	0.98
chains B, Y	0.79	0.80	0.99

DOI: 10.1021/jacs.5b11598
J. Am. Chem. Soc. 2016, 138, 1601–1610

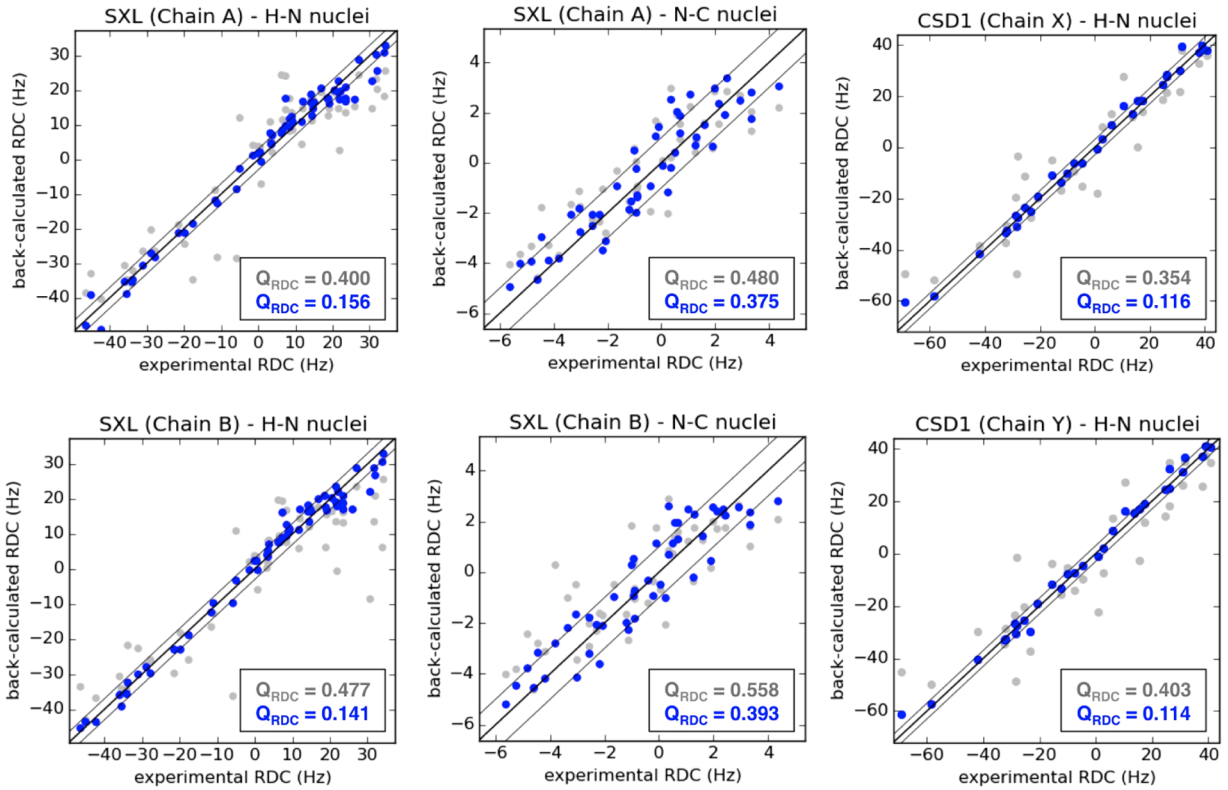


Figure 1. Correlation plots between experimental and back-calculated RDCs for the original structure (gray dots) and for the refined structure (blue dots).

measurements of the Sxl and CSD1 data involving Sxl- and CSD1-isotope-labeled complexes, respectively. This explanation is supported by the following considerations: (a) the tensor is reduced but has the same orientation and anisotropy, which is a hint of just a smaller alignment, and (b) it is unlikely that Sxl, which is larger, with a more anisotropic shape and intimately connected with the highly charged RNA molecule, could show smaller alignment than the smaller and more isotropic CSD1. Hence, a uniform scaling by an empirical factor 0.8 was applied to CSD1 RDC values in order to perform a new REFMAC-NMR refinement calculation using a single alignment tensor for the complete RDC data set. The refinement showed that the RDC Q factor increases only marginally (from 0.124 to 0.144, see Table 1) on passing from the use of three independent tensors for the three individual units of the Sxl–Unr complex to the use of a single tensor, remaining much smaller with respect to the Q factor of 0.440 calculated for the original X-ray model. No appreciable differences are observed for the structural statistics of the X-ray data (see Table 1).

In summary, these results indicate that the refined crystal structures provide a very good fit of the NMR data and thus represent also a good model of the Sxl–Unr complex in solution. The correlation plots reporting the agreement of the experimental RDCs with the refined structural model are shown in Figure 1. A good overall agreement is observed for both Sxl–Unr complexes (chains A–P–X and B–C–Y) found in the asymmetric unit of the crystal, with a slight preference for chains B and Y with respect to chains A and X. Taking into account the measurement errors (3 and 1 Hz for H^N–N and C–N, respectively), H^N–N RDCs collected for CSD1 and C–N RDCs for Sxl reveal optimal fit of the available data, with $\chi^2_{reduced}$ values of 1.003 (chain X), 0.97 (chain Y), 1.04 (chain A), and 1.12 (chain B); Sxl H^N–N RDCs show $\chi^2_{reduced}$ of 1.458 for chain A and 1.192 for chain B. Whereas C–N RDCs show agreement between calculated and observed data almost within the experimental error, it is immediately apparent that larger deviations exist in H^N–N RDCs. These deviations, which are discussed in detail in the Possible Differences between Crystal and Solution Structures section, show a peculiar effect of the joint refinement (see Table 3 and Figure S1). The refinement produces a sizable overall reduction of the differences between observed and calculated values, but still some of the values largely exceed the experimental uncertainty. Most notably, those residues that stand out after refinement were not violating more than others in the original structure,

Table 3. Residues Showing after Refinement a Deviation between Observed and Calculated Data Larger than the Standard Deviation Compared with Their Deviation before the Refinement (see Figure S1)

residue	original	refined
137	A, 0.300; B, 0.246	A, 1.353; B, 1.412
138	A, 0.261; B, 0.068	A, 0.989; B, 0.894
142	A, 1.364; B, 1.307	A, 1.920; B, 1.584
146	A, 2.234; B, 2.598	A, 1.260; B, 1.110
188	A, 0.404; B, 0.313	A, 1.317; B, 1.179
190	A, 0.506; B, 0.124	A, 0.966; B, 0.426
191	A, 1.859; B, 4.463	A, 2.384; B, 2.582
195	A, 0.208; B, 0.492	A, 0.875; B, 0.888
212	A, 0.812; B, 0.963	A, 2.925; B, 2.536
238	A, 1.138; B, 0.763	A, 1.967; B, 1.890
240	A, 0.762; B, 1.199	A, 1.743; B, 1.394
257	A, 1.264; B, 1.043	A, 2.560; B, 2.617

DOI: 10.1021/jacs.5b11598
J. Am. Chem. Soc. 2016, 138, 1601–1610

whereas they do after the violations of the others that have been successfully reduced by refinement.

Comparison of the refined structure with respect to the original model (Table 4) showed slightly improved fits of crystallographic data and Ramachandran scores with an increase in the percentage of residues belonging to the core (from 86.0% to 87.2%) and a decrease of those in the allowed (from 13.3% to 12.3%) and in the generously allowed regions (from 0.5% to 0.2%).

Table 4. Quality Evaluation for the Original Structure and for the REFMAC-NMR Refined Structure As Calculated by the wwPDB Validation Server (wwpdb-validation.wwpdb.org)

	original	refined
R^{free}	0.236	0.235
RSRZ outliers	5.4%	5.0%
Clashscore	4	5
Ramachandran outliers	core: 86.0%	core: 87.2%
	allowed: 13.3%	allowed: 12.3%
	generous: 0.5%	generous: 0.2%
	disallowed: 0.2%	disallowed: 0.2%
RNA backbone	0.35	0.34

Notably, REFMAC-NMR refinement produced an effective improvement of the structural model with NMR data. As a proof that the major contribution to the improvement is not due to simple in-plane or out-of-plane distortions of the H^N−N bonds (even if within the standard limits, see Figure S2a,b) the protons were removed from the refined structure and added back by using automatic methods available from common software (i.e., Molprobity[44] and MOLMOL 2K). Evaluation of the "reprotonated" structure shows that the agreement with NMR data is clearly maintained (Table 5), whereas adding protons with the same programs to the original structure does not provide any improvement. This demonstrates how much the uncertainty in heavy atom coordinates can affect the automatic positioning of hydrogen atoms and thus the orientation of H^N−N bonds to which RDCs mostly refer to.

Novel Insights from the Joint Refinement. REFMAC-NMR does not produce any global difference in the refined structure with respect to the original one (backbone coordinate rmsd of 0.066, 0.075, 0.093, and 0.094 Å for chains A, B, X, and Y, respectively). Indeed, complexes A−P−X and B−C−Y, which slightly differ in their relative positions (backbone coordinate rmsd of the secondary structure is 0.470 and 0.444 Å, respectively, between chains A and B and X and Y), get only slightly closer after the refinement (backbone coordinate rmsd of the secondary structure is 0.463 and 0.400 Å, respectively, between chains A and B and X and Y) (Table 6). However, by looking at the structural changes in more detail, it can be observed that small but significant local differences between the two complexes in the asymmetric unit exist (e.g., around residue 60 in CSD1, see Figure S3). Local differences are even more apparent in the backbone dihedral angles of the two complexes (Figure S4), which become more similar (average decrease of the rmsd in the secondary structure elements around 12% for both φ and ψ angles, see Table 6).

Table 6. RMSD between the Two Complexes in the Asymmetric Unit before and after Refinement

rmsd	original	refined
backbone atoms[a]	A−B, 0.470; X−Y, 0.444	A−B, 0.463; X−Y, 0.400
backbone angles[b]	φ A−B, 15.686; X−Y, 12.324	φ A−B, 13.015; X−Y, 8.156
	ψ A−B, 13.855; X−Y, 7.681	ψ A−B, 12.123; X−Y, 7.199
backbone atoms[c]	A−B, 0.508; X−Y, 0.488	A−B, 0.502; X−Y, 0.446

[a]rmsd calculated upon independent alignment of chains A and B and of chains X and Y. [b]rmsd calculated for the secondary structure elements. [c]rmsd calculated upon alignment of complexes A−P−X and B−C−Y.

Indeed, inclusion of RDC data mostly aids the local refinement of the two structures. Figures 2 and 3 report the backbone rmsd per residue between original and refined structures. Some differences of modest extent (never exceeding 0.2 Å) are present for Sxl (chains A and B). The most relevant changes involve residues N152, I230, and A271 for chain A and residues R146, N152, V185, T190, V191, Q239, and K240 for chain B, scattered in different regions of the two RRM domains. Slight differences are also detected in the recognition of secondary structure elements by the DSSP software,[45] the most significant involving a better definition of a new β-strand constituted by residues S285−L288 and of a helix-3 for residues E206−I208 (Figure 6). The improved definition of these structural elements is likely a consequence of additional information provided by the joint refinement.

In contrast, for the CSD1 domain (chains X and Y) structural variations are found to be all grouped in the well-defined loop region constituted by residues Y56−P63 (Figure 2), for which a conformational difference is observed upon joint refinement (Figures 3 and 6). Refinement of the loop conformation is reflected also in the slight rearrangement of some of the side chains, the most relevant one being R59 (Figure 4a and 4b). This residue is of particular importance as it forms an essential contact with Sxl Y164. Substitution of either residue by alanine severely impairs or abolishes the formation of the entire

Table 5. Comparison of the Q_{RDC} for the Sxl−CSD1 Complex for the Refined Structure when H^N Atoms Are Added by REFMAC-NMR by MolProbity and by MOLMOL 2K as Well as for the Original Structure

	refined			original
	H^N added by REFMAC-NMR Q_{RDC} (H^N−N, C−N)	H^N added by MolProbity Q_{RDC} (H^N−N, C−N)	H^N added by MOLMOL Q_{RDC} (H^N−N, C−N)	H^N added by MolProbity Q_{RDC} (H^N−N, C−N)
SXL (chain A) RRM1	0.144 (0.140, 0.390)	0.161 (0.157, 0.390)	0.197 (0.194, 0.389)	0.425 (0.425, 0.470)
SXL (chain A) RRM2	0.180 (0.176, 0.360)	0.195 (0.192, 0.360)	0.178 (0.175, 0.354)	0.366 (0.364, 0.491)
CSD1 (chain X)	0.116	0.135	0.189	0.354
SXL (chain B) RRM1	0.134 (0.129, 0.398)	0.155 (0.151, 0.398)	0.192 (0.189, 0.396)	0.560 (0.561, 0.478)
SXL (chain B) RRM2	0.161 (0.156, 0.379)	0.170 (0.165, 0.379)	0.182 (0.178, 0.376)	0.336 (0.331, 0.633)
CSD1 (chain Y)	0.114	0.136	0.200	0.403

DOI: 10.1021/jacs.5b11598
J. Am. Chem. Soc. 2016, 138, 1601−1610

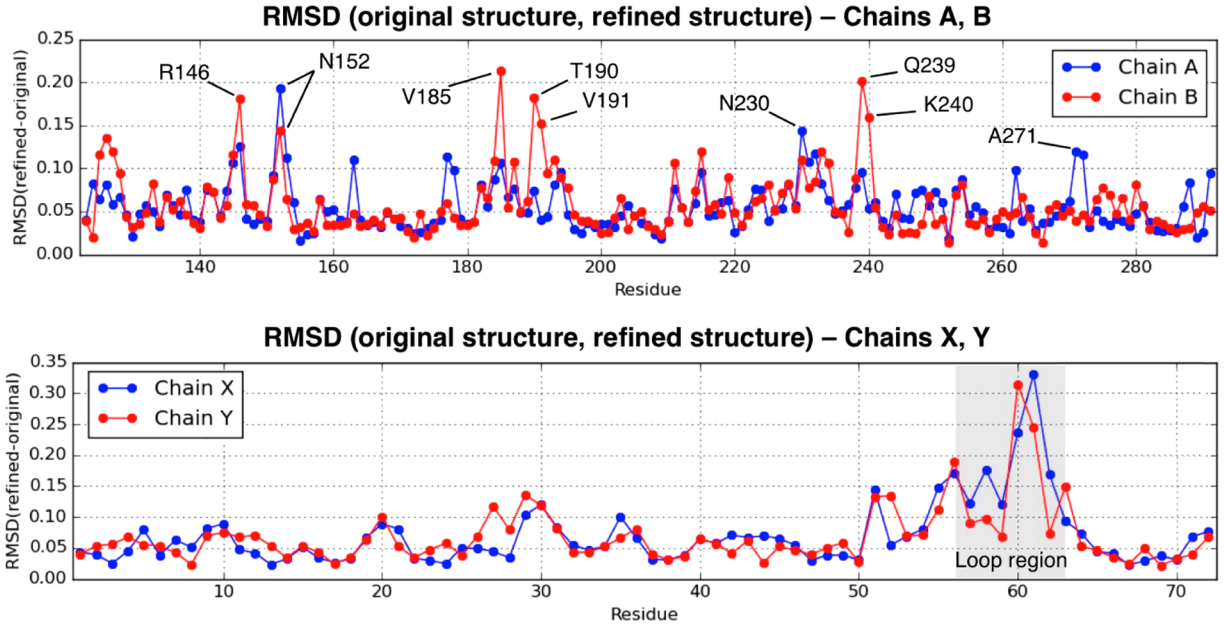


Figure 2. rmsd calculated from backbone C, C_α and N nuclear positions between the original and the refined structure for chains A, B (top), and X, Y (bottom).

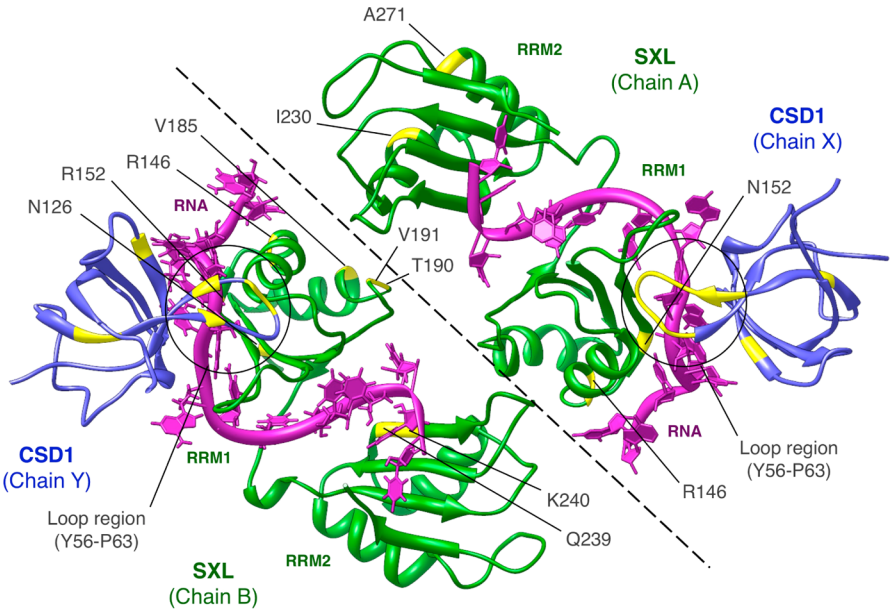


Figure 3. Two conformations of Sxl–Unr complex structure (chains A–P–X and B–C–Y) as found in the asymmetric crystal unit (pdb file 4QQB). Residues showing a rmsd in the backbone C, C_α and N nuclear positions between the original and the refined structures above 0.12 Å are shown in yellow.

complex.[33] It is also interesting to observe that in the free CSD1 domain residues R58 and R59 exhibit slightly higher conformational flexibility on subnanosecond time scales as indicated by the low heteronuclear $\{^1H\}-^{15}N$ NOE data (Figure S5) when compared to other residues in the free state. This likely correlates with the observed conformational differences detected by REFMAC-NMR, suggesting that the refinement is more effective where electron density is less determined. Further indication of conformational rearrangement of CSD1 loop is provided by the fact that Y164 of Sxl, which interacts directly with R59 of CSD1, adopts two distinct conformations in the crystal by chains A–P–X and B–C–Y (Figure 4c and 4d). Indeed, while in the A–P–X complex the side chain of Y164 is stacked against the side chain of CSD1–R59, in the B–C–Y complex it is flipped to the other side and interacts with the RNA base of U7. The former is not observed in the absence of CSD1 but the latter while bound to *transformer* mRNA.[46] This indicates that both conformations of Y164 are energetically accessible from the solution conformation. Bringing together the information about the two partners, this suggests that the high flexibility of the CSD1 loop, coupled to the conformational heterogeneity detected through the side chain of Y164, may play an important role for complex formation. Of note, in this case REFMAC-NMR refinement gave access to useful information about the CSD1 loop, otherwise impossible to retrieve by NMR relaxation measure-

DOI: 10.1021/jacs.5b11598
J. Am. Chem. Soc. 2016, 138, 1601–1610

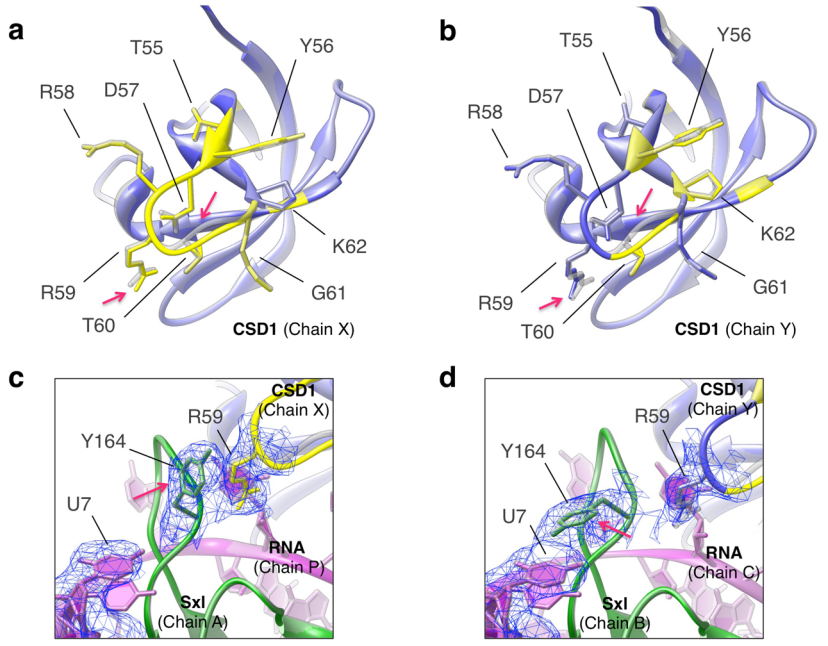


Figure 4. Detail of the CSD1 loop at the interface with RNA (a) for chains A–P–X and (b) B–C–Y. The original X-ray structure is shown in gray, whereas the refined structure is colored as in Figure 3. Relevant variations are indicated by red arrows. Zoomed view indicating conformational differences for CSD1 R59 and Sxl Y164 for chains (c) A–P–X and (d) B–C–Y. The flipping side chain of Y164 in the two conformations is indicated by a red arrow. Electron density for R59, Y164, and U7 and for neighboring residues/bases is shown in blue mesh lines at contour level of sigma 1.

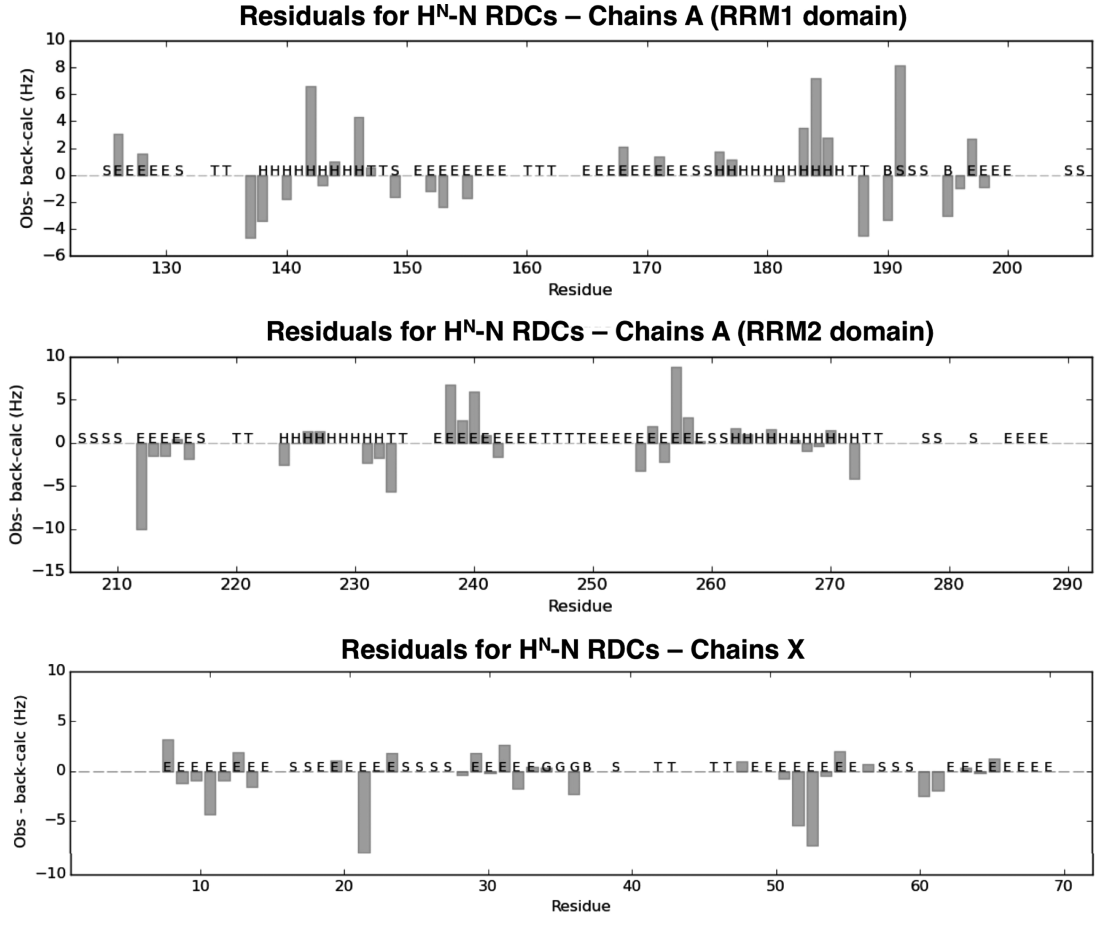


Figure 5. Residuals computed as difference between experimental and back-calculated RDCs for H^N–N nuclei of chains A and X.

DOI: 10.1021/jacs.5b11598
J. Am. Chem. Soc. 2016, 138, 1601–1610

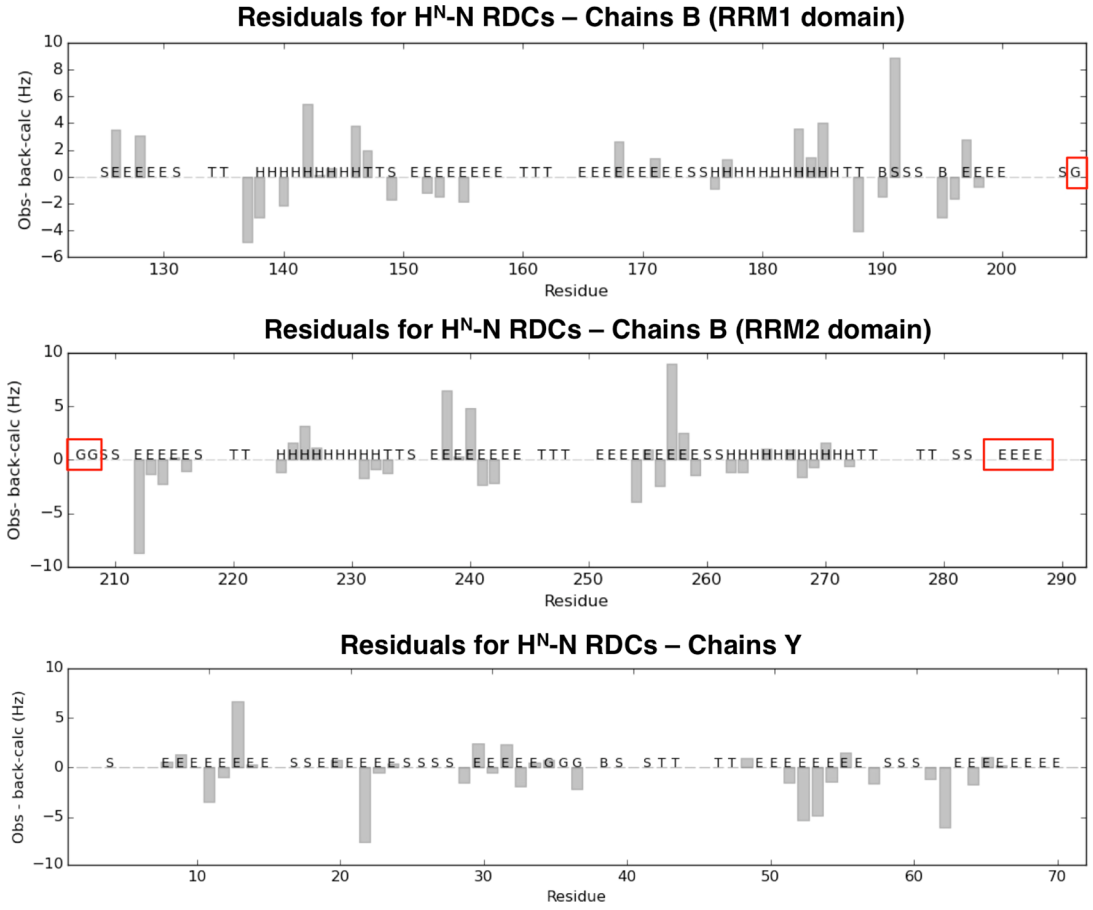


Figure 6. Residuals computed as difference between experimental and back-calculated RDCs for H^N–N nuclei of chains B and Y. Highlighted residues S285–L288 and E206–I208 are detected by DSSP software to change fold after REFMAC-NMR refinement.

ments of the complex due to line broadening of the amide resonance.

Possible Differences between Crystal and Solution Structures. Despite the refined structure presenting an overall very good fit for the available NMR data, a number of violations can still be observed, especially for some H^N–N RDC values belonging to Sxl (see section REFMAC-NMR Refinement). Figures 5–7 report the differences between experimental and back-calculated data, referred to as residuals, for the H^N–N RDC of both Sxl and CSD1. Interestingly, differences are found to be mainly clustered in two groups comprising residues T137, D138, Y142, and R146 (Figure 7b) and N212, V238, K240, and V257 (Figure 7c), which are located in the RRM1 and RRM2 domains of Sxl, respectively, and at the RNA binding interface. In particular, residues Y142 and R146 in RRM1 are of special significance, as chemical shift perturbations and mutational analyses confirmed their key role during RNA binding and complex formation[33] in vitro and during functional activity in vivo. This illustrates how the joint refinement provides novel structural insight for regions that are not well defined by the individual methods.

Minor but significant discrepancies are also observed for residues G188, T190, V191, and R195, located in the loop of Sxl (Figure 7d). This loop represents the region in which the two complexes A–P–X and B–C–Y are closer in space: the interaction of R192 in chain A with the loop K246–R250 of chain B results in the formation of two H bonds (R192–K246, R192–L247). Therefore, the presence of these interactions in the crystalline state may be at the origin of the structural differences observed for this region with respect to the solution structure. Of note, although at low concentrations this complex is monomeric, a concentration-dependent weak dimerization is observed for Sxl in solution, as indicated by the increased rotational correlation time (Figure S6). Consistently, residues around I189, including G188, T190, and V191 (relaxation rates of R195 could not be assessed), exhibit longer apparent rotational correlation times derived from ^{15}N R_2/R_1 relaxation rates at higher concentrations. During crystallization of the complex the concentration is even higher, and the same residues seem to be involved in the interaction between both complexes in the asymmetric unit. This interaction does not take place at lower concentrations at which RDCs were acquired and may thus explain the discrepancy observed. On the other hand, residues around K246 are more flexible at both concentrations. In the ternary complex studied here differences between crystal and solution data are observed for residues that interact with the 3′ region of the RNA. As in the cellular context the full-length RNA sequence extends beyond A20; additional factors might stabilize the RNA contacts.

CONCLUSIONS

Here, we show that joint X-ray and NMR refinement can be effectively used to probe if a crystal structure reflects the conformation found in solution at a residue-by-residue level (depending on the NMR data used). Moreover, the joint refinement can reveal local differences between solution and

DOI: 10.1021/jacs.5b11598
J. Am. Chem. Soc. 2016, 138, 1601–1610

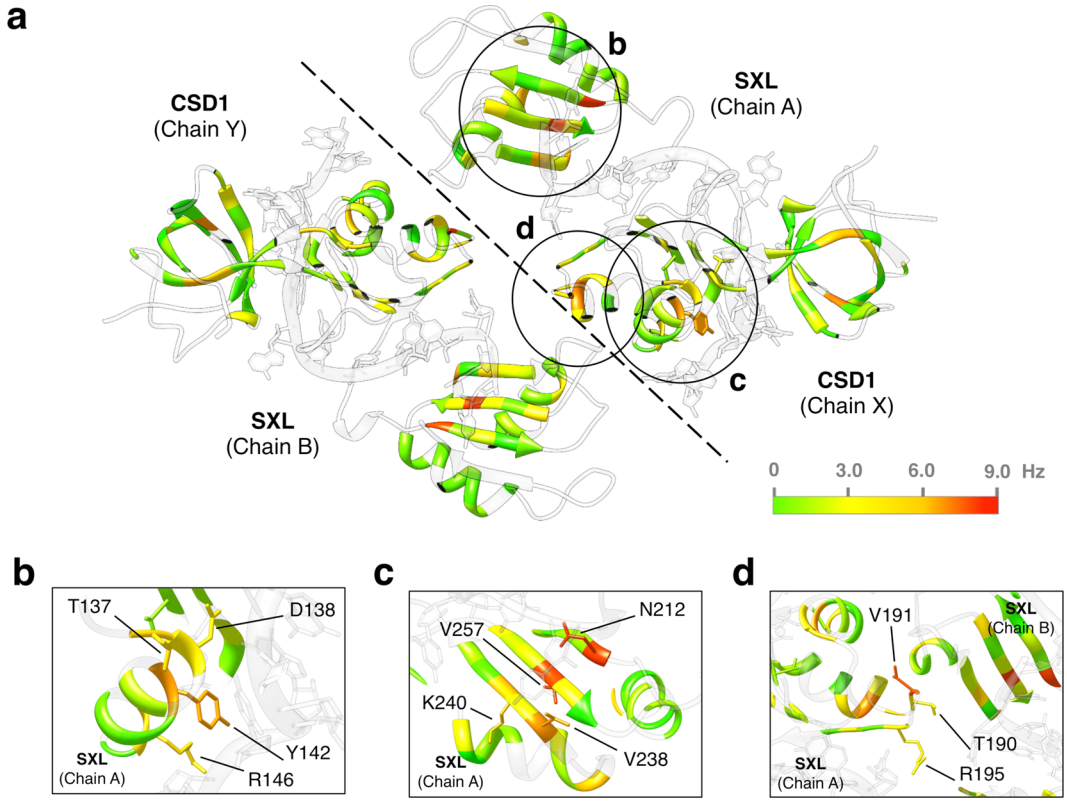


Figure 7. (a) Differences between experimental RDCs and values back-calculated from the original crystal structure for H^N–N interactions (in absolute value) are mapped onto two X-ray models of the Sxl–Unr complex according to the color code reported in the bar. (b–d) Zoomed views for distinct residues where large deviations (i.e., sizably exceeding the experimental uncertainty) are observed at the protein–RNA interface.

crystal state conformations and thus provide complementary information. Local conformational inaccuracies can arise from the uncertainty in the exact position of atoms within the electron density maps of X-rays, referred to as structural noise. These inaccuracies can be detected by NMR data, which provide complementary information about bond angles and moiety orientations. NMR data provide a direct experimental handle on these structural properties.

In the present example of Sxl–Unr translation regulatory complex, where an unsatisfactory fit of diamagnetic RDCs was obtained against the original X-ray structure, REFMAC-NMR refinement produced an effective improvement in the quality of the fit, with a drop of the *Q* factor from 0.440 to 0.144. This confirms that the solution and crystal structures of this complex are sufficiently close, and thus, the general conclusion drawn from analysis of the crystal structure are unaffected. On the other hand, the small but significant changes are not trivially due to adjustments of N–H bond vector orientation but rather reflect non-negligible backbone rearrangements. In turn, these backbone rearrangements, although small, imply some adjustments in the corresponding side chains. Comparison of the tensors calculated independently for the different subunits reasonably exclude interdomain motion effects. Moreover, the results from the step-by-step refinement confirm that both the intradomain conformations and the interdomain positions observed in the crystal are very good models also for the solution state. We also demonstrated that the reduction of the structural noise (i.e., of the uncertainty in the atomic positions within the electron density map, which imposes systematic errors in the calculated values of RDCs[26]) accomplished by the inclusion of RDCs in the structural refinement can reveal minor but interesting conformational differences between crystal and solution conformations, which help to better understand the structural biology of the studied complex. These few differences could only be spotted after successful reduction of the many differences that could be taken care of by a minor perturbation of the structure within the X-ray structure uncertainty. Indeed, joint refinement yields access to local structural differences that escaped detection when using the two methods separately and points to differences in the protein–RNA interface, which may be relevant for understanding the biological function of the complex. Detecting differences between solution and crystalline states can also help to rule out effects and artifacts from crystal packing and to identify interactions that may be important in the assembly of higher order complexes and thus guide follow-up studies. Thus, simultaneous refinement helps to understand phenomena observed in solution, which cannot be directly explained from the crystal structure alone.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b11598.

Tables of force constants and tolerances used for the refinement calculations and of alignment tensors; figures of RDC residuals, dihedral angles, residue-by-residue backbone RMSD, heteronuclear $\{^1H\}-^{15}N$ NOE values for the free CSD1, ^{15}N relaxation data for free Sxl (PDF)

Refined PDB structure (PDB)

DOI: 10.1021/jacs.5b11598
J. Am. Chem. Soc. 2016, 138, 1601–1610

AUTHOR INFORMATION

Corresponding Authors

*sattler@helmholtz-muenchen.de

*claudioluchinat@cerm.unifi.it

Present Address

[‖]Structural and Computational Biology Unit, EMBL Heidelberg, 69117 Heidelberg, Germany

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by The Deutsche Forschungsgemeinschaft (DFG, SFB1035 and GRK1721 to M.S.), the Ente Cassa di Risparmio di Firenze (C.L.), MIUR PRIN 2012SK7ASN (C.L), European Commission (284209 to C.L., 317127 C.L. and G.P.), the EU ESFRI Instruct Core Centre CERM, and the Center for Integrated Protein Science Munich (CIPSM). J.H. acknowledges the EMBO for a long-term fellowship (ALTF-276-2010) and the Swedish Research Council (Vetenskapsrådet) for a postdoctoral fellowship. E.R. holds a FIRC fellowship "Guglielmina Locatello e Gino Mazzega" (17941).

REFERENCES

(1) Brunger, A. T. *Nat. Struct. Biol.* **1997**, *4* (Suppl), 862.

(2) Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.* **2001**, *8*, 990.

(3) Goto, N. K.; Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Mol. Biol.* **2001**, *308*, 745.

(4) Sikic, K.; Tomic, S.; Carugo, O. *Open Biochem. J.* **2010**, *4*, 83.

(5) Volkov, A. N.; Worrall, J. A. R.; Holtzmann, E.; Ubbink, M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 18945.

(6) Ryabov, Y. E.; Fushman, D. *J. Am. Chem. Soc.* **2007**, *129*, 3315.

(7) Tang, C.; Schwieters, C. D.; Clore, G. M. *Nature* **2007**, *449*, 1078.

(8) Bertini, I.; Kursula, P.; Luchinat, C.; Parigi, G.; Vahokoski, J.; Wilmanns, M.; Yuan, J. *J. Am. Chem. Soc.* **2009**, *131*, 5134.

(9) Bashir, Q.; Volkov, A. N.; Ullmann, G. M.; Ubbink, M. *J. Am. Chem. Soc.* **2010**, *132*, 241.

(10) Mackereth, C. D.; Madl, T.; Bonnal, S.; Simon, B.; Zanier, K.; Gasch, A.; Rybin, V.; Valcárcel, J.; Sattler, M. *Nature* **2011**, *475*, 408.

(11) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179.

(12) Tian, F.; Valafar, H.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 11791.

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

(14) Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T. *Nature* **2009**, *462*, 669.

(15) Tugarinov, V.; Kay, L. E.; Ibraghimov, I.; Orekhov, V. Y. *J. Am. Chem. Soc.* **2005**, *127*, 2767.

(16) Riek, R.; Pervushin, K.; Wüthrich, K. *Trends Biochem. Sci.* **2000**, *25*, 462.

(17) Matzapetakis, M.; Turano, P.; Theil, E. C.; Bertini, I. *J. Biomol. NMR* **2007**, *38*, 237.

(18) Blackledge, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23.

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

(20) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098.

(21) Meiler, J.; Peti, W.; Griesinger, C. *J. Am. Chem. Soc.* **2003**, *125*, 8072.

(22) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923.

(23) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471.

(24) Maltsev, A. S.; Grishaev, A.; Roche, J.; Zasloff, M.; Bax, A. *J. Am. Chem. Soc.* **2014**, *136*, 3752.

(25) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128.

(26) Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2002**, *23*, 127.

(27) Han, B.; Liu, Y.; Ginzinger, S.; Wishart, D. S. *J. Biomol. NMR* **2011**, *50*, 43.

(28) Li, F.; Lee, J. H.; Grishaev, A.; Ying, J.; Bax, A. *ChemPhysChem* **2015**, *16*, 572.

(29) Chou, J. J.; Li, S.; Bax, A. *J. Biomol. NMR* **2000**, *18*, 217.

(30) Prestegard, J. H.; Mayer, K. L.; Valafar, H.; Benison, G. C. *Methods Enzymol.* **2005**, *394*, 175.

(31) Simon, B.; Madl, T.; Mackereth, C. D.; Nilges, M.; Sattler, M. *Angew. Chem., Int. Ed.* **2010**, *49*, 1967.

(32) Rinaldelli, M.; Ravera, E.; Calderone, V.; Parigi, G.; Murshudov, G. N.; Luchinat, C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, *70*, 958.

(33) Hennig, J.; Militti, C.; Popowicz, G. M.; Wang, I.; Sonntag, M.; Geerlof, A.; Gabel, F.; Gebaur, F.; Sattler, M. *Nature* **2014**, *515*, 287.

(34) Hennig, J.; Wang, I.; Sonntag, M.; Gabel, F.; Sattler, M. *J. Biomol. NMR* **2013**, *56*, 17.

(35) Bertini, I.; Case, D. A.; Ferella, L.; Giachetti, A.; Rosato, A. *Bioinformatics* **2011**, *27*, 2384.

(36) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kilossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T., Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; University of California: San Francisco, CA, 2012.

(37) Rinaldelli, M.; Carlon, A.; Ravera, E.; Parigi, G.; Luchinat, C. *J. Biomol. NMR* **2015**, *61*, 21.

(38) Wassenaar, T. A.; van Dijk, M.; Loureiro-Ferreira, N.; van der Schot, G.; de Vries, S. J.; Schmitz, C.; van der Zwan, J.; Boelens, R.; Giachetti, A.; Ferella, L.; Rosato, A.; Bertini, I.; Herrmann, T.; Jonker, H. R. A.; Bagaria, A.; Jaravine, V.; Guntert, P.; Schwalbe, H.; Vranken, W. F.; Doreleijers, J. F.; Vriend, G.; Vuister, G. W.; Franke, D.; Kikhney, A.; Svergun, D. I.; Fogh, R. H.; Ionides, J.; Laue, E. D.; Spronk, C.; Jurksa, S.; Verlato, M.; Badoer, S.; Dal Pra, S.; Mazzucato, M.; Frizziero, E.; Bonvin, A. M. J. J. *J. Grid Computing* **2012**, *10*, 743.

(39) Losonczi, J. A.; Prestegard, J. H. *J. Biomol. NMR* **1998**, *12*, 447.

(40) Bertini, I.; Del Bianco, C.; Gelis, I.; Katsaros, N.; Luchinat, C.; Parigi, G.; Peana, M.; Provenzani, A.; Zoroddu, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6841.

(41) Russo, L.; Maestre-Martinez, M.; Wolff, S.; Becker, S.; Griesinger, C. *J. Am. Chem. Soc.* **2013**, *135*, 17111.

(42) 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.

(43) Ahlner, A.; Carlsson, M.; Jonsson, B. H.; Lundstrom, P. *J. Biomol. NMR* **2013**, *56*, 191.

(44) Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12.

(45) Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.; Joosten, R. P.; Vriend, G. *Nucleic Acids Res.* **2015**, *43*, D364−D368.

(46) Handa, N.; Nureki, O.; Kurimoto, K.; Kim, I.; Sakamoto, H.; Shin, D. I.; Muto, Y.; Yokoyama, S. *Nature* **1999**, *398*, 579.

DOI: 10.1021/jacs.5b11598
J. Am. Chem. Soc. 2016, 138, 1601−1610