

Solution structure of tRNA^{Val} from refinement of homology model against residual dipolar coupling and SAXS data

Alexander Grishaev · Jinfa Ying · Marella D. Canny · Arthur Pardi · Ad Bax

Received: 5 August 2008 / Accepted: 12 August 2008 / Published online: 12 September 2008
© US Government 2008

Abstract A procedure is presented for refinement of a homology model of *E. coli* tRNA^{Val}, originally based on the X-ray structure of yeast tRNA^{Phe}, using experimental residual dipolar coupling (RDC) and small angle X-ray scattering (SAXS) data. A spherical sampling algorithm is described for refinement against SAXS data that does not require a globbic approximation, which is particularly important for nucleic acids where such approximations are less appropriate. Substantially higher speed of the algorithm also makes its application favorable for proteins. In addition to the SAXS data, the structure refinement employed a sparse set of NMR data consisting of 24 imino N–H^N RDCs measured with Pf1 phage alignment, and 20 imino N–H^N RDCs obtained from magnetic field dependent alignment of tRNA^{Val}. The refinement strategy aims to largely retain the local geometry of the 58% identical tRNA^{Phe} by ensuring that the atomic coordinates for short, overlapping segments of the ribose-phosphate backbone and the conserved base pairs remain close to those of the starting model. Local coordinate restraints are enforced

using the non-crystallographic symmetry (NCS) term in the XPLOR-NIH or CNS software package, while still permitting modest movements of adjacent segments. The RDCs mainly drive the relative orientation of the helical arms, whereas the SAXS restraints ensure an overall molecular shape compatible with experimental scattering data. The resulting structure exhibits good cross-validation statistics (jack-knifed $Q_{\text{free}} = 14\%$ for the Pf1 RDCs, compared to 25% for the starting model) and exhibits a larger angle between the two helical arms than observed in the X-ray structure of tRNA^{Phe}, in agreement with previous NMR-based tRNA^{Val} models.

Keywords NMR · RDC · Refinement · Rigid body · SAXS · tRNA

Abbreviations

MSA Magnetic susceptibility anisotropy
NCS Non-crystallographic symmetry
RDC Residual dipolar coupling
SAXS Small angle X-ray scattering
rms Root mean square

Electronic supplementary material The online version of this article (doi:10.1007/s10858-008-9267-x) contains supplementary material, which is available to authorized users.

A. Grishaev (✉) · J. Ying · A. Bax (✉)
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA
e-mail: AlexanderG@intra.niddk.nih.gov

A. Bax
e-mail: bax@nih.gov

M. D. Canny · A. Pardi (✉)
Department of Chemistry and Biochemistry, 215 UCB,
University of Colorado, Boulder, Boulder, CO 80309-0215, USA
e-mail: Arthur.Pardi@Colorado.edu

Introduction

Structure determination of larger nucleic acids by conventional solution NMR methods, based on NOEs and J-couplings, can be challenging since these highly helical structures exhibit low proton density and few, if any, NOEs between elements of secondary structure (Allain and Varani 1997). While the NOEs and through-hydrogen-bond J couplings (Dingley et al. 1999) provide valuable information for resonance assignment and identification of

base-pair partners, they usually provide little information on the relative positioning of individual helices in a multi-helix system. For larger nucleic acid structures, the use of ^{13}C labeling is limited due to the associated considerable ^1H line broadening, arising from the strong ^1H – ^{13}C dipolar interactions, and the very limited ^{13}C chemical shift dispersion in the helical regions. As a consequence, measurement of an extensive number of $^1\text{D}_{\text{CH}}$ residual dipolar couplings (RDCs) can be challenging, and often only $^1\text{D}_{\text{HN}}$ RDCs for base-paired imino protons are readily accessible in such systems.

Since the number of experimental imino $^1\text{D}_{\text{NH}}$ RDCs restraints is far smaller than the total number of torsion angles in the molecule, de novo structures cannot be determined from these restraints alone. An alternate strategy, which supplements the sparse set of NMR restraints with prior structural information, can yield reasonable nucleic acid models that agree with the experimental NMR restraints (Mollova et al. 2000; Vermeulen et al. 2000, 2005; Lukavsky et al. 2003; D'Souza et al. 2004; Getz et al. 2007).

Here, we report an extension of this approach, which describes a systematic procedure for refining a starting structure, based on homology modeling, by combining a sparse set of experimental $^1\text{D}_{\text{HN}}$ RDCs, with SAXS data. The idea behind the refinement protocol is to ensure that, at the local level, structural changes relative to the starting model are kept to a minimum by restraining the coordinates of short overlapping sections of the refined model to remain close to those of the starting homology model. At the same time, potentially larger changes in global structure but only smaller local changes along the backbone of the oligonucleotide ensure agreement with experimental RDCs as well as a molecular shape compatible with the SAXS data. Similar approaches, lacking SAXS input, previously have been used successfully for proteins (Chou et al. 2001; Ulmer et al. 2003). We demonstrate the refinement protocol for the 76-nt tRNA^{Val} using only the NMR restraints obtained from 24 imino N–H vectors, measured under two different alignment orientations. This low number of experimental NMR data required restraining the local geometry of the RNA close to that of the X-ray structure (PDB entry 1EHZ, (Shi and Moore 2000)) of the highly homologous tRNA^{Phe} (58% sequence identity). The orientational information contained in the RDCs was complemented by global shape information obtained from SAXS data, recorded under conditions very similar to that of the NMR data acquisition. Several recent studies (Grishaev et al. 2005, 2008; Gabel et al. 2006; Schwieters and Clore 2007; Zuo et al. 2008) have demonstrated the power of combining the molecular shape information, encoded in SAXS data, with global orientational restraints obtained from RDCs. For tRNA^{Val}, the refinement procedure is shown to result in a structure with excellent RDC

cross validation statistics ($Q_{\text{free}} = 14\%$), compared to 25% for the starting model.

Materials and methods

Sample preparation

Samples of uniformly ^{15}N -enriched native tRNA^{Val} from *E. coli* were prepared as described previously (Latham et al. 2008). The NMR sample contained 0.5 mM tRNA^{Val} in a buffer containing 10 mM sodium phosphate, pH 6.8, 80 mM NaCl, 5 mM MgCl₂, 0.1 mM EDTA, in 10% D₂O, 90% H₂O. For measurements in the Pf1-aligned phase, ca. 8 mg ml⁻¹ Pf1 was added, yielding a ^2H solvent quadrupole splitting of 7.9 Hz. Samples for SAXS data collection were prepared by dialysis against a buffer that contained 10 mM sodium phosphate, 150 mM NaCl, 5 mM MgCl₂ and 0.1 mM EDTA at pH 7.0, with the final stock concentration of tRNA^{Val} at 10 mg ml⁻¹.

RDC data collection and processing

Resonance assignments are taken from (Vermeulen 2003), but small changes in the resonance positions of the imino resonances of G22, G43, and G53 resulted from slightly different solvent conditions, causing assignment ambiguity for these resonances. Even though for the assignments depicted in Fig. 1 the measured RDCs for these nucleotides are in excellent agreement with predicted data, these RDCs were not used at any stage of the analysis or refinement. $^1\text{D}_{\text{HN}}$ RDCs in Pf1 medium were collected from a set of interleaved 800 MHz TROSY-HSQC and regular ^1H – ^{15}N HSQC spectra (Fig. 1), where the frequency difference in the ^{15}N dimension corresponds to $(^1\text{J}_{\text{NH}} + ^1\text{D}_{\text{NH}})/2$. Although in principle the $(^1\text{J}_{\text{NH}} + ^1\text{D}_{\text{NH}})/2$ splittings can be obtained independently either from the relative peak displacements in the ^{15}N dimension or in the ^1H dimension, in practice the measurement is best carried out in the ^{15}N dimension where line widths are narrowest and line shapes are most symmetric. In the ^1H dimension, the presence of unresolved ^1H – ^1H dipolar couplings and relaxation interference between ^1H – ^{15}N and ^1H – ^1H dipolar interactions can give rise to a slight asymmetry of the line shape that adversely impacts the accuracy of the measured $(^1\text{J}_{\text{NH}} + ^1\text{D}_{\text{NH}})/2$ splitting (Fig. 1b). Isotropic $^1\text{J}_{\text{NH}}$ splittings, which also include a very small dipolar contribution resulting from alignment due to tRNA's magnetic susceptibility anisotropy (MSA), were measured previously (Ying et al. 2007). For obtaining the Pf1 RDCs, the isotropic $(^1\text{J}_{\text{NH}} + ^1\text{D}_{\text{NH}}^{\text{MSA}})$ splittings, measured previously at the same magnetic field strength, were subtracted from the corresponding splittings, $^1\text{J}_{\text{NH}} + ^1\text{D}_{\text{NH}}$, measured in the

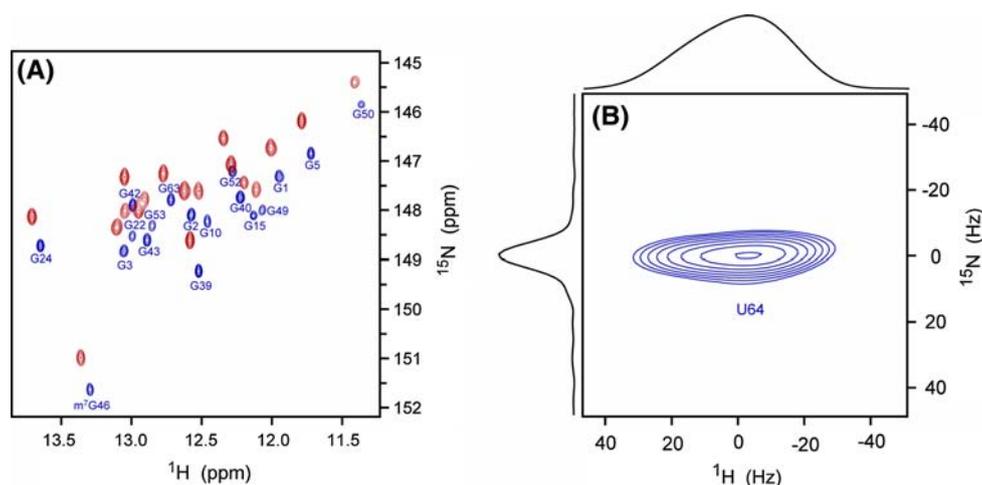


Fig. 1 800 MHz ^1H - ^{15}N shift correlation of tRNA^{Val}. **(a)** Part of the imino region of the superimposed 2D ^{15}N - ^1H TROSY-HSQC (blue) and regular HSQC (red) correlation spectra of tRNA^{Val} recorded at 800 MHz, 28°C, showing correlations for the imino groups of G nucleotides. **(b)** Expanded cross peak for the TROSY imino

correlation of U64, which forms a wobble base pair with G50 and experiences extensive homonuclear ^1H - ^1H RDC with adjacent protons. ^{15}N - $^1\text{H}/^1\text{H}$ - ^1H dipolar cross correlated relaxation results in asymmetry of the unresolved ^1H - ^1H multiplet

presence of Pf1. $^1\text{D}_{\text{HN}}^{\text{Pf1}}$ values thus obtained (Supplementary Material) only include the Pf1-induced RDC contribution and correlate closely (Pearson's correlation coefficient $R_P = 0.994$) with values measured previously (Ying et al. 2007).

SAXS data collection

Solution scattering data were acquired on a SAXSess instrument from Anton-Paar, which includes a Kratky camera equipped with high-flux multilayer optics and a wide angle measurement extension. A sealed fine-focus tube (Princeton Instruments), operating at 40 kV and 50 mA, served as the X-ray source. An elliptically bent multilayer mirror selected radiation at the Cu $K\alpha$ wavelength (1.542 Å). A 1-mm inner diameter quartz capillary of 10 mm length was used as the sample cell and kept at 25°C with a Peltier element. An X-ray beam of ca. 9 mm width, parallel to the sample capillary, was generated by adjustment of the collimator slit. Data were collected as series of sequential 2 h acquisitions on the tRNA^{Val} sample, followed immediately by the matching dialysate buffer. Due to fast signal relaxation in the first few minutes after exposure, the imaging plates were read out with a 5 min delay after each data collection on a Cyclone Plus scanner from Perkin Elmer. Data at three RNA concentrations (2.29, 4.93 and 10.0 mg ml⁻¹) were acquired in order to evaluate the magnitude of the inter-particle structure factor. The recorded scattering profiles spanned a q -range from ~ 0.02 to ~ 2.8 Å⁻¹, where $q = 4\pi\sin(\theta)/\lambda$, 2θ is the scattering angle, and λ is the wavelength of the incident radiation. The raw 2D images were converted to

1D scattering profiles by radial integration within 5 mm strips aligned at the center of the incident beam. 1D profiles were then mapped onto the q -axis by reference to the position of the primary beam, attenuated by the semi-transparent beam stop of the instrument. The converted profiles were corrected for the scanner readout noise and normalized to the recorded intensities of the primary beam. The scattering curves from the buffer were then subtracted from the scattering curves of the tRNA sample. The resulting scattering intensity curves were averaged over two independent sample/buffer data acquisitions. The line-collimation 1D profiles were desmeared using GNOM software (Svergun 1992; Svergun et al. 2001), taking into account the length and width profiles of the incident beam. The resulting point-collimation-like data were used for the subsequent structural analysis in the q interval from 0.03 to 0.35 Å⁻¹ (crystallographic resolutions between ~ 200 and ~ 18 Å). Evaluations of the quality of the fit of the scattering data to the various structural models were made with the program Crysol, version 2.5 (Svergun et al. 1995).

SAXS refinement using quasi-uniform angular averaging

A new module was developed for fitting RNA SAXS data via XPLOR-NIH or CNS which no longer requires the use of a globbic approximation and its associated correction terms, employed in several of our previous protein studies (Grishaev et al. 2005, 2008; Parsons et al. 2008). The new algorithm, following the standard description (Koch et al. 2003), represents scattering intensity predicted from a structure as

$$I(q) = \left\langle |\mathbf{F}_a(\mathbf{q}) - \rho_o \mathbf{F}_s(\mathbf{q})|^2 \right\rangle_{\Omega} \quad (1)$$

where $\mathbf{F}_a(\mathbf{q})$ and $\mathbf{F}_s(\mathbf{q})$ are the scattering amplitudes for the macromolecule and the excluded volume, respectively, ρ_o is the solvent electron density, and $\langle \rangle_{\Omega}$ denotes the solid angle average over all orientations of the momentum transfer vector \mathbf{q} for the fixed norm q .

Using a previously described dummy-solvent approximation, which assumes that displaced solvent resides exactly at the atomic positions in the macromolecule (Fraser et al. 1978), the solvent-subtracted complex scattering amplitude then becomes

$$\mathbf{F}_d(\mathbf{q}) = \sum_{j=1}^N g_j(q) \exp(i\mathbf{q}\mathbf{r}_j) \quad (2)$$

with $g_j(q)$ representing solvent-subtracted atomic scattering amplitudes and the summation extending over all atomic coordinates \mathbf{r}_j . The advantage of this expression is that it scales linearly with the number of atoms, compared to the quadratic scaling inherent in the Debye formula, used in our earlier work. The calculation is then accelerated by approximating the exact angular average of Eq. 1 by a summation over a finite number of orientations, evenly distributed on the surface of a sphere (Schwieters and Clore 2007). A spiral grid algorithm with a total of 90 angular directions gives a robust representation of the scattering data within the experimental q range (0–0.35 Å⁻¹) used in this study. In order to suppress systematic errors resulting from the finite number of equidistant angular directions, the set of spiral grid vectors is rotated by a random angle around a random axis every 50 time steps of the molecular dynamics trajectory. With this formalism, the force acting on atom m , when the experimental scattering data is $I^o(q)$ and the scattering data predicted from the current model is $I(q)$, becomes

$$\begin{aligned} \nabla_m \chi^2 = & \frac{4}{N_{\text{dat}} N_{\text{grid}}} \sum_{j=1}^{N_{\text{dat}}} c_j \frac{c_j I(q_j) - I^o(q_j)}{\sigma_j^2} g_m(q_j) \sum_{k=1}^{N_{\text{grid}}} \mathbf{q}_{jk} \\ & \times \{ \cos(\mathbf{q}_{jk} \cdot \mathbf{r}_m) \text{Im}[F_d(\mathbf{q}_{jk})] - \sin(\mathbf{q}_{jk} \cdot \mathbf{r}_m) \\ & \times \text{Re}[F_d(\mathbf{q}_{jk})] \} \end{aligned} \quad (3)$$

where c_j are the bound solvent corrections, σ_j are the experimental uncertainties, and the sums run over all data points and \mathbf{q} vector grid directions. The real and imaginary parts of the scattering amplitude in the above expression are calculated for a particular direction of the \mathbf{q} vector on the equi-spaced grid.

tRNA^{Val} structure refinement

Building a homology model for the tRNA^{Val} structure and its further refinement against experimental restraints was

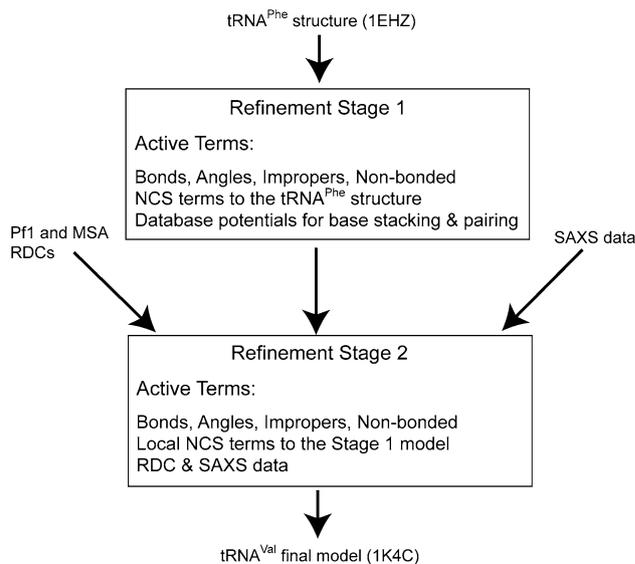


Fig. 2 Flow diagram of the two-stage process to generate the refined tRNA^{Val} structure

carried out in two stages, summarized in Fig. 2. During the first stage, a homology-based model was built starting from the X-ray structure of tRNA^{Phe} (Shi and Moore 2000). In the second stage, this homology model was refined against RDC and solution small angle X-ray scattering data. The two stages are described in detail below.

Generation of tRNA^{Phe}-based stage 1 model

A regularized tRNA^{Val} model was built on the basis of the 1.93 Å resolution X-ray structure of yeast tRNA^{Phe}, PDB code 1EHZ (Shi and Moore 2000), with hydrogens added with the program Reduce (Word et al. 1999). Generation of this “first stage” model comprised a Cartesian simulated annealing protocol, performed using XPLOR-NIH (Schwieters et al. 2003), including active energy terms for bonds, angles, improvers, repulsive-only non-bonded interactions, non-crystallographic symmetry (NCS) terms, base pairing planarity restraints (Kuszewski et al. 1997), as well as database potentials of mean force (PMF) for base stacking, pairing, and backbone dihedral angle correlations (Cai et al. 2003). The NCS module of the XPLOR-NIH program, with a force constant of 10 kcal Å⁻², was used to keep the structures of the 1EHZ X-ray reference structure and the coordinates of the tRNA^{Val} working model very close to one another. A single NCS term included all non-hydrogen atoms in the ribose-phosphate backbone and base non-hydrogen atoms for the 44 nucleotides which are identical in tRNA^{Phe} and tRNA^{Val}. Specifics of the NCS terms and their violation statistics are listed in the Supplementary Material section. The force constants for the database potentials were adjusted to yield matching PMF

energies for the tRNA^{Val} homology model and the 1EHZ tRNA^{Phe} structure. The empirical force field terms for the bonds, angles and impropers were used at their default settings. Non-bonded interactions were modeled by a repulsive-only quartic term with the van der Waals radii scaled by a factor of 0.85 and a standard force constant multiplier of 4.0 kcal Å⁻⁴.

Structure refinement with RDC and SAXS data

During the “stage 2” refinement, the above derived homology model was refined against the experimental data, comprising 24 Pf1 RDCs, 20 MSA RDCs, and the SAXS profile. To allow moderate reorientation of the helices relative to one another, while preserving the relative geometries of the stacked and base-paired bases as much as possible, a large number of local NCS terms were defined by reference to their respective parts in the rigidly held stage 1 model. The terms included all sequential pairs of nucleotides, except for the connections between the acceptor stem/D arm and anticodon stem-loop/T ψ C arm which were kept flexible, as well as all base pairs within the helices, and long-range interactions that define the three-dimensional tRNA fold. Specifics of the NCS terms are listed, along with their violation statistics, in the Supplementary Material section. The refinement calculations were carried out using a CNS (Brunger et al. 1998) torsion angle dynamics simulated annealing protocol, with the temperature ramped in 80 steps from 2001 to 1 K, and one thousand 2-fs integration steps at each temperature stage. The experimental SAXS data extended from 0.03 to 0.35 Å⁻¹ and were sparsened to 33 data points, prior to input for the refinement calculations.

Results and discussion

The accuracy of experimental data is a key consideration during any structure refinement, and this aspect becomes particularly critical when the number of experimental observables is well below the number of degrees of freedom in generating the structure, as applies to the current study. We therefore first discuss the uncertainties in the experimental input data, prior to evaluating the final structures.

RDC data quality

The close correlation relative to previously measured RDCs in Pf1 medium for tRNA^{Val} (pairwise rmsd 1.1 Hz after scaling by a factor of 0.77 to account for differences in Pf1 concentration) suggests a random error ≤ 1 Hz in either set of values. The measurement error in the MSA

RDCs, reported here as the difference between isotropic ¹J_{NH} splittings at 800 and 500 MHz, previously was estimated to be 0.3 Hz (Ying et al. 2007). The final values of the magnitude and rhombicity of the two corresponding alignment tensors, optimized during structure refinement, are 16.2 Hz and 0.570 for the Pf1 data and 0.908 Hz and 0.195 for the MSA alignment data. A ratio of 10:1 between the force constants for the MSA and Pf1 data was used in refinement, which reflects the much higher relative error of the MSA data. Note that a ratio in force constants equal to ca. 400:1 would be needed to give the two types of RDCs equal importance if their relative uncertainties would have been the same. Despite the higher relative uncertainty of the MSA data, they have a beneficial albeit small impact on the structure refinement of tRNA^{Val}.

SAXS data quality

The scattering data recorded at concentrations of 2.29, 4.93 and 10.0 mg ml⁻¹ show the presence of a non-negligible structure factor at all concentrations, presumably due to the high charge carried by the RNA, even though relatively high salt concentration (150 mM NaCl) was used in the buffer. In order to remove the effects of interparticle interference from the data, a linear extrapolation to zero concentration was performed based on the three measured concentration points. Briefly, the scattering curves at all three concentrations were aligned using the data from 0.10 to 0.35 Å⁻¹, where the effects of structure factor are negligible. The aligned data (see inset to Fig. 3a) shows the presence of structure factor at $q \leq 0.07$ Å⁻¹. Therefore, linear extrapolation was performed point-by-point below 0.07 Å⁻¹ and the extrapolated data were then merged with the 10 mg ml⁻¹ data above that value. None of the collected data show any indication of aggregation, as evidenced by $P(r)$ distributions that decay smoothly at the highest inter-atomic vector values. The extrapolated data were used for all subsequent structure analyses, with d_{\max} set to 95 Å for GNOM (Svergun 1992; Svergun et al. 2001) desmearing (Fig. 3b). The uncertainty of the scattering data, evaluated from the photon counting statistics, ranges from $\sim 0.4\%$ at $q = 0.03$ Å⁻¹ to $\sim 18\%$ at $q \sim 0.35$ Å⁻¹.

Refinement of the homology model

The regularized homology model exhibits a 0.3 Å backbone rmsd to the tRNA^{Phe} 1EHZ X-ray structure, with virtually unchanged relative positions of the four individual helices. This stage 1 model represents a near-optimal starting point for deriving a refined tRNA^{Val} structure because the 58% sequence identity with tRNA^{Phe} is the highest among tRNAs for which complete coordinates are

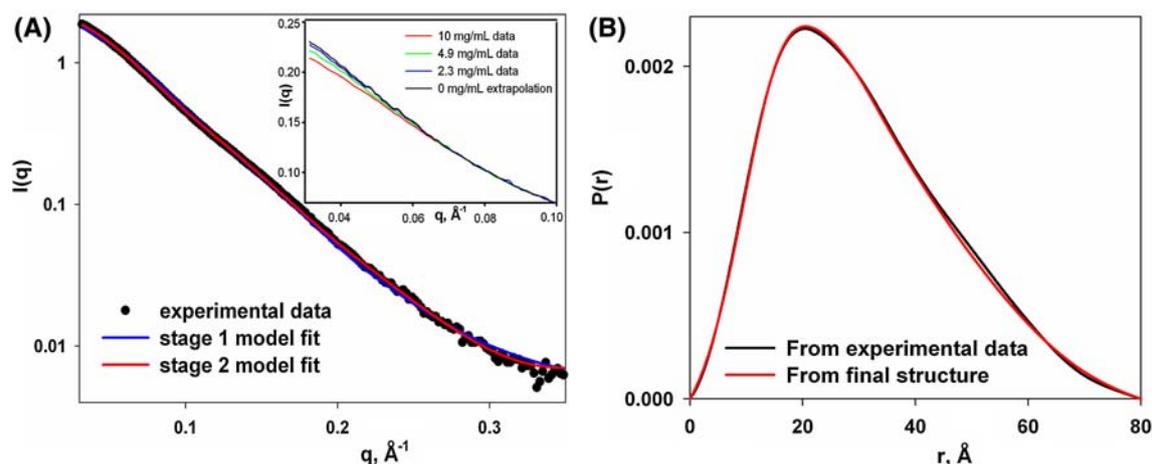


Fig. 3 SAXS data for tRNA^{Val}. (a) Desmeared experimental scattering data extrapolated to zero concentration, and their fit to the starting (stage 1; $\chi \sim 2.8$) and refined (stage 2; $\chi \sim 0.9$) tRNA^{Phe}-based models. The inset shows the effect of the structure

available; moreover, all nucleotides involved in the long-range interactions responsible for the tertiary fold of tRNA^{Phe} are strictly conserved between tRNA^{Phe} and tRNA^{Val}. For nucleotides lacking identity to the corresponding one in tRNA^{Phe}, application of the empirical database potentials (Clore and Kuszewski 2003) which impact both sequential base stacking and relative positions of the base-paired elements is used to optimize the quality of the starting model.

Although this stage 1 model fits both Pf1 RDCs (rmsd 3.02 Hz; $Q = 0.191$) and MSA RDCs (rmsd 0.34 Hz; $Q = 0.401$) very well, it should be borne in mind that such fitting, carried out by singular value decomposition (SVD), includes five adjustable parameters (Losonczi et al. 1999; Sass et al. 1999) and therefore underestimates the true error, especially when only 20–24 RDCs are being fitted. A fairer way to evaluate the errors for a set of N RDCs, which also makes comparison to the analogous results on the refined model more straightforward, uses $N - 1$ RDCs to fit the alignment tensor by SVD, and calculates the difference between the remaining observed and predicted RDC. This procedure is repeated N times, each time leaving out a different RDC, and the rmsd between the observed and predicted non-fitted couplings is then evaluated. This jack-knifing procedure results in an rmsd of 4.03 (Pf1) and 0.45 Hz (MSA) for these two sets of data (Fig. 4), corresponding to jack-knifed Q factors of 25% and 53%, respectively.

Refinement of the stage 1 model against both RDC and SAXS data resulted in a narrow bundle of structures (coordinate rmsd to average of 0.3 Å) that exhibit a ~ 2.8 Å rmsd (nt. 1–72) relative to the stage 1 model (Table 1). These refined structures are characterized by an increase in the angle between the two arms of the L-shaped tRNA from ~ 81 to $\sim 98^\circ$ (Fig. 5). An analogous, slightly

larger increase in the angle between these two arms previously was obtained by rigid-body optimization against the Pf1 RDCs alone (Vermeulen et al. 2005). Although application of only the SAXS restraints results in a slightly smaller increase in the angle between the two arms than application of just the RDC restraints (Table 1), this smaller change in global structure simply reflects the minimum change needed to get acceptable agreement with the SAXS data. The fit to the SAXS data remains equally good when the SAXS terms and RDC restraints are applied simultaneously, even though the inter-arm angles for the SAXS-only and SAXS + RDC structures differ by about 6° . Addition of the SAXS data also does not significantly impact the Q_{free} factors of the resulting structures over the use of RDC restraints alone. On the other hand, it is important to note that when omitting the RDC restraints from the stage 2 structure refinement, inclusion of the SAXS data results in improved agreement with the RDCs, as manifested in a decrease of the jack-knifed Q value from 25% to 21% (Table 1).

The final rmsd between the experimental and best-fitted RDCs is ~ 1.2 Hz for Pf1 data and ~ 0.34 Hz for the MSA data, comparable to their estimated experimental uncertainties. Bound surface water corrections, which include the scattering by counter-ions, used during the SAXS data fits required six cycles for convergence to a final value of $0.066 \text{ e} \text{ \AA}^{-3}$. This is about two-fold higher than the typical 10% solvent density increase often seen in the water layer surrounding proteins, and may reflect the presence of counter-ions associated with the high charge density of RNA.

Structural statistics of the final refined models are summarized in Table 2. Objective evaluation of the quality of the models requires cross-validation statistics where the refinement is repeated, with a given Pf1 RDC left out of the

Fig. 4 Agreement between experimental RDCs and tRNA^{Val} models. RDC^{Pf1} (a), or RDC^{MSA} (b) versus stage 1 model, generated without SAXS or RDC input data. (c) RDC^{Pf1} versus refined (stage 2) model; (d) RDC^{MSA} versus refined model. In all cases, filled symbols represent RDCs included when carrying out the SVD fit between the experimental data and the structure; open symbols are (a, b) jack-knifed RDC values, predicted from the alignment tensor when that particular RDC was not included in the SVD fit, and (c) cross-validated RDC values, predicted for the refined model calculated without that particular RDC

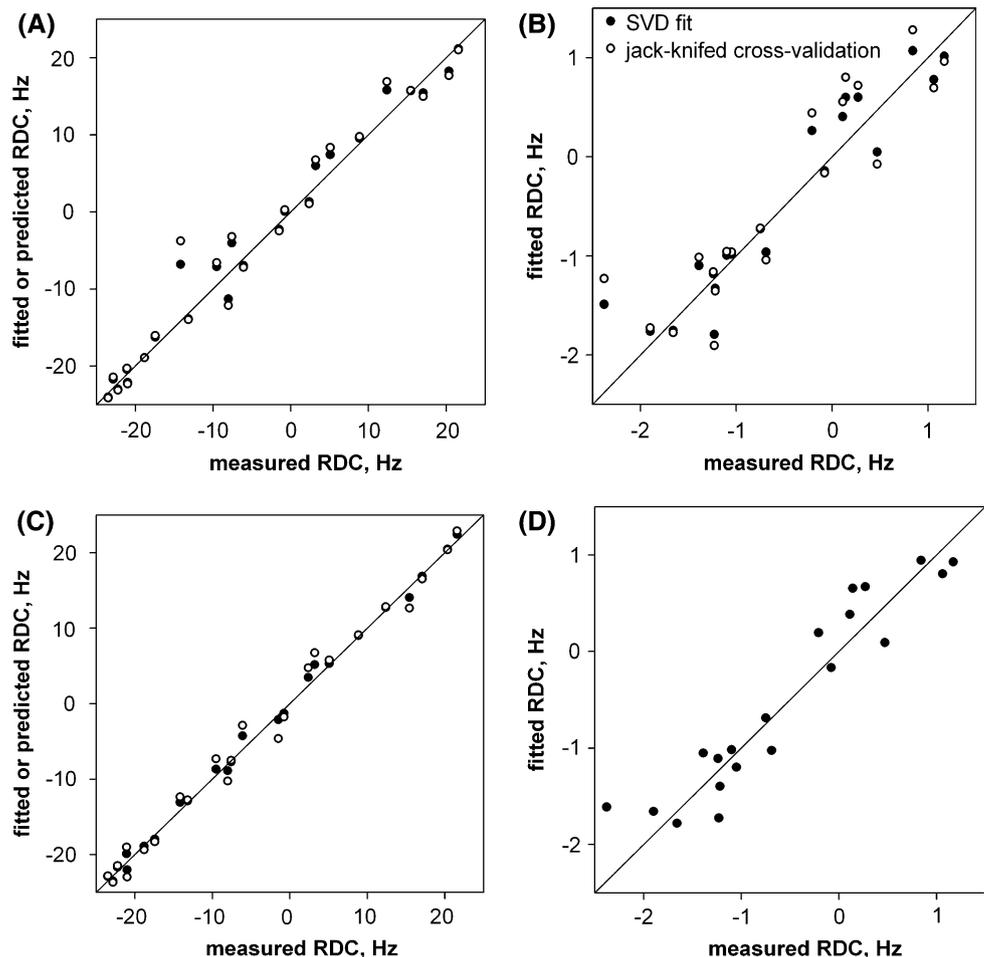


Table 1 Impact of different types of restraints on tRNA^{Val} structure during structure refinement

	$Q_{\text{free}}^{\text{Pf1}}$ (%)	Opening angle (°)	RMSD to stage 1 all heavy atoms, nt. 1–72 (Å)
Stage 1	24.7	81	
SAXS	21.1	92	2.61 ± 0.26
RDC ^{Pf1}	14.3	99	2.31 ± 0.10
RDC ^{MSA}	18.2	96	2.17 ± 0.17
RDC ^{Pf1/MSA}	13.9	98	2.21 ± 0.19
SAXS + RDC ^{Pf1/MSA}	14.2	98	2.81 ± 0.21

refinement, and an SVD fit to the remaining ones is used to predict the value of the RDC not used during refinement. This refinement protocol is then repeated 24 times for each of the Pf1 RDCs. When SAXS data are not being fitted, such jack-knifed cross-validation yields a Q_{free} of 13.9% when MSA RDCs are included and 14.3% without them. The small magnitude of the difference between the two Q_{free} values results from the relatively high uncertainty of

the MSA RDC data, and their correspondingly weak weighting factor. The final structure shows very low interatomic clashing scores (~ 5 clashes >0.4 Å per 1,000 atoms, versus ~ 18 for the stage 1 model, and ~ 23 for 1EHZ), as evaluated by Molprobity (Davis et al. 2004).

Concluding remarks

SAXS is increasingly being used to provide structural information in RNAs (Lipfert and Doniach 2007; Putnam et al. 2007). Recent applications include yeast tRNA^{Phe}, the P4–P6 domain of the *Tetrahymena* ribozyme, a glycine riboswitch and a SAM riboswitch (Lipfert et al. 2007a, b; Putnam et al. 2007). However, in the absence of other structural restraints, the scattering data only provide low-resolution structural information and cannot uniquely define 3D structure. The goal of the present study is to extend these studies by combining the SAXS data with RDC and structural restraints to the homologous tRNA^{Val} to generate a refined model for tRNA^{Val}. The RDC and SAXS data are largely complementary, where SAXS reflects the overall molecular shape and the RDCs provide

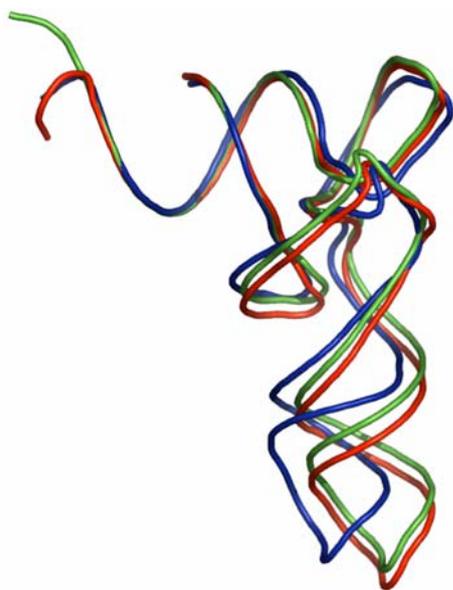


Fig. 5 Backbone structures of tRNA^{Val} obtained by refinement of a tRNA^{Phe}-based homology model. The regularized starting model (stage 1) is shown in blue; the model refined against Pf1 and MSA RDCs in green, and the model that includes both the RDC and SAXS data in red. The models are superimposed by best-fitting nucleotides 1–7 and 66–72. The absence of restraints for nucleotides 73–76 results in their disorder during the refinement simulated annealing protocol when using only RDC data, whereas they adopt a more compact shape when SAXS data are active

Table 2 Structural statistics for the RDC- and SAXS-refined tRNA^{Val} homology model

Deviations from experimental data	
Pf1 RDCs (Hz)	1.16 ± 0.04
MSA RDCs (Hz)	0.328 ± 0.004
NCS restraints (Å)	0.56 ± 0.06
SAXS data fit (χ^a)	0.91 ± 0.03
Deviations from idealized geometry	
Bond lengths (Å)	0.0046 ± 0.0001
Bond angles (°)	0.606 ± 0.001
Improper (°)	0.407 ± 0.002
Molprobrity clash score ^b	5.3 ± 0.6
Average pairwise rmsd (Å), all heavy atoms, nt. 1–72	0.31 ± 0.07

^a χ is the square root of the normalized χ^2 difference between observed and predicted scattering data

^b Clashes per 1,000 atoms

orientational constraints for the helical domains. In our study, the RDCs tightly constrain the possible orientations of the helical arms, but provide no translation information on the distance between these arms. On the other hand, the SAXS data tightly constrain the distance between the two arms, and are less sensitive to small changes in interhelical angle or twisting about a helical axis.

Refinement of any structural model on the basis of limited experimental data, each with their own inherent uncertainty, can be challenging. For example, if NOE data were to be used for such refinement, calibration of the reference distance used for extracting distances from NOE intensities can cause systematic errors. Similarly, when using SAXS data, unrecognized interparticle interference effects or transient aggregation could result in a systematic bias during refinement. For RDCs, a potential systematic problem can arise when the magnitude or rhombicity of the alignment tensor used during refinement deviates from its true value. In our refinement procedure, these values as well as the orientation of the alignment tensor were allowed to float to give the best agreement with all the experimental data (Sass et al. 2001). Furthermore, the first five RDCs have no restraining value as there are five independent parameters required to define the alignment tensor, or three parameters if the system under study were to exhibit three-fold or higher axial symmetry.

The improved cross-validation statistics obtained upon inclusion of the RDC restraints indicates higher quality of the refined model compared to the starting structure. Clearly, however, when using a very small set of experimental restraints the cross-validation statistics attainable for the refined model strongly depend on the quality of the starting structure. For example, starting from a homology model that is based on the 1.9-Å X-ray structure of yeast tRNA^{Phe} (58% identity) yields better statistics than starting from a more general model, generated on the assumption of idealized A-form helices (Supplementary Material). Starting from this latter model, which yields relatively poor agreement with the RDCs ($Q_{\text{free}} = 53\%$), refinement against RDC and SAXS data again yields considerable improvement. The final refined structure falls close (2.1 Å coordinate rmsd) to that of our refined homology model (PDB entry 2K4C), albeit with less favorable cross validation statistics ($Q_{\text{free}} = 28\%$ instead of 14%; see Supplementary Material). In this respect, it is important to note that Q_{free} simply reports on the orientations of imino N–H vectors relative to the alignment frame, which are impacted by both the global structure (e.g. interhelical angles) and by local structural noise (Zweckstetter and Bax 2002). Because the number of experimental restraints is far smaller than the number of parameters that define N–H vector orientations, the refinement protocol is fundamentally limited in its ability to remove local structural noise. Improvements in cross validation are therefore dominated by the more global changes in structure.

It is interesting that even the use of only SAXS data during refinement already results in a significant improvement in the fit of the high precision Pf1 RDC data to the model. A similar improvement in the fit of the Pf1 RDC data occurs when using only the MSA RDC values during

refinement (Table 1). Due to the relatively large fractional measurement error in the very small MSA RDCs, they are only enforced with a weak force constant to prevent introduction of local distortions, and their impact on changing the global structure during refinement is therefore limited.

The procedure used in our refinement aims to keep local structure close to that of the starting model by requiring similar geometries for short, overlapping segments in the polymer, and conserved hydrogen bonding where indicated by homology. At the same time, these local geometries are not completely frozen and permit gradual changes along the polymer backbone. In principle, more abrupt changes are also easily accommodated, and this may be appropriate when indicated by a lack of homology in a certain region or by another perturbation such as a ligand binding event, marked by a chemical shift change. Although computationally quite demanding, the refinement approach used in our study strikes a balance between full-fledged structure calculations, which would require far more experimental input parameters, and the widely used procedure of rigid body refinement (Wang et al. 2000; Clore and Bewley 2002; Cai et al. 2003; Jain et al. 2004; Vermeulen et al. 2005; Tang et al. 2006; Bhatnagar et al. 2007).

Our refinement procedure relies on the use of a large number of NCS terms that serve to minimize local structural changes. Therefore, the refinement procedure reaches the solution closest to the starting structure (in terms of local rmsd) that is in satisfactory agreement with the experimental data. These NCS terms also extend to tertiary interactions that define pairing of the helices and tertiary structure of the tRNA. In cases where the secondary information is available but the tertiary fold is not known, a similar approach may be applicable, but whether or not a unique (and correct) solution can be obtained ultimately depends on the amount and quality of the available data and the specifics of the particular structure. In such cases, the NCS terms can be applied in the same way for the helical segments, but not for segments involving any unknown long-range tertiary interaction. Serious complications can arise when the inter-helical linkages are flexible, resulting in the absence of fixed orientations and/or translations between the individual helical units. In favorable cases, where a sufficient number of RDCs is available for each helical segment to determine an alignment tensor, such flexibility can be recognized if the alignment strengths of the helices differ, or it may manifest itself by different relaxation characteristics of the helical segments (Zhang et al. 2006). Although detailed information regarding such flexible structures can be obtained from NMR data by resorting to cleverly chosen modifications of the molecular system (Zhang et al. 2006; Bailor et al. 2007), data collected for a single molecule in a single liquid crystalline medium generally will be insufficient to uniquely define average orientations.

For a rigid system consisting of N helical segments, the degeneracy of the RDCs with respect to 180° rotations around each of the three principal axes of the alignment tensor results in 4^{N-1} distinct conformations (Al-Hashimi et al. 2000; Latham et al. 2008). Whether or not a unique solution can be selected from such a set depends on whether all but one of the 4^{N-1} conformations can be ruled out due to steric clashes, linkage strain, etc. SAXS data will also aid in filtering out incorrect conformations, but there is no guarantee that a unique solution will emerge. In either case, such a solution would have to be validated by the analysis of additional data which might include observed NOEs or comparison between the alignment tensor parameters predicted from a procedure such as PALES (Zweckstetter et al. 2004) and the experimentally observed ones. Although NOE analysis at the early stages of structure determination is often hampered by extensive resonance overlap in A-form RNA, once the set of solutions is restricted to a small number of structures, identification of long-range NOEs can become much easier. These considerations suggest that our “hybrid” approach to refining a multi-helical A-form RNA structure against a small number of RDCs and SAXS data may be applicable even in cases where the inter-helical connections are not known a priori.

Supplementary information available

Description of the NCS terms used in the refinement and their violation statistics; description of the refinement procedure starting from the idealized A-form tRNA model and its results; table with RDCs observed in tRNA^{Val}.

Coordinates deposited to the RCSB Protein Data Bank under reference number 2K4C.

Software available

Scripts used for model refinement and source code for Xplor-NIH and CNS modules for SAXS data refinement via procedures described in this paper can be downloaded from <http://spin.niddk.nih.gov/bax/software/>.

Acknowledgments This work was supported by the Intramural Research Program of the NIDDK, NIH, and by the Intramural AIDS-Targeted Antiviral Program of the Office of the Director, NIH and NIH grant AI33098 (AP).

References

- Al-Hashimi HM, Valafar H, Terrell M, Zartler ER, Eidsness MK, Prestegard JH (2000) Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *J Magn Reson* 143:402–406

- Allain FHT, Varani G (1997) How accurately and precisely can RNA structure be determined by NMR? *J Mol Biol* 267:338–351
- Bailor MH, Musselman C, Hansen AL, Gulati K, Patel DJ, Al-Hashimi HM (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat Protoc* 2:1536–1546
- Bhatnagar J, Freed JH, Crane BR (2007) Rigid body refinement of protein complexes with long-range distance restraints from pulsed dipolar ESR. *Meth Enzymol* 423:117–133
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921
- Cai ML, Williams DC, Wang GS, Lee BR, Peterkofsky A, Clore GM (2003) Solution structure of the phosphoryl transfer complex between the signal-transducing protein IIA(Glucose) and the cytoplasmic domain of the glucose transporter IICBGlucose of the *Escherichia coli* glucose phosphotransferase system. *J Biol Chem* 278:25191–25206
- Chou JJ, Li SP, Klee CB, Bax A (2001) Solution structure of Ca²⁺-calmodulin reveals flexible hand-like properties of its domains. *Nat Struct Biol* 8:990–997
- Clore GM, Bewley CA (2002) Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings. *J Magn Reson* 154:329–335
- Clore GM, Kuszewski J (2003) Improving the accuracy of NMR structures of RNA by means of conformational database potentials of mean force as assessed by complete dipolar coupling cross-validation. *J Am Chem Soc* 125:1518–1525
- Davis IW, Murray LW, Richardson JS, Richardson DC (2004) MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* 32:W615–W619
- Dingley AJ, Masse JE, Peterson RD, Barfield M, Feigon J, Grzesiek S (1999) Internucleotide scalar couplings across hydrogen bonds in Watson-Crick and Hoogsteen base pairs of a DNA triplex. *J Am Chem Soc* 121:6019–6027
- D'Souza V, Dey A, Habib D, Summers MF (2004) NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J Mol Biol* 337:427–442
- Fraser RDB, Macrae TP, Suzuki E (1978) Improved method for calculating contribution of solvent to X-ray-diffraction pattern of biological molecules. *J Appl Crystallogr* 11:693–694
- Gabel F, Simon B, Sattler M (2006) A target function for quaternary structural refinement from small angle scattering and NMR orientational restraints. *Eur Biophys J Biophys Lett* 35:313–327
- Getz M, Sun XY, Casiano-Negroni A, Zhang Q, Al-Hashimi HM (2007) NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers* 86:384–402
- Grishaev A, Wu J, Trewella J, Bax A (2005) Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J Am Chem Soc* 127:16621–16628
- Grishaev A, Tugarinov V, Kay LE, Trewella J, Bax A (2008) Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *J Biomol NMR* 40:95–106
- Jain NU, Wyckoff TJO, Raetz CRH, Prestegard JH (2004) Rapid analysis of large protein-protein complexes using NMR-derived orientational constraints: the 95 kDa complex of LpxA with acyl carrier protein. *J Mol Biol* 343:1379–1389
- Koch MHJ, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* 36:147–227
- Kuszewski J, Gronenborn AM, Clore GM (1997) Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J Magn Reson* 125:171–177
- Latham MP, Hanson P, Brown DJ, Pardi A (2008) Comparison of alignment tensors generated for native tRNA(Val) using magnetic fields and liquid crystalline media. *J Biomol NMR* 40:83–94
- Lipfert J, Doniach S (2007) Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu Rev Biophys Biomol Struct* 36:307–327
- Lipfert J, Chu VB, Bai Y, Herschlag D, Doniach S (2007a) Low-resolution models for nucleic acids from small-angle X-ray scattering with applications to electrostatic modeling. *J Appl Crystallogr* 40:S229–S234
- Lipfert J, Das R, Chu VB, Kudaravalli M, Boyd N, Herschlag D, Doniach S (2007b) Structural transitions and thermodynamics of a glycine-dependent riboswitch from *Vibrio cholerae*. *J Mol Biol* 365:1393–1406
- Losonczi JA, Andrec M, Fischer MWF, Prestegard JH (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson* 138:334–342
- Lukavsky PJ, Kim I, Otto GA, Puglisi JD (2003) Structure of HCVIRES domain II determined by NMR. *Nat Struct Biol* 10:1033–1038
- Mollova ET, Hansen MR, Pardi A (2000) Global structure of RNA determined with residual dipolar couplings. *J Am Chem Soc* 122:11561–11562
- Parsons LM, Grishaev A, Bax A (2008) The periplasmic domain of To1R from haemophilus influenzae forms a dimer with a large hydrophobic groove: NMR solution structure and comparison to SAXS data. *Biochemistry* 47:3131–3142
- Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 40:191–285
- Sass J, Cordier F, Hoffmann A, Rogowski M, Cousin A, Omichinski JG, Lowen H, Grzesiek S (1999) Purple membrane induced alignment of biological macromolecules in the magnetic field. *J Am Chem Soc* 121:2047–2055
- Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S (2001) An easy way to include weak alignment constraints into NMR structure calculations. *J Biomol NMR* 21:275–280
- Schwieters CD, Clore GM (2007) A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry* 46:1152–1166
- Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160:65–73
- Shi HJ, Moore PB (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 angstrom resolution: a classic structure revisited. *RNA-Publ. RNA Soc* 6:1091–1105
- Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25:495–503
- Svergun D, Barberato C, Koch MHJ (1995) CRYSOLE—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773
- Svergun DI, Petoukhov MV, Koch MHJ (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80:2946–2953
- Tang C, Iwahara J, Clore GM (2006) Visualization of transient encounter complexes in protein-protein association. *Nature* 444:383–386

- Ulmer TS, Ramirez BE, Delaglio F, Bax A (2003) Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *J Am Chem Soc* 125: 9179–9191
- Vermeulen A (2003) Determining nucleic acid global structure by application of NMR residual dipolar couplings. PhD, University of Colorado, Boulder
- Vermeulen A, Zhou H, Pardi A (2000) Determining DNA global structure and DNA bending by application of NMR residual dipolar couplings. *J Am Chem Soc* 122:9638–9647
- Vermeulen A, McCallum SA, Pardi A (2005) Comparison of the global structure and dynamics of native and unmodified tRNA. *Biochemistry* 44:6024–6033
- Wang GS, Louis JM, Sondej M, Seok YJ, Peterkofsky A, Clore GM (2000) Solution structure of the phosphoryl transfer complex between the signal transducing proteins HPr and IIA(Glucose) of the *Escherichia coli* phosphoenolpyruvate: sugar phosphotransferase system. *EMBO J* 19:5635–5649
- Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 285: 1735–1747
- Ying JF, Grishaev A, Latham MP, Pardi A, Bax A (2007) Magnetic field induced residual dipolar couplings of imino groups in nucleic acids from measurements at a single magnetic field. *J Biomol NMR* 39:91–96
- Zhang Q, Sun XY, Watt ED, Al-Hashimi HM (2006) Resolving the motional modes that code for RNA adaptation. *Science* 311: 653–656
- Zuo XB, Wang JB, Foster TR, Schwieters CD, Tiede DM, Butcher SE, Wang YX (2008) Global molecular structure and interfaces: refining an RNA: RNA complex structure using solution X-ray scattering data. *J Am Chem Soc* 130:3292–3293
- Zweckstetter M, Bax A (2002) Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *J Biomol NMR* 23: 127–137
- Zweckstetter M, Hummer G, Bax A (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophys J* 86:3444–3460