

## RNA Structure Determination Using SAXS Data

Sichun Yang,<sup>†</sup> Marc Parisien,<sup>‡</sup> François Major,<sup>\*,‡</sup> and Benoît Roux<sup>\*,†</sup>

Department of Biochemistry and Molecular Biology, 929 East 57th Street, University of Chicago, Chicago, Illinois 60637, and Institute for Research in Immunology and Cancer and Department of Computer Science and Operations Research, University of Montreal, Montreal, Canada

Received: June 21, 2010

Exploiting the experimental information from small-angle X-ray solution scattering (SAXS) in conjunction with structure prediction algorithms can be advantageous in the case of ribonucleic acids (RNA), where global restraints on the 3D fold are often lacking. Traditional usage of SAXS data often starts by attempting to reconstruct the molecular shape *ab initio*, which is subsequently used to assess the quality of a model. Here, an alternative strategy is explored whereby the models from a very large decoy set are directly sorted according to their fit to the SAXS data. For rapid computation of SAXS patterns, the method developed here makes use of a coarse-grained representation of RNA. It also accounts for the explicit treatment of the contribution to the scattering of water molecules and ions surrounding the RNA. The method, called Fast-SAXS-RNA, is first calibrated using a tRNA (tRNA-val) and then tested on the P4–P6 fragment of group I intron (P4–P6). Fast-SAXS-RNA is then used as a filter for decoy models generated by the MC-Fold and MC-Sym pipeline, a suite of RNA 3D all-atom structure algorithms that encode and exploit RNA 3D architectural principles. The ability of Fast-SAXS-RNA to discriminate native folds is tested against three widely used RNA molecules in molecular modeling benchmarks: the tRNA, the P4–P6, and a synthetic hairpin suspected to assemble into a homodimer. For each molecule, a large pool of decoys are generated, scored, and ranked using Fast-SAXS-RNA. The method is able to identify low-rmsd models among top ranking structures, for both tRNA and P4–P6. For the hairpin, the approach correctly identifies the dimeric state as the solution structure over the monomeric state and alternative secondary structures. The method offers a powerful strategy for recognizing native RNA conformations as well as multimeric assemblies and alternative secondary structures, thus enabling high-throughput RNA structure determination using SAXS data.

## Introduction

Small-angle X-ray solution scattering (SAXS) is an experimental method that can provide low-resolution structural information about macromolecules in solution.<sup>1–7</sup> The method is powerful because it allows for structural characterization of biomolecular complexes in various solution conditions, without the need of growing ordered crystals as in X-ray crystallography. Ribonucleic acids (RNA), given their key role in living cells, are often the subject of such SAXS analysis.<sup>8,9</sup>

SAXS data can be exploited in three different ways to advance our knowledge of RNA structures. For example, one can attempt to determine an *ab initio* molecular shape capable of reproducing the experimental SAXS profile.<sup>10</sup> The goodness of fit of a RNA three-dimensional (3D) model structure is then measured by performing a rigid-body docking of the model into this shape (shape-models comparison). *Ab initio* molecular shapes are generally constructed from a collection of beads and are obtained at a high computational cost. One concern is that different *ab initio* shapes can yield similar scattering profiles, giving rise to the problem of nonuniqueness of the solution. Furthermore, current *ab initio* shape generators have been designed often for globular proteins and may not be appropriate for RNA characterization with their branch-like shapes. An alternative strategy consists of deriving a real-space pair-density distribution func-

tion  $P(r)$  from the SAXS profile, which can then be interpreted as the probability of observing a pairwise distance  $r$  of two particles. The goodness of fit of a model structure is then measured by comparing its  $P(r)$  distribution to the experimentally derived one. Difficulties arise from the fact that the calculation of  $P(r)$  requires extra assumptions about the value of  $D_{\max}$ , the maximum pairwise distance within the macromolecule. In other words, although  $P(r)$  is related to the Fourier transform of the SAXS profile  $I(q)$  (where  $q$  is a scattering distance in reciprocal space,  $q = 2\pi/d$ , and  $d$  is the Bragg spacing), its accurate determination requires such a *prior*  $D_{\max}$  value (e.g., see ref 1), which might be nontrivial to obtain for macromolecules such as flexible multidomain protein complexes. Another difficulty is that the derived  $P(r)$ , because of the scattering contribution of hydration in  $I(q)$ , includes the contribution not only from the macromolecular itself but also from the surrounding hydration layer. Both the shape reconstruction and the  $P(r)$  approaches seek to transform the SAXS profile into a real-space quantity to measure the goodness of fit of an RNA 3D model, and thus attempt to solve the “inverse” problem of converting  $I(q)$  to  $P(r)$ . However, the inverse problem encounters difficulties in the introduction of the aforementioned assumption on  $D_{\max}$  as well as the completeness of the SAXS spectrum needed for the Fourier transform, which ultimately limit and undermine the full use of SAXS data. To circumvent those difficulties, one can use a “forward” approach by generating putative 3D structures and then test them directly for their goodness of fit to the raw SAXS data.<sup>8,11–13</sup> This approach has the clear advantage of avoiding any extra as-

\* To whom correspondence should be addressed. E-mail: F.M., major@iro.umontreal.ca; B.R., roux@uchicago.edu.

<sup>†</sup> University of Chicago.

<sup>‡</sup> University of Montreal.

sumptions in the interpretation of the experimental data that are required to solve the inverse scattering problem.

Several RNA 3D structure prediction methods are now available, ranging from simplified bead models to atomistically detailed models.<sup>14–20</sup> In this context, a rapid computational method for calculating the theoretical scattering profile from a given RNA model for a direct comparison with experimental SAXS profile offers clear advantages. Any given model is attributed a score,  $\chi^2$ , which allows sorting, filtering, and subsequent identification of the native fold among a pool of decoys. In this scenario, it is mandatory to develop a computational method that is capable of computing SAXS profiles from a given RNA model in an efficient and accurate fashion. Existing approaches to calculate SAXS profiles treat macromolecules at the atomic level, while the contribution from the solvent at the molecular surface is taken into account implicitly.<sup>21</sup> These approaches are too slow as well as insufficiently accurate. Given the low-resolution nature of SAXS, coarse-graining both solvent molecules and the solute macromolecule is more efficient. With this strategy, a Fast-SAXS method has been previously developed for protein scattering,<sup>22</sup> resulting in a significant reduction of computational costs without comprising any accuracy. In this paper, we extend the Fast-SAXS method to RNA scattering calculations.

## Methods

**Theoretical Background.** X-ray scattering from proteins and nucleic acids in solution essentially measures the electron density contrast between the macromolecules and bulk solvent. The scattering profile,  $I(q)$ , is determined from the scattering of macromolecular samples after the subtraction of the background buffer contribution.<sup>1</sup> Theoretically, the scattering intensity from dilute samples is proportional to the spherically averaged scattering of a single molecule, minus the excluded volume contributions and plus the excess electron density in the hydration layer,<sup>3</sup>

$$I(q) = \langle |A_m(\mathbf{q}) - \rho_s A_s(\mathbf{q}) + \Delta\rho_b A_b(\mathbf{q})|^2 \rangle_{(\Omega)} \quad (1)$$

where the amplitude of the wavevector transfer  $q = |\mathbf{q}| = 2\pi/d = 4\pi \sin \theta/\lambda$  ( $d$  is the Bragg spacing,  $\theta$  is half of the scattering angle, and  $\lambda$  is the X-ray wavelength).  $A_m(\mathbf{q})$  is the scattering amplitude from the macromolecule in vacuum,  $A_s(\mathbf{q})$  is from the solvent with an excluded volume displaced by the macromolecule, and  $A_b(\mathbf{q})$  is from the shell of bound waters reflected in the density excess ( $\Delta\rho_b$ ) relative to the bulk phase.<sup>23</sup> The brackets  $\langle \dots \rangle_{(\Omega)}$  stand for an average over all possible orientations of the macromolecule. Equation 1 provides the theoretical basis for solution scattering.

Similar to the Fast-SAXS method for protein,<sup>22</sup> a coarse-grained method is developed and applied on a given RNA conformation where the scattering contribution arises from the RNA molecule itself in vacuum, the excluded volume of solvent displaced by the RNA, and the excess electron density at the RNA surface. These three aspects are briefly described as follows. First, the scattering  $I(q)$  from the RNA itself is calculated using the Debye formula,

$$I(q) = \langle |A_m(\mathbf{q})|^2 \rangle_{(\Omega)} = \sum_{i,j=1}^n f_i(q) f_j(q) \frac{\sin(qr_{ij})}{qr_{ij}} \quad (2)$$

where  $f_i$  are atomic form factors ( $i = 1, \dots, n$ , where  $n$  is the number of atoms), analytically provided by the Cromer–Mann scattering-factor coefficients.<sup>24</sup> At the limit of  $q = 0$ ,  $f_i$  is the electron number of atom  $i$ , and  $r_{ij}$  is the distances between atoms  $i$  and  $j$ . Second, the excluded solvent effect can be incorporated into the scattering factors by assigning a Gaussian sphere approximated for the atoms,<sup>25</sup>

$$f'_i(q) = f_i(q) - v_i \rho_s \exp(-\pi v_i^{2/3} q^2) \quad (3)$$

where  $v_i$  are the observed atomic volumes from experiments.<sup>25</sup> Therefore, the scattering from the macromolecule, taking into account the excluded volume, is given by

$$I(q) = \langle |A_m(\mathbf{q}) - \rho_s A_s(\mathbf{q})|^2 \rangle = \sum_{i,j=1}^n f'_i(q) f'_j(q) \frac{\sin(qr_{ij})}{qr_{ij}} \quad (4)$$

where  $f'_i(q)$  are the scattering factors after the consideration of the excluded volume effect as in eq 3. Third, the excess electron density in the solution layer at the RNA surface also contributes to the total scattering.<sup>23,26–28</sup> This excess density gives rise to the third term in the total scattering of  $I(q)$  in eq 1. An explicit solvent treatment is implemented into Fast-SAXS calculations, as described below.

**Coarse-Grained Representation.** The low resolution intrinsic to SAXS data allows us to simplify each nucleotide into a two-particle model, where one particle represents the sugar–phosphate backbone group and the other represents the base side chain group. The positions of the particles are taken from the coordinates of the atoms that are closest to the center-of-scattering ( $\mathbf{R}_{\text{COS}}$ ) of each group, as defined by

$$\mathbf{R}_{\text{COS}} = \frac{1}{\sum_i |f'_i(0)|^2} \sum_i |f'_i(0)|^2 \mathbf{r}_i \quad (5)$$

where  $f'_i(0)$  is the solvent-corrected atomic scattering factor at  $q = 0$  and  $\mathbf{r}_i$  is the position of atom  $i$ .

One also has to replace the atomic scattering factors with effective structure factors that account for the internal details of individual nucleic electron densities, in a fashion similar to that for amino acids in protein scattering.<sup>22,29–32</sup> For each nucleotide, two effective structure factors,  $F^{\text{CG}}(q)$ , defined for their backbone and side chain groups, are derived from a set of high-resolution nucleic atomic coordinates, using the Debye formula,<sup>33</sup>

$$F^{\text{CG}}(q) = \left\langle \sum_{i,j=1}^m f_i(q) f_j(q) \frac{\sin(qr_{ij})}{qr_{ij}} \right\rangle_{\text{PDB}}^{1/2} \quad (6)$$

where  $m$  is the number of atoms within each group. The brackets  $\langle \dots \rangle_{\text{PDB}}$  indicate that the scattering factor is averaged over a set of backbone conformers and side chain rotamers taken from the Protein Data Bank.

**Explicit Solvent Representation.** In Fast-SAXS calculations, a layer of explicit water molecules is placed around the RNA surface to account for the excess electron density. For a TIP3P water molecule,<sup>34</sup> an effective scattering factor has already been derived,<sup>22</sup>

$$F_w^{\text{CG}}(q) = \left[ \sum_{i,j=1}^3 f_i(q) f_j(q) \frac{\sin(qr_{ij})}{qr_{ij}} \right]^{1/2} \quad (7)$$

where  $r_{ij}$  is the distance between water particles  $i$  and  $j$ . In practice, these water molecules are represented by dummy atoms centered at the oxygen positions, in bulk solvent density ( $\rho_s = 0.334 \text{ e}/\text{\AA}^3$  at 20 °C). These dummy atoms are taken from a pre-equilibrated TIP3P waterbox if their positions are within 3.5–6.5 Å away from the RNA coarse-grained particles. To effectively model the excess electron density in the hydration layer, the scattering factor of dummy water molecules is assigned a proper weight  $w$  by

$$F^{\text{CG}}(q) = w \times F_w^{\text{CG}}(q) \quad (8)$$

The optimal value of the weighting factor  $w$  for RNA is calibrated using the experimental SAXS data of tRNA-Val.<sup>35</sup> With this strategy, the total scattering of a given RNA structure is conveniently and accurately represented by its  $2N$  nucleic two-particle plus the surrounding  $M$  explicit dummy water molecules. Therefore, the total calculated scattering corresponding to the Fast-SAXS-RNA method is given by the following Debye formula,

$$I_{\text{cal}}(q) = \sum_{i,j=1}^{2N+M} F_i^{\text{CG}}(q) F_j^{\text{CG}}(q) \frac{\sin(qr_{ij})}{qr_{ij}} \quad (9)$$

where  $F_i^{\text{CG}}(q)$  are the effective coarse-grained scattering factors for both nucleic two-particle and dummy water molecules (eqs 6 and 8). The parameters  $N$  and  $M$  are the number of nucleotides and water molecules, respectively.

**RNA Conformational Sampling.** RNA conformational sampling is performed using the recently introduced MC-Fold and MC-Sym pipeline.<sup>16</sup> The conformational sampling via MC-Sym is enhanced by the use of NCMs,<sup>16,36,37</sup> which makes it possible to build structures for RNA sequences of up to 120 nucleotides. The advantage of MC-Sym is to build all-atom models with proper excluded volumes and electronic densities. Two RNA molecules with available experimental SAXS data, the *Escherichia coli* tRNA-VAL (tRNA)<sup>35</sup> and the P4–P6 domain of *Tetrahymena thermophila* group I intron (P4–P6),<sup>38</sup> are used. Interestingly, these two RNA molecules belong to two different structural classes: coplanar for the tRNA and bipolar for the P4–P6.<sup>39</sup> Additional tests include a RNA dimer that is suspected to form a dimeric complex in solution.<sup>40</sup>

For tRNA, two sets of 9999 decoys were made: tRNA\_high and tRNA\_low. The tRNA\_low set was produced using knowledge of the secondary structure, in-stem-only noncanonical base pairs (14–21, 26–44, 54–58), and local 3D motifs such as the T-loop<sup>41,42</sup> and the anticodon loop,<sup>43</sup> thus without any explicit long-range distance information. The single-stranded acceptor terminator, nucleotides 73–76, is assumed to be in the A-RNA conformation. Furthermore, the coaxial stacking of the four helical domains was used in the tRNA sampling. The tRNA\_low set is to identify the rough global L-shape of the tRNA fold. The tRNA\_high set was produced with the knowledge of tRNA\_low, in addition to long-range interactions (15–48, 18–55, 19–56) and base triples (8–14–21, 9–23–12, 22–46–13), as predicted from multiple-sequence analysis.<sup>44,45</sup> The purpose of this tRNA\_high set is to test the SAXS approach

to identify high-resolution three-dimensional models among closely related L-shaped structures.

For P4–P6, two sets of decoy models were made: the P4–P6\_low set (9999 decoys) and the P4–P6\_high set (8867 decoys). The P4–P6\_high set was generated by assembling stems/hairpins with the knowledge of a long-range contact between the tip of the P5b stem, a GNRA tetraloop, and its receptor in the P6 stem. This interaction was predicted by Murphy and Cech (MC)<sup>46</sup> and later confirmed in the crystal structure of the P4–P6 domain.<sup>47</sup> Here, the stems are sampled in the context of the crystal structure, with the distance constraint imposed by the MC contact. The group I domain is assembled by coaxially stacking the stems P6, P4, and P5 in axis 1, and the stems P5b on P5a in axis 2. The two axes are put relative to one another via an explicit sugar/sugar base pair relation (LW nomenclature<sup>48</sup>) between nucleotides 153 and 250. Final modeling steps link the two axes (Supporting Information Figure 11, magenta). The P4–P6\_low set was generated by arranging the stems into coaxially stacked axes, and by connecting the two axes via the 123–125/197–198 junctions. Because the distance constraint imposed by the MC contact was not imposed, the two axes are free to move in any directions to generate a broad range of model structures, such as the native U-shape, but also L-shapes and fully extended conformations.

For the RNA dimeric complex, we built various decoy sets representing different multimeric states and secondary structures. RNA hairpin monomers were first built using NCMs (Supporting Information Figure 12, set #1). The receptor module was taken from PDB file 1GID.<sup>38</sup> Then, the pairs of monomers were assembled into dimers in 3D space by using a sugar–sugar base pair edge interaction (LW nomenclature) between nucleotides A23 of monomer A and G8 of monomer B (Supporting Information Figure 12, set #2). The symmetric docking of nucleotide A23 of monomer B into nucleotide G8 of monomer A was achieved by using two distance constraints which simulate a sugar–sugar edge base pair interaction. This produces homodimers which are doubly docked. Another set was made in which the symmetric docking is not enforced. Only one tetraloop–receptor is modeled explicitly (Supporting Information Figure 12, set #4). This produces singly docked configurations, but doubly docked ones can occur fortuitously.

The last decoy set was made by using the MC-Fold computer program to generate alternative secondary structures from a tandem repeat of the monomer sequences X and Y, attached by an UUCG linker and to yield a 5′-X-UUCG-Y-3′ sequence. The linker is used to coax the sequence into a long extended hairpin. From MC-Fold, the UUCG linker is subtracted and further checked. Additional alternative secondary structures forming a 4-way junction were obtained (Supporting Information Figure 12; set #3). Interestingly, the extended duplex was not found in the top 20 MC-Fold solutions, as it features too many noncanonical base pairs to compete against more potent structures. However, this possibility still had to be explored, as RNA hairpins in solution can hybridize to one another to form extended double helices.<sup>49</sup> The alternative fold, which comprise a 4-way junction, is made up of two monomer units in which the basal stem would unzip up to just after the tetraloop receptor, hence unfolding it, then hybridize with complementary strands from another partially unzipped monomer (notice that the basal stems of monomers A and B in Supporting Information Figure 12 set #3 is made of two different colors or units). Similar to the tRNA fold, these four branches were organized into two main axes by coaxially stacking two stems, on their 5′ sides.



The 3D models of both axes were first built using MC-Sym, then assembled into complete structures in a second step.

All 3D models were refined using the Tinker molecular modeling package version 4.2.<sup>50</sup> A steepest-descent minimization algorithm, combined with the Amber99 force field<sup>51</sup> in a gas phase, was used until a gradient rmsd of 5.0 kcal/(mol Å) or a maximum of 200 steps was reached. Nucleobase atom positions were kept fixed using a spring constant of 1001 kcal/(mol Å). Sugar pucker in the C3' -endo anti configuration is enforced using a spring constant of 0.025 kcal/(mol °C) on the torsion angle (C3', C4', O4', C1') with mean 22.2 °C and standard deviation 4.3 4.3 °C.

**SAXS-Based  $\chi^2$  Scoring.** A ranking score,  $\chi^2$ , is used to measure the similarity between the experimental SAXS data and its theoretically computed Fast-SAXS-RNA profile,

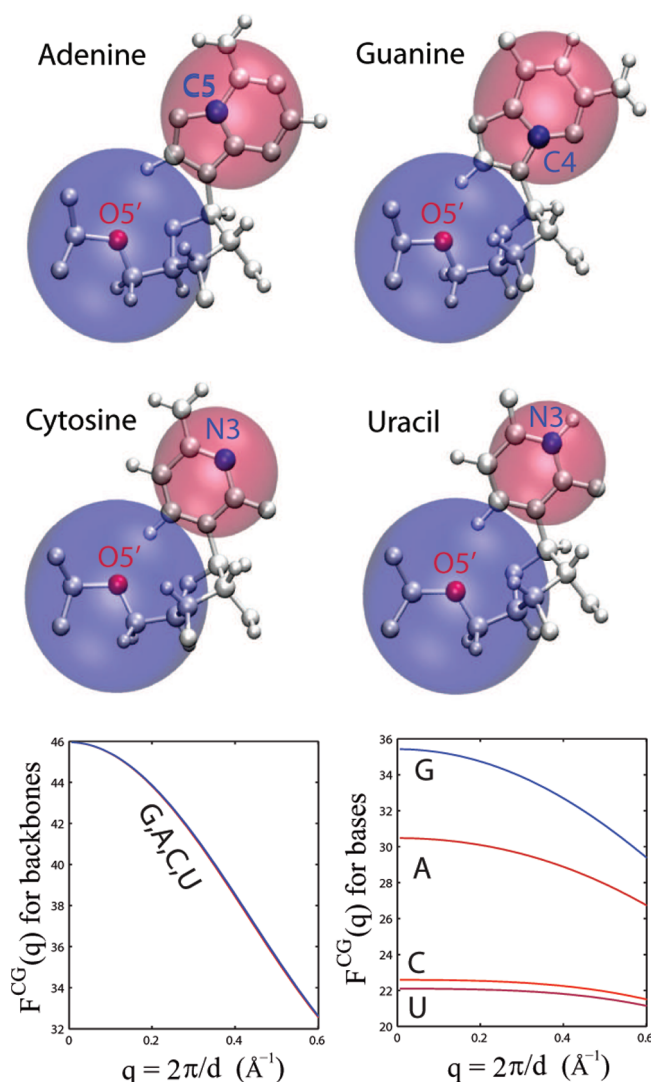
$$\chi^2 = \sum_{q_{\min}}^{q_{\max}} \frac{1}{\delta I_{\log}^2(q)} (\log I_{\text{cal}}(q) - \log I_{\text{exp}}(q) - \Delta)^2 \quad (10)$$

where  $q_{\min}$  and  $q_{\max}$  are the lower and upper limits of the  $q$ -range from the experimental scattering profile  $I_{\text{exp}}(q)$  and  $\delta I_{\log}(q)$  are the experimental uncertainties on  $\log I_{\text{exp}}(q)$ . The value of  $\Delta$  is the offset between the theoretical and experimental SAXS profiles ( $\log I_{\text{cal}}$  and  $\log I_{\text{exp}}$ ) at  $q = q_{\min}$ .

## Results

**Coarse-Grained Representation.** Recognizing the intrinsic low-resolution nature of SAXS data, the goal is to develop a coarse-grained approach for computing RNA scattering profiles. The coarse-graining development is built upon the theoretical basis of SAXS, as described in the Methods section. The nucleic units of RNA (adenine, guanine, cytosine, and uracil) are simplified into a two-particle model, each representing the scattering of the sugar–phosphate group of the backbone and the side chain group of the base. Figure 1 (top) shows the schematics of the two-particle representation for the four nucleotides. Subsequently, effective scattering structure factors for the resulting eight pseudoparticles are derived according to eq 6, based on their atomistically detailed representations. In practice, the effective scattering is obtained by averaging a set of high-resolution RNA atomistic coordinates, such as those deposited in the Protein Data Bank, to account for nucleotide conformational heterogeneity. The averaging converges when a set of 100 nucleotide coordinates are used. Therefore, these effective structure factors are derived by taking advantage of the low-resolution nature of SAXS. In the derived effective structure factors, an excluded volume effect for each atom is taken into account by assigning a Gaussian sphere (eq 3) such that it effectively accounts for the electron density contrast between RNA molecules and the bulk solvent.<sup>25</sup> Figure 1 (bottom) shows that the effective structure factors for sugar–phosphate backbone groups are nearly identical for the four RNA nucleotide units but are quite different for the side chain groups to the bases.

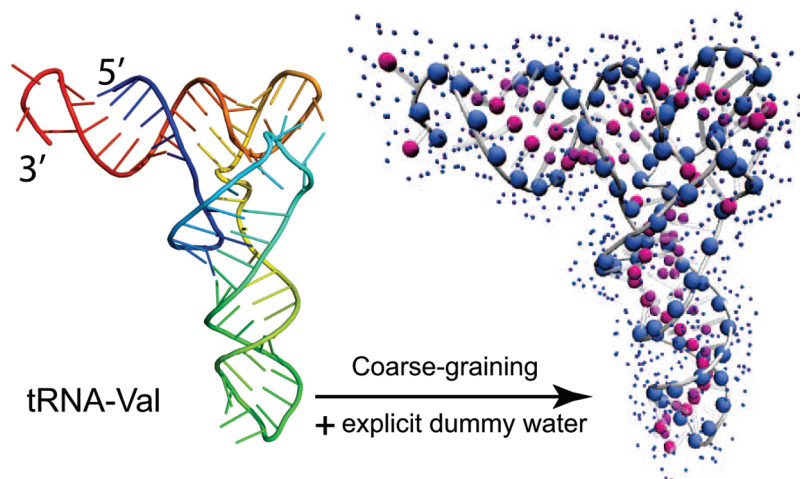
Equation 5 is used to determine the optimal placement of the pseudoparticles. The site placement is determined by the proximity to the group center of effective electron density, or center-of-scattering. It results in selecting the positions of O5' atoms for the backbone groups. The positions selected for the side chain groups are C5 for adenine, C4 for guanine, and N3 for both cytosine and uracil. Figure 1 depicts the placement of the pseudoparticles. The different positioning for the side chain groups between adenine and guanine is simply due to the



**Figure 1.** Two-particle model of RNA for SAXS computing. (Top) Schematic representation of the four RNA nucleic units: adenine (A), guanine (G), cytosine (C), and uracil (U). Each nucleotide is simplified into a two-particle model, where one particle accounts for the scattering of the phosphate and sugar groups, the backbone group, and the other for the side chain group. The location of the site is determined by the position of the atom closest to the center-of-scattering. Backbone positions are highlighted by red dots, and side chain ones in blue. The relative sizes of these particles are drawn approximately proportional to their scattering intensity at  $q = 0$ . (Bottom) Derived structure factors. On the left are shown the structure factors for the backbone groups and on the right are the structure factors for the side chain groups.

different location of the amine group of the base, close to C5 in adenine but close to C4 in guanine. It is shown below that the reduced representation is sufficient to reproduce the scattering curves of solved RNA structures. It is noteworthy that a reduced representation with one particle per residue was found sufficient for the scattering of proteins.<sup>10,22,52</sup>

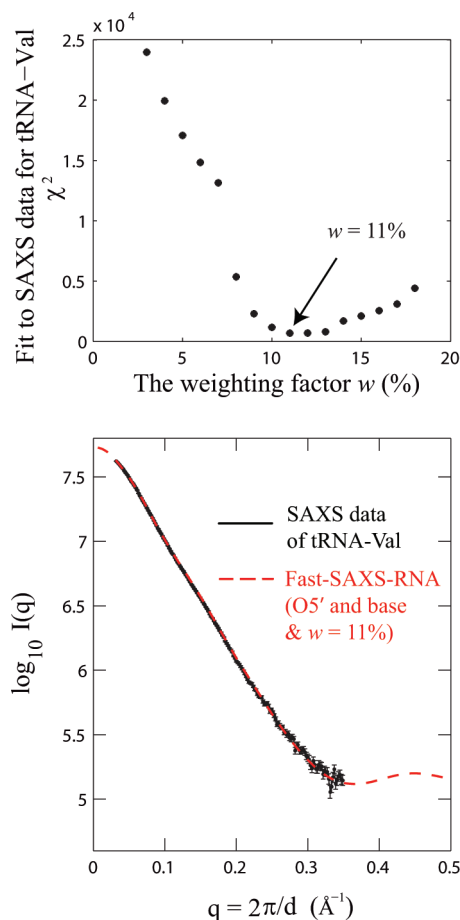
**Explicit Solvent Representation.** In addition to the RNA molecule itself, the hydration layer at the RNA surface also contributes to the total scattering.<sup>23,26,27,53,54</sup> The bound water molecules in the hydration layer collectively display a higher electron density than that of the bulk solvent. Here, the excess electron density, relative to the bulk, is modeled by soaking the RNA molecule in a layer of explicit water molecules. In the same fashion as for protein scattering calculations,<sup>22</sup> an explicit solvent representation is used for RNA scattering to account for the excess electron in the hydration layer. Both the



**Figure 2.** Coarse-graining for RNA scattering. (left) Representation of the 3D structure of valine tRNA (tRNA; PDB entry 2K4C), colored from blue at the 5' end to red at the 3' end. (Right) 3D structure coarse-grained into the two-particle representation, where the backbone particles are large blue spheres and the side chain in red. In addition, explicit water molecules are placed surrounding the RNA molecule and shown as small blue dots.

RNA and water are effectively coarse-grained to take advantage of the low-resolution nature of SAXS data. An effective structure factor for TIP3P water molecules can be determined by the use of eq 7 and has already been parametrized.<sup>22</sup> The water molecules are taken from a pre-equilibrated TIP3P waterbox at bulk solvent density ( $\rho_s = 0.334 \text{ e}/\text{\AA}^3$  at 20 °C). Water particles overlapping with the RNA molecule are deleted, and only those within 3.5–6.5 Å away from the RNA are kept. It results in a hydration layer of about 3 Å in thickness. The pre-equilibrated waterbox facilitates the solvation of the biomolecule, while preserving the explicit representation of the hydration layer. The water particles make a collective contribution to the total scattering, by representing excess hydration electron density as in eq 8. In this sense, they are not actual water molecules. Figure 2 shows the water particles representation for tRNA. At this point, a full solvated RNA molecular system is ready for SAXS calculations according to eq 9.

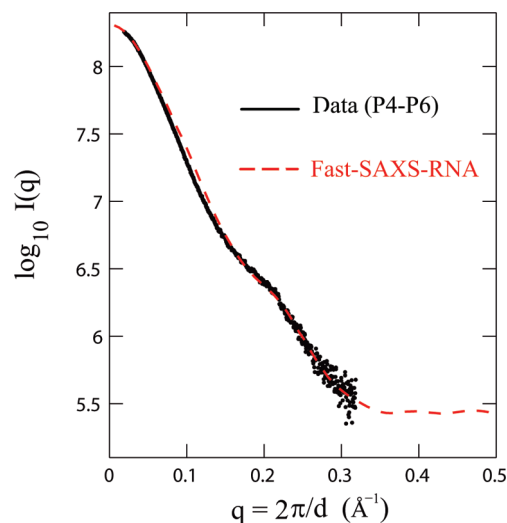
To calibrate the excess hydration electron density around RNA, an optimal value for the weighting factor  $w$  of the hydration contribution can be assigned using available experimental SAXS data. The recently solved tRNA-Val (PDB entry 2K4C) was used as a model system for this calibration.<sup>35</sup> Under the Fast-SAXS-RNA framework (eq 9), SAXS profiles of tRNA are calculated for various values of  $w$ . A  $\chi^2$  score is then calculated according to eq 10 to account for the difference between theoretical and experimental profiles. Finally, the scattering amplitude of water molecules is determined by best-fitting to SAXS data. Figure 3 shows the  $\chi^2$  difference between the computed and experimental SAXS profiles as a function of  $w$ . The result indicates that a value of  $w = 11\%$  is optimal to reproduce the experimental SAXS data of tRNA (Figure 3). It also suggests that the excess hydration electron density is higher for the tRNA ( $w = 11\%$ ) than for the lysozyme protein ( $w = 3\%$ ).<sup>22</sup> The higher excess hydration electron density might be caused by high occupancy of water molecules in both the major and minor grooves of the RNA helices.<sup>55–57</sup> Since each nucleotide has three pairing faces (Watson–Crick, Sugar and Hoogsteen, according to the Leontis–Westhof nomenclature), only one face is involved in base pairs (usually the Watson–Crick face) and the other two polar faces are left exposed to hydration (the Hoogsteen faces in major grooves, and the Sugar faces in minor grooves). In addition, each phosphate group has a net negative charge in solution, which attracts mono and divalent



**Figure 3.** Hydration layer contribution for RNA scattering. The contribution of the hydration layer is calibrated using the experimental SAXS data and the solution structure for a tRNA (tRNA; PDB entry 2K4C). (Top) Fit of the solution structure to the SAXS data as a function of the weighting factor  $w$ . A value of  $w = 11\%$  is optimal to fit the SAXS data. (Bottom) Comparison of the experimental (black) and computed (red) SAXS profiles.

cations and contributes to the excess electron density in the hydration layer.

We note that computing SAXS profiles is not sensitive to the detailed water placement in the hydration layer; instead, it depends only on the collective effect that the hydration layer



**Figure 4.** Model validation of the Fast-SAXS-RNA method. The two-particle model with the hydration layer is validated using the experimental SAXS data and the crystal structure for the P4–P6 fragment of group I intron (P4–P6; PDB entry 1GID). Comparison of the experimental (black) and Fast-SAXS-RNA computed (red) SAXS profiles. The  $\chi^2$  difference (eq 10) between these two curves is  $1.8 \times 10^{-3}$ .

makes as a whole to the total scattering. To illustrate this point, the total effective excess electron density is modeled by two factors: the weighting factor  $w$  and a hydration layer thickness  $d$ . As illustrated in the case of the lysozyme protein,<sup>22</sup> varying the weighting factor  $w$  and the thickness  $d$  can still capture the essential features of computed SAXS profile. In other words, the hydration contributions to the total scattering, as long as the net excess electron contribution is conserved, remain similar. In addition, we observe that the coarse-graining of RNA molecules using Fast-SAXS-RNA can well reproduce theoretical SAXS profiles using an all-atom-based approach (Supporting Information Figure 8). Taken together, an experimentally calibrated hydration contribution is obtained for a complete Fast-SAXS-RNA method. The calibration procedure is then tested on the SAXS data of the P4–P6 group I intron fragment by its crystal structure (Figure 4). This test provides the validation on the contribution of the hydration layer to the total scattering profile.

**RNA Conformational Sampling.** To test the ability of the Fast-SAXS-RNA method in identifying the native fold and evaluating the usefulness of SAXS data in RNA structure prediction, large decoy sets were generated for three different RNA molecules, namely tRNA, P4–P6, and a RNA dimer. The decoy sets were generated using the MC-Fold and MC-Sym pipeline,<sup>16</sup> which produces all-atom RNA 3D structures. The all-atom models allow for proper nucleotide volumes and electron densities. We capitalize on the generator's built-in knowledge of RNA architectural principles, essentially encoded in nucleotide cyclic motifs (NCMs) and in the ability of MC-Sym to explicitly model noncanonical and long-range base triples. For each RNA molecule, two sets of ensemble models were generated, in high and low resolutions to cover different regimes of the fold space (see Methods). For the RNA dimer, the method was further tested with the use of SAXS by making decoy models of alternate secondary structures.

**Application to tRNA.** For the tRNA conformational sampling, the set tRNA\_low has models with rmsd ranging from 5.6 to 25.5 Å (O5'-only), hence at a low resolution regime. The set tRNA\_high has models ranging from 3.1 to 12.7 Å, at a

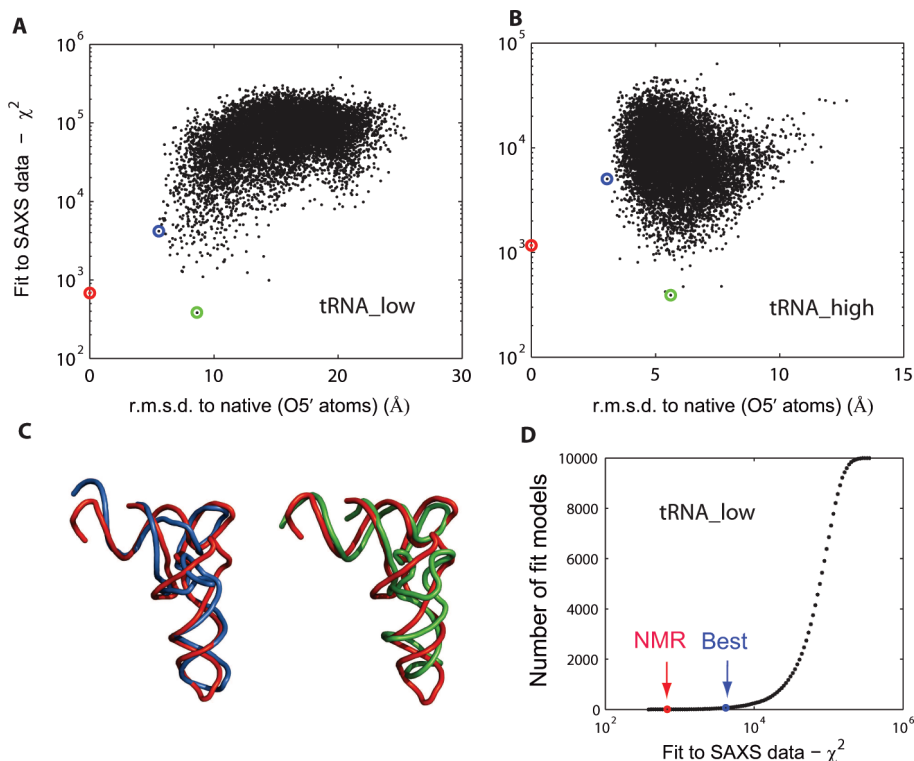
higher resolution regime. Supporting Information Figure 9 shows the conformational sampling for tRNA. The latter set includes only L-shaped models, while the former features several additional shapes such as H- and T-shapes. Surprisingly, decoys from the tRNA\_high set, with rmsd values of up to 12.7 Å, still look native-like. rmsd values of less than 3 Å seem to be quite difficult to obtain, even by making use of an all-atom model generator and explicit long-range and base triple pairings.

For each decoy within a set, a theoretical SAXS profile is first computed by Fast-SAXS-RNA, according to eq 9 (optimal weighting factor  $w = 11\%$ ). The theoretical profile is then compared to the experimental data via eq 10, by producing a  $\chi^2$  score. A lower  $\chi^2$  score indicates higher fidelity to the experimental SAXS profile. Decoys within a set are subsequently sorted by their  $\chi^2$  score. Figure 5A,B shows the plot of  $\chi^2$  against rmsd for both the tRNA\_low and tRNA\_high sets. It shows that the SAXS-based  $\chi^2$  scores serve as a good indicator for selecting top structural candidates from decoys in the tRNA\_low set. In fact, only 61 models (out of ten thousand) have a lower  $\chi^2$  score than the lowest-rmsd model in the tRNA\_low set. The result suggests that SAXS data are capable of filtering 3D models in RNA structure prediction in a low-resolution regime and significantly accelerate the search for the native fold. The ability of identifying the top 1% decoy tRNA\_val models by  $\chi^2$  scores demonstrates that one can pinpoint only this small subset as representative folds. These low-resolution models can further offer a starting point for a next-step structural refinement (which is beyond the scope of this work). As illustrated by Figure 5C, the best models according to a  $\chi^2$  criterion have similar overall shapes. The cumulative plot in Figure 5D further shows that the lowest-rmsd model is quite close to the solution structure, with a certain  $\chi^2$  cutoff. For the high-resolution set, a ranking based on a  $\chi^2$  score is not capable of capturing the finer structural differences between decoys (Figure 5B). Because of the low-resolution nature of SAXS data, most models in the tRNA\_high set have similar overall shapes and thus have undistinguished  $\chi^2$  scores. Nonetheless, the correlation between rmsd and  $\chi^2$  in the tRNA\_low set suggests that as a proof-of-principle study SAXS data can be used as input for a score function to locate these low-resolution models, especially when a native structure is not available.

We also note that the use of the full SAXS spectrum of tRNA\_val up to  $q \sim 0.3 \text{ Å}^{-1}$  can be advantageous. Traditional use of SAXS data is to derive the radius of gyration ( $R_g$ ) from the low- $q$  region using a Guinier fit.<sup>58</sup> Here, the  $\chi^2$  scores, which make use of the full SAXS spectrum, are strongly correlated with the rmsd values of decoy models in the tRNA\_low set (Figure 5A). In contrast, such a funnel-like correlation feature does not show up in the plot between  $R_g$  and rmsd (Supporting Information Figure 10), suggesting that  $R_g$  is not capable of discriminating native-like folds of tRNA\_val. Therefore, the SAXS-based  $\chi^2$  scores serve well as an indicator for correlating with native tRNA models by taking advantage of the full SAXS spectrum.

**Application to P4–P6.** Two sets of models were generated for the P4–P6, representing low and high resolution regimes, respectively. The set P4–P6\_low has models with rmsd ranging from 13.3 to 49.7 Å (O5' only). The set P4–P6\_high has models with rmsd from 6.9 to 14.2 Å. Even though the range of rmsd in the P4–P6\_high set is large compared to that for tRNA\_high, one has to keep in mind that P4–P6 has twice as many nucleotides than the tRNA, and that small modeling errors in connecting the stems produce large deviations. Nonetheless, the





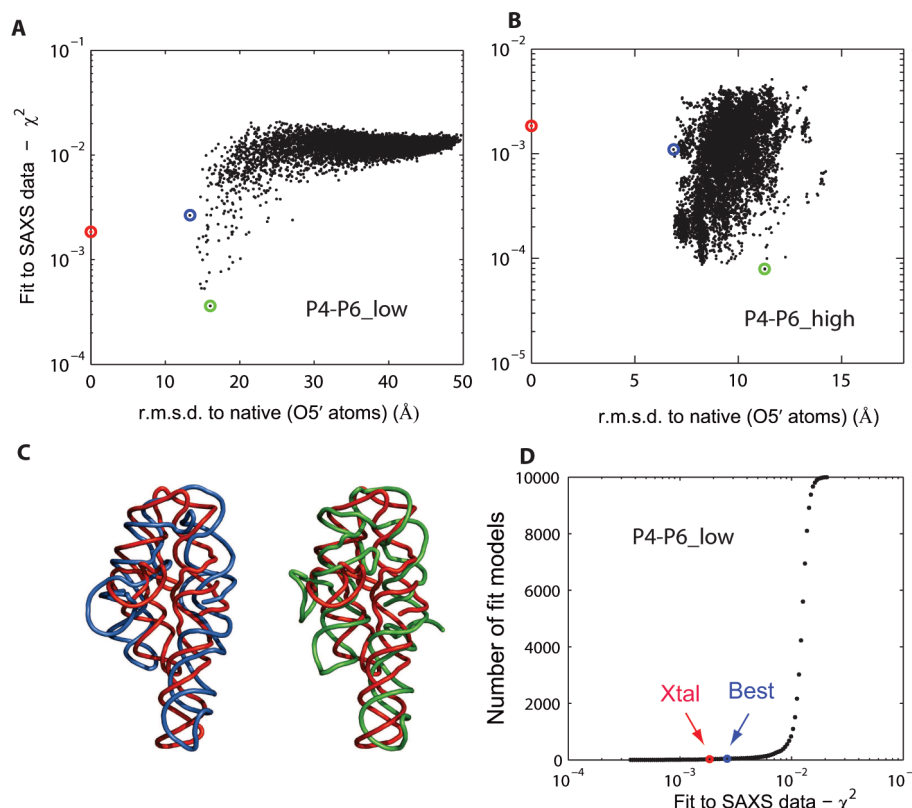
**Figure 5.** Application to tRNA. Two decoy sets are used: tRNA\_low and tRNA\_high. Each set contains 10 000 models (including the solution structure) covering different regimes of the conformational space. Plot of fit to the SAXS data  $\chi^2$  as a function of rmsd to the solution structure (PDB entry 2K4C), for tRNA\_low (A) and tRNA\_high (B), respectively. Lower  $\chi^2$  values indicate a better fit. rmsd is computed over all O5' atoms. The solution structure is indicated in red, the lowest-rmsd model in blue and the best SAXS-fit model in green. (C) Representative tRNA\_low decoy set models. Optimal superposition of the best lowest-rmsd model (blue; 5.6 Å) and the best SAXS-fit model (green; 8.6 Å) on the solution structure (red). (D) Cumulative model plot as a function of the fit to SAXS data  $\chi^2$ . Indicated are the positions of the solution structure (NMR) and the best lowest-rmsd model (Best). There are 61 models (out of 10,000) that display better  $\chi^2$  values than the best lowest-rmsd model. The  $q$ -ranges  $q_{\min} = 0.05 \text{ \AA}^{-1}$  to  $q_{\max} = 0.32 \text{ \AA}^{-1}$  were used for tRNA  $\chi^2$ -score calculations.

models compare favorably to those constructed by other methods (e.g., FARNAS<sup>59</sup> and NAST<sup>18</sup>). Supporting Information Figure 11 shows the conformational sampling for P4–P6.

The P4–P6\_low decoy set offers the most wide range of conformations (Figure 6A). Most decoy models are in an extended conformation, where two stems (P6–P4–P5 and P5a–P5c) form a long double-helical structure with a 180° angle (Figure 11a). Other decoys include the native-like U-shape, where a zero degree angle between these two stems (Figure 11b), as well as some decoys with a relative angle between 0 and 180°. Hence, it is striking that the use of SAXS data is capable to identify the U-shaped native fold of P4–P6 from the low-resolution decoys (Figure 6A). In contrast, the P4–P6\_high set, which is generated with the explicit use of the tetraloop-receptor long-range 3D contact, offers a different decoys set in Figure 6B, where the funnel-like correlation between  $\chi^2$  and rmsd is lost (compared to Figure 6A) although the slope of the distribution of  $\chi^2$  values against rmsd remains positive. Figure 6C shows the superimposition between the native U-shape structure (red), the best rmsd model (blue), and the best SAXS fit with the lowest  $\chi^2$  (green). Similar to tRNA-val results, these results demonstrate that the SAXS data P4–P6 is able to identify its native U-shape fold from the pool of decoys in the P4–P6\_low set. From the standpoint of structural modeling, one can use the resulting top models to further design new local structure probing experiments (such as cross-linking), which can provide insights into details about structural organizations of aligned sequences. Nonetheless, since the plot of  $\chi^2$  against rmsd shows a sign of positive correlation (Figure 6B), it appears that  $\chi^2$  can still keep track of the extended shape of P4–P6 in this

P4–P6\_high set. From the cumulative plot, only 40 models (out of ten thousand) have better  $\chi^2$  than the lowest-rmsd model (Figure 6D). As more quality SAXS data become available,<sup>60</sup> this result indicates that one can expect the use of SAXS data offers the top 1% best SAXS fit models in the search for native-like folds.

**Application to an RNA Dimer.** In this section, the ability of the method to identify the multimeric state and various secondary structures of an RNA hairpin is tested. The hairpin contains both a GNRA tetraloop and its receptor, allowing dimerization in solution (the tetraloop of one monomer would dock into the receptor in the other monomer, and vice versa). Supporting Information Figure 12 shows all the multimeric states and secondary structures tested, in a total of four different decoy sets. Figure 7A shows the fit to SAXS data for the four conformational states corresponding to the four sets. Conformational set #2 with its rmsd range close to the native structure has the best fit to the SAXS data, since it features dimers that are doubly docked. Interestingly, several models from this set display lower SAXS  $\chi^2$  scores than the solution structure (PDB entry 2JYH).<sup>40</sup> This suggests that conformational set #1 has the poorest  $\chi^2$  scores and the largest deviation from SAXS data among the four sets. It is well separated from the native homodimer set #2 in  $\chi^2$  values. This separation demonstrates that one should be able to tell from SAXS data that the dimeric form is more favorable than the monomeric form. Conformational set #4 includes models with only one tetraloop-receptor interaction formed and overlaps with set #2. Most of models in set #3 with the four-way junction have poorer SAXS ranking scores than those in set #4, suggesting that this state is less



**Figure 6.** Application to P4–P6. Two decoy sets are used: P4–P6\_low and P4–P6\_high. Each set contains approximately 10,000 and 9000 models (including the crystal structure). Plot of fit to the SAXS data  $\chi^2$  of P4–P6 as a function of rmsd to its crystal structure (PDB entry 1GID), for P4–P6\_low (A) and P4–P6\_high (B), respectively. The crystal structure is indicated in red, the lowest-rmsd model in blue, and the best SAXS-fit in green. (C) Representative P4–P6\_low decoy set models. Optimal superposition of the best lowest-rmsd model (blue; 13.3 Å) and the best SAXS fit model (green; 16.0 Å) on the crystal structure (red). (D) Cumulative model plot as a function of the fit to the SAXS data ( $\chi^2$ ). Indicated are the positions of the crystal structure (Xtal) and the best lowest-rmsd model (Best). There are 40 models that display better  $\chi^2$  values than the best rmsd model. The  $q$ -ranges  $q_{\min} = 0.02 \text{ Å}^{-1}$  to  $q_{\max} = 0.32 \text{ Å}^{-1}$  were used for P4–P6  $\chi^2$ -score calculations.

likely to be observed in solution. Figure 7B shows the superposition of the models with the best ranking score (green) and the lowest rmsd (blue) with the solution structure (red). Figure 7C shows the plot of  $\chi^2$  vs rmsd for set #4. The results suggest that the SAXS ranking helps to identify the models of the RNA dimer with lower rmsd (though keeping in mind that the doubly docked state can happen fortuitously in that decoy set).

Given the intrinsic low-resolution nature of SAXS, the information context is sufficient to distinguish the native state among many multimerization states for a small RNA hairpin. The symmetric homodimer state (2-docked) is identified as the best fit to the SAXS data against the asymmetric one (1-docked) and an alternative secondary structure, even though the decoy sets have similar radii of gyration  $R_g$  or maximum distances  $D_{\max}$ . Clearly, the usage of the full SAXS spectrum instead of single parameters like  $D_{\max}$  or  $R_g$  is required for a proper identification (Supporting Information Table 1).

## Concluding Discussion

A rapid coarse-grained method for calculating the SAXS profile from complex nucleic acid structures has been developed. The method has been termed Fast-SAXS-RNA. The computer program Fast-SAXS-RNA will be released under the GNU General Public License.

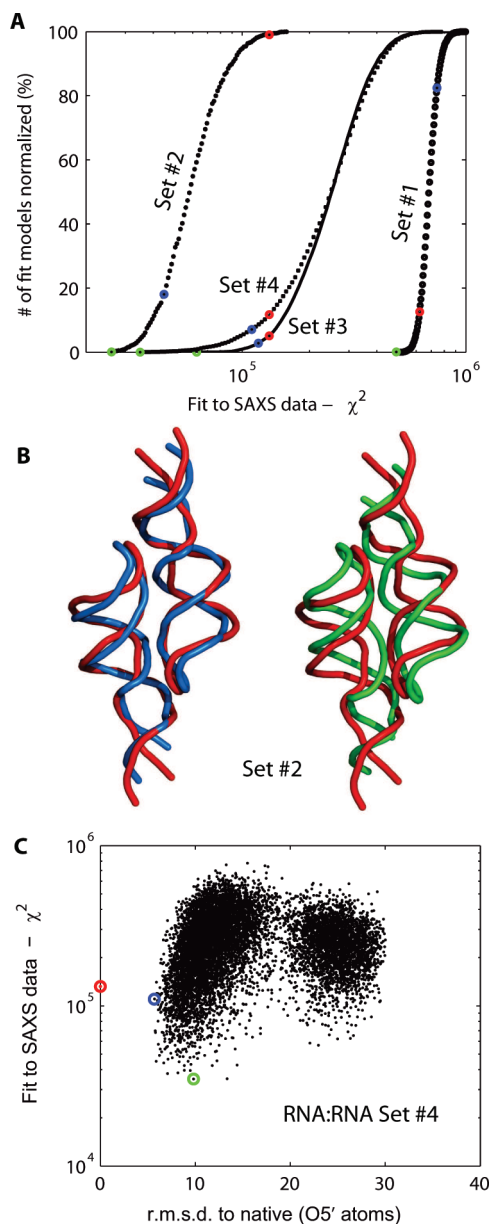
The method is built on a reduced two-particle representation for each nucleotide. While it is possible to treat proteins based on a reduced representation with a single particle per residue,<sup>22</sup> the present results indicate that a two-particle representation to

achieve the desirable accuracy in the calculation of SAXS patterns in the case of tRNA structures. A single-particle-based SAXS calculation is unable to reproduce accurately the experimental profile, particularly beyond  $q = 0.1 \text{ Å}^{-1}$  (Supporting Information Figure 8). Furthermore, the excess electron density at the RNA surface is taken into account in the Fast-SAXS-RNA calculation by explicit construction of a hydration layer. The amplitude of water scattering was calibrated from tRNA-Val data, and further tested on P4–P6 domain with good agreement with SAXS data.

As expected, coarse-graining RNA molecules and their surrounding solvent greatly accelerate the computation of SAXS profiles. For example, within the framework of the Debye formalism, computing the scattering of a protein without surrounding water molecules is faster by a factor of  $N^2$  (where  $N = 16$  is the average number of atoms per group). It should be noted that the gain in speed is less pronounced when explicit dummy water molecules are included in the calculation, although a direct comparison with atomistically detailed solvent has not been attempted.

The utility of Fast-SAXS-RNA becomes obvious when it is combined with RNA 3D structure generators. Here, the advantage of using MC-Sym was to build all-atom models with proper excluded volumes and electronic densities. Besides, the MC-Pipeline offers a series of utility tools for managing RNA 3D structure decoys, including a general filtering process that allows one to upload various types of structural data. Because of the significant reduction in computational cost for testing a model, it is possible to rank, order, and filter large ensemble of 3D





**Figure 7.** Application to an RNA dimer. (A) Cumulative model plot as a function of the fit to SAXS data  $\chi^2$ . A total of four decoy sets, from #1 to #4, are compared. These decoy sets address various multimerization states and secondary structures (see Methods). Since the number of models in each decoy set is different, the cumulative plot has been normalized to the fraction of total number of decoys in the set. For each set, the solution structure is highlighted in red (PDB entry 2JYH), the best lowest-rmsd model in blue, and the best SAXS-fit in green. For decoy set #1, the solution structure is considered to be the monomer unit. (B) Representative models from decoy set #2. The best SAXS-fit model (green; 7.5 Å) and best rmsd model (blue; 4.7 Å) optimally superimposed on the solution structure (red). (C) Plot of fit to SAXS data  $\chi^2$  against rmsd, for models in decoy set #4. The solution structure is highlighted in red, blue for the best rmsd model, and green for the best SAXS-fit model.

models extremely efficiently. For tRNA, P4–P6, and the RNA dimeric complex, usage of SAXS data as a constraint succeeds to filter out most non-native folds and identify well-fit models at a low-resolution level. Hence, the incorporation of SAXS data in modeling offers an alternative mean to dramatically narrow down the conformational search. It is our hope that this strategy will enable the rapid characterization and structural refinement of RNA models by exploiting increasingly available SAXS data.

**Acknowledgment.** We thank Drs. Alexander Grishaev and Ad Bax for providing the SAXS data of tRNA-Val, Drs. Jan Lipfert and Sebastian Doniach for providing the SAXS data of P4–P6, and Drs. Xiaobing Zuo, Samuel Butcher, and Yu-Xing Wang for providing the SAXS data of RNA dimer. This work was supported by the National Institute of Health (S.Y. and B.R.) via a NCI grant (No. CA093577) and by the Canadian Institutes of Health Research (CIHR) and the Natural Sciences and Engineering Research Council (NSERC) of Canada (FM). M.P. holds Ph.D. scholarships from the NSERC and the Fonds Québécois de la Recherche sur la Nature et les Technologies. F.M. is a member of the Centre Robert-Cedergren of the Université de Montréal.

**Supporting Information Available:** List of parameters for the RNA dimer decoy sets. Figures showing SAXS scattering profiles, conformations of tRNA, P4–P6, and the RNA dimer, and  $R_g$  vs rmsd. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## References and Notes

- (1) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. *Q. Rev. Biophys.* **2007**, *40*, 191–285.
- (2) Doniach, S. *Chem. Rev.* **2001**, *101*, 1763–1778.
- (3) Koch, M. H. J.; Vachette, P.; Svergun, D. I. *Q. Rev. Biophys.* **2003**, *36*, 147–227.
- (4) Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole, F. L.; Tsutakawa, S. E.; Jr, F. E. J.; Classen, S.; Frankel, K. A.; Hopkins, R. C.; jae Yang, S.; Scott, J. W.; Dillard, B. D.; Adams, M. W. W.; Tainer, J. A. *Nat. Meth.* **2009**, *6*, 606–612.
- (5) Tsuruta, H.; Irving, T. *Curr. Opin. Struct. Biol.* **2008**, *18*, 601–608.
- (6) Ibel, K.; Stuhmann, H. B. *J. Mol. Biol.* **1975**, *93*, 255–265.
- (7) Jacques, D. A.; Trewella, J. *Protein Sci.* **2010**, *19*, 642–657.
- (8) Lipfert, J.; Chu, V. B.; Bai, Y.; Herschlag, D.; Doniach, S. *J. Appl. Crystallogr.* **2007**, *40*, s229–s234.
- (9) Lipfert, J.; Doniach, S. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 307–327.
- (10) Svergun, D. I.; Petoukhov, M. V.; Koch, M. H. *Biophys. J.* **2001**, *80*, 2946–2953.
- (11) Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* **2007**, *129*, 5656–5664.
- (12) Zheng, W.; Doniach, S. *J. Mol. Biol.* **2002**, *316*, 173–187.
- (13) Ali, M.; Lipfert, J.; Seifert, S.; Herschlag, D.; Doniach, S. *J. Mol. Biol.* **2010**, *396*, 153–165.
- (14) Das, R.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14664–14669.
- (15) Ding, F.; Sharma, S.; Chalasani, P.; Demidov, V. V.; Broude, N. E.; Dokholyan, N. V. *RNA* **2008**, *14*, 1164–1173.
- (16) Parisien, M.; Major, F. *Nature* **2008**, *452*, 51–55.
- (17) Martinecz, H. M.; Jacob, V.; Maizel, J.; Shapiro, B. A. *J. Biomol. Struct. Dyn.* **2008**, *25*, 669–84.
- (18) Jonikas, M. A.; Radmer, R. J.; Laederach, A.; Das, R.; Pearlman, S.; Herschlag, D.; Altman, R. B. *RNA* **2009**, *15*, 189–199.
- (19) Frellsen, J.; Moltke, I.; Thim, M.; Mardia, K. V.; Ferkinghoff-Borg, J.; Hamelryck, T. *PLoS Comput. Biol.* **2009**, *5*, e1000406.
- (20) Walther, D.; Cohen, F. E.; Doniach, S. *J. Appl. Crystallogr.* **2000**, *33*, 350–363.
- (21) Svergun, D.; Barberato, C.; Koch, M. H. *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
- (22) Yang, S.; Park, S.; Makowski, L.; Roux, B. *Biophys. J.* **2009**, *96*, 4449–4463.
- (23) Bragg, L.; Perutz, M. F. *Proc. R. Soc. A* **1952**, *213*, 425–435.
- (24) Cromer, D. T.; Mann, J. B. *Acta Crystallogr. A* **1968**, *24*, 0567–7394.
- (25) Fraser, R. D. B.; MacRae, T. P.; Suzuki, E. *J. Appl. Crystallogr.* **1978**, *11*, 693–694.
- (26) Svergun, D. I.; Richard, S.; Koch, M. H. J.; Sayers, Z.; Kuprin, S.; Zaccari, G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2267–2272.
- (27) Merzel, F.; Smith, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5378–5383.
- (28) Koizumi, M.; Hirai, H.; Onai, T.; Inoue, K.; Hirai, M. *J. Appl. Crystallogr.* **2007**, *40*, s175–s178.
- (29) Harker, D. *Acta Crystallogr.* **1953**, *6*, 731–736.
- (30) Guo, D. Y.; Smith, G. D.; Griffin, J. F.; Langs, D. A. *Acta Crystallogr. A* **1995**, *51*, 945–947.

- (31) Guo, D. Y.; Blessing, R. H.; Langa, D. A.; Smith, G. D. *Acta Crystallogr. D* **1999**, *55*, 230–237.
- (32) Grishaev, A.; Wu, J.; Trehwella, J.; Bax, A. *J. Am. Chem. Soc.* **2005**, *127*, 16621–16628.
- (33) Debye, P. *Ann. Phys. (Leipzig)* **1915**, *46*, 809–823.
- (34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (35) Grishaev, A.; Ying, J.; Canny, M. D.; Pardi, A.; Bax, A. *J. Biomol. NMR* **2008**, *42*, 99–109.
- (36) Major, F.; Turcotte, M.; Gautheret, D.; Lapalme, G.; Fillion, E.; Cedergren, R. *Science* **1991**, *253*, 1255–1260.
- (37) Major, F. *Comput. Sci. Eng.* **2003**, *5*, 44–53.
- (38) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678–1685.
- (39) Laederach, A.; Chan, J. M.; Schwartzman, A.; Willgoos, E.; Altman, R. B. *RNA* **2007**, *13*, 643–650.
- (40) Zuo, X.; Wang, J.; Foster, T. R.; Schwieters, C. D.; Tiede, D. M.; Butcher, S. E.; Wang, Y.-X. *J. Am. Chem. Soc.* **2008**, *130*, 3292–3293.
- (41) Lee, J. C.; Cannone, J. J.; Gutell, R. R. *J. Mol. Biol.* **2003**, *325*, 65–83.
- (42) Krasilnikov, A. S.; Mondragon, A. *RNA* **2003**, *9*, 640–643.
- (43) Fuller, W.; Hodgson, A. *Nature* **1967**, *215*, 817–821.
- (44) Levitt, M. *Nature* **1969**, *224*, 759–763.
- (45) Klingler, T. M.; Brutlag, D. L. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1993**, 225–233.
- (46) Murphy, F. L.; Cech, T. R. *Biochemistry* **1993**, *32*, 5291–5300.
- (47) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Szewczak, A. A.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1696–1699.
- (48) Leontis, N. B.; Westhof, E. *RNA* **2001**, *7*, 499–512.
- (49) Bouchard, P.; Lacroix-Labonte, J.; Desjardins, G.; Lampron, P.; Lisi, V.; Lemieux, S.; Major, F.; Legault, P. *RNA* **2008**, *14*, 736–748.
- (50) Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016–1024.
- (51) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (52) Chacón, P.; Morán, F.; Díaz, J.; Pantos, E.; Andreu, J. *Biophys. J.* **1998**, *74*, 2760–2775.
- (53) Perkins, S. J. *Biophys. Chem.* **2001**, *93*, 129–139.
- (54) Hubbard, S.; Hodgson, K.; Doniach, S. *J. Biol. Chem.* **1988**, *263*, 4151–4158.
- (55) Auffinger, P.; Hashem, Y. *Curr. Opin. Struct. Biol.* **2007**, *17*, 325–333.
- (56) Chu, V. B.; Bai, Y.; Lipfert, J.; Herschlag, D.; Doniach, S. *Curr. Opin. Chem. Biol.* **2008**, *12*, 619–625.
- (57) Chen, S.-J. *Annu. Rev. Biophys.* **2008**, *37*, 197–214.
- (58) Guinier, A.; Fournet, G. *Small-angle scattering of x-rays*; Wiley: New York, 1955.
- (59) Das, R.; Kudaravalli, M.; Jonikas, M.; Laederach, A.; Fong, R.; Schwans, J. P.; Baker, D.; Piccirilli, J. A.; Altman, R. B.; Herschlag, D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4144–4149.
- (60) Rambo, R. P.; Tainer, J. A. *RNA* **2010**, *16*, 638–646.

JP1057308