# 8aSupporting Information for

## Determination of the conformational ensemble of the TAR RNA by x-ray scattering interferometry

**Xuesong Shi, Peter Walker, Pehr B. Harbury, and Daniel Herschlag**

### Supplementary Note S1. Step-by-step protocol for the post synthesis cleavage and de-protection of synthetic RNA containing amino-modifier (N2-Amino-Modifier C6 dG, Glen-research)

1.) Following synthesis of the desired RNA oligonucleotide, dry the resin briefly and transfer to a clean, dry 2 ml Eppendorf tube with a screw cap.
2.) Pipette 1 ml of Ammonia:Propylamine (1:1) into the tube. Allow the mixture to sit at room temperature overnight.
3.) Transfer the supernatant to a new Eppendorf tube and proceed to dry the sample in a vacuum centrifuge with no heat.
4.) To the dried sample add 115 uL of fresh DMSO and incubate at 65 °C for 20 min to dissolve the sample.
5.) Add 60 µL of concentrated triethylamine (Sigma #T0886) followed by 75 µL of triethylamine trihydrofluoride (Sigma #344648) and allow the sample to incubate for 2.5 hours at 65 °C.
6.) Following incubation, allow the sample to cool for 30 min and then add 1.7 ml of Quenching Buffer (Glen Research #60-4120-82).
7.) Purify the sample using an RNA purification column (Glen Research #60-6100).

### Supplementary Note S2. Generation of a solution model for RNA helices and internally labeled Au nanocrystals

To generate a solution model of the RNA helix and a model of internally labeled Au nanocrystals, X-ray interferometry (XSI) measurements were carried out on RNA duplexes with Au nanocrystal probes placed at 20 different pairs of positions (Supplementary Figure S3) at two salt conditions. We use the resulting distance distributions (Supplementary Figure S4) to obtain a solution model of RNA helices and of the internally labeled Au nanocrystals under the two salt conditions. These models include solution properties of RNA helices –i.e., the average twist, roll and tilt per base step, RNA elastic properties for bending and twisting persistence lengths, and positions of internally labeled Au nanocrystals with respect to the helices (Supplementary Figure S5 and Supplementary Table S1, also see Supplementary notes S3 and S4 below for a detailed description of the data analysis procedure).

As the addition of 10 mM $MgCl_2$ gives only small differences in the properties of the RNA/Au nanocrystal system (Supplementary Table S1), we adopted a previously used simple model(1) to account for this salt dependence. This model assumes that each of the nine parameters defining the RNA/Au nanocrystal system is correlated with the relative electrostatic screening strength of the solution ($x$) with a single linear correlation, as in equation (1).

$$Para_{i=1-9}(x) = Para_i(x_{low}) + (x - x_{low}) * \frac{Para_i(x_{high}) - Para_i(x_{low})}{(x_{high} - x_{low})} \quad (1)$$

In equation (1), x is the relative salt strength of the solution; $x_{low}$ and $x_{high}$ are relative salt strength of the two salt conditions with experimentally determined parameter sets available. We previously used RNA helices with end-labeled gold nanocrystals to generate a simple formula to estimate the relative salt strength of a solution (Eq. 2)(1). The solution properties of an RNA/Au system under different salt conditions can be predicted from equation (1) and the parameters in Supplementary Table S1.

$$x = [divalent] + [monovalent]/46 \quad (2)$$

To test the above salt dependence model, we carried out additional XSI measurements using four different gold pairs (Supplementary Figure S3, duplex 2) under a third salt condition (10 mM NaCl, 30 mM Tris-HCl, pH 7.4, 10 mM sodium ascorbate; the low salt condition in the main-text). The mean Au-Au distance values predicted using the above salt dependence model, using equation (2), agreed well with the observed values with an average deviation of ~0.4 Å. A better fit can be achieved when the value of x is allowed to deviate slightly from [divalent] + [monovalent]/46, and this best-fit model with x = 27 instead of x = 46 was used herein to predict the corresponding TAR-XSI under this salt condition (10 mM NaCl, 30 mM Tris-HCl, pH 7.4, 10 mM sodium ascorbate; the low salt condition in the main-text).

## Supplementary Note S3. Generation of the gold probe position cloud in the coordinate of its attachment base.

To predict the Au-Au distance for any given RNA conformation requires knowing the location of the Au probe in the coordinate of its attachment base. The average position of the Au probe can be defined by three Au probe parameters, D, $\theta_0$ and $axial_0$, as previously described (2,3). Briefly, in the coordinate of the base-pair (Figure 1b), D is the radial displacement of the probe from the z-axis, $\theta_0$ is the azimuthal angular rotation of the probe away from the short axis of the base pair, and $axial_0$ is the displacement along the z-axis of the probe from the base pair. As the probe linker is not completely rigid, the Au nanocrystal is expected to move and the coordinates of the Au probes are better described as a Au cloud around an average position. This Au position cloud was modeled as a sphere around the average position in our previous studies on DNA(2,4). In this work, we took the atomic detail of the probe linker and RNA helix geometry into consideration and generated a more realistic model of the Au position cloud (Supplementary Figure S2).

Specifically, a starting Au position cloud was generated by a pseudo all-atom Monte Carlo sampling of linker conformations. With an RNA helix fixed in the coordinate of the probe attachment nucleotide, we built a linker starting from the fixed attachment atom by random sampling a set of torsional angles for rotatable linker bonds. The bond length and bond angles were fixed to consensus values and torsional rigidity was ignored to model a completely flexible linker; the Au nanocrystal was modeled as a pseudo atom with a S-$Au_{center}$ bond length of 7 Å. Thus, a set of linker torsional angles determines the coordinate of the Au probe center. In the simulation, we began with a randomly generated linker conformation and

sampled the Au position space by randomly changing the set of torsional angles in small steps. A move was accepted if there was no steric clash between the Au center and RNA helix atoms, between Au center and linker atoms, or between linker atoms and RNA helix atoms, with clashing distances set to 11, 8 and 2 Å, respectively. A starting Au nanocrystal position cloud of 10,000 conformations was generated using the above procedure.

To model the actual Au nanocrystal cloud with experimental data, we introduced four parameters, D, $\theta_0$, $axial_0$, and $\varepsilon_{Au}$. The first three parameters, as noted above, define the average position of the Au nanocrystal cloud. The fourth parameter, $\varepsilon_{Au}$, adjusts the size of the Au cloud. Specifically, with the probability for each Au coordinate in the starting space being $p_i$ = 1/10,000, we introduced a pseudo energy potential that favors Au coordinates that are closer to the center of the Au nanocrystal cloud, reweighting the Au cloud as $p_i = (1/10{,}000)*\exp(-d*\varepsilon_{Au})$, where d is the distance from the cloud center and $\varepsilon_{Au}$ adjusts the strength of this energy potential and consequently the size of the cloud. $\varepsilon_{Au}$ is set to be non-negative; when $\varepsilon_{Au}$ equals to zero, the Au nanocrystal cloud is unchanged from the starting ensemble, and larger values of $\varepsilon_{Au}$ result in a smaller or more compact Au nanocrystal cloud and reflect a more rigid linker. The entire Au nanocrystal cloud was then translated to match the cloud center with the position defined by D, $\theta_0$ and $axial_0$. The resulting Au nanocrystal cloud was then used to calculate the Au-Au distance distribution for any given RNA conformation as described below.

The Au nanocrystal cloud is relatively insensitive to salt. Under two experimental conditions with and without 10 mM $MgCl_2$, $\varepsilon_{Au}$ values were similar and the average Au probe position differed by only ~0.6 Å (Supplementary Table S1).

## Supplementary Note S4. Fitting the gold probe and RNA helix parameters

***RNA model.*** We modeled the RNA helices as continuous linear elastic rods (see ref. (2)), where fluctuations of base step parameters, [twist, tilt, roll], are assumed to be governed by elastic potentials. The experimental data were used to parameterize five RNA parameters, including three helical parameters: the average [twist, tilt, roll] per base step, and two elastic parameters: the bending (B) and twisting (C) persistence length. The other three helical parameters, [shift, slide and rise], were set to literature values derived from crystal structure database (ref. (5), Supplementary Table S1) as their values are roughly constant across different literature models of RNA (Supplementary Table S2) and as initial modeling attempts suggested that the data may be underpowered to determine all six helical parameters simultaneously.

***Prediction of the Au-Au center-to-center distance distribution for a pair of Au nanocrystal probes on an RNA helix.*** The procedure for predicting the Au-Au center-to-center distance distribution for a pair of Au nanocrystal probes on an RNA helix was the same as described previously for DNA(2), except that an RNA model was used. Briefly, a virtual RNA chain of $10^6$ base pairs was constructed using a previously described procedure(2). Each base pair was built upon the previous one using a set of [twist, tilt, roll] randomly sampled from a Gaussian set of distributions defined by the RNA model above. For each base pair, the Au nanocrystal position was randomly picked from the Au nanocrystal cloud generated as described above. This chain was then used to compute the distribution of probe-probe distance for

base-step separations between -35 and 35. The mean Au-Au distance and Au-Au distance variance for the predicted Au-Au center-to-center distance distribution was compared with the experimental values to obtain the optimal values for Au nanocrystal and RNA parameters (see below).

***Fitting the Au nanocrystal parameters and RNA helix parameters.*** There are nine parameters to be optimized, consisting of four gold parameters [D, $\theta_0$, axial$_0$, $\varepsilon_{Au}$], defined in Supplementary Note S3, and five RNA parameters [twist, tilt, roll, B, C]. Twist, tilt and roll are the average twisting and bending angles, respectively, of a base step, as defined in ref. (6); B and C are the bending and twisting persistence length, respectively. These nine parameters are divided into three groups. The first group is sensitive to the average Au-Au distance and includes the three helical parameters, average [twist, tilt, roll] per base step, and the three probe position parameters, [D, $\theta_0$, axial$_0$]. The second group is sensitive to the Au-Au distance variance and includes the two RNA elasticity parameters, the RNA bending (B) and twisting (C) persistence length. The third group is $\varepsilon_{Au}$, which is most sensitive to the Au-Au distance variance when the number of base steps are small (e.g., $N < 7$), where the variance from probe dynamics is significant compared to helical dynamic variance.

For each choice of parameter values, a model distribution was generated as described above and a $\chi^2$ statistic quantifying the goodness-of-fit between the mean Au-Au distance and Au-Au distance variance of the model distributions and the observed distributions was then computed. The total $\chi^2$ was calculated as $\chi^2$ (mean distance) + $\chi^2$ (distance variance). The optimal parameters were defined as those that minimized the $\chi^2$ statistic identified by a Monte Carlo-type stepwise numerical search using MATLAB. In the stepwise numerical search, starting from a set of randomly assigned initial parameter values, the parameter values were randomly increased or decreased in small increments. A step was always accepted if $\chi^2 < \chi^2$ (last step); if $\chi^2 > \chi^2$ (last step), the step (change) was accepted with a probability of $e^{\chi^2 - \chi^2 \text{ (last step)}}$. To accelerate the search for the global minimum, the search was carried in stages. In each stage, only one of the three parameter groups (see above) was allowed to vary, and for group 3, only data with base step number of $-7 < N < 7$ were used (see above). In each stage, the first 300 moves were used for equilibration and the last 1200 moves were used to calculate the parameters to be used as input for the next stage. The search stages were repeated as stage (group3) → stage (group2) → stage (group1) → stage (group3) and so on. The search was carried out for 7 rounds and 21 stages and converged after a few rounds (Supplementary Figure S6).

## Supplementary Note S5. The Au nanocrystal label does not substantially perturb RNA structure.

To independently test possible effects from the Au labels, we compared circular dichroism (CD) spectra and melting temperatures of RNA duplexes with and without Au modification. The CD spectra showed no observable difference (Supplementary Figure S11), and single Au labels had only minor effects on duplex thermostability (Supplementary Table S3); the decrease in $T_m$ by 2-3 °C upon single Au labeling could arise from the 2-deoxyribose at the position of Au labeling and from a desolvation penalty associated with the reduction in solvent accessibility of the thiol-glucose shell of the Au nanocrystal upon duplex formation, which was previously suggested to account for a 1-2 °C Tm decrease upon Au labeling of DNA(2). The effect of the double Au labels on duplex thermostability is additive within error (Supplementary Table S3).

The demonstrated ability of partial datasets at large Au-Au separation to accurately predict Au-Au distances at smaller separation (Supplementary Figure S10a) and of partial dataset from one RNA sequence to accurately predict the mean Au-Au distance of a different RNA sequence and length (Supplementary Figure S10b) indirectly but strongly suggests that the RNA structure is not perturbed by the Au labels.

## Supplementary Note S6. Comparison of the abilities of the XSI and NMR-RDC methods to determine ensembles

To learn more about the relative abilities of XSI-RNA and NMR-RDC measurements to reveal conformational ensembles, we carried out a comparison with a synthetic data set. Ten distinct target ensembles and corresponding synthetic XSI and RDC data were generated as described in Methods (Supplementary Figure S9a and S9b). The synthetic XSI and RDC data were then separately used to recover the target ensemble (Methods). The recovered ensembles using synthetic XSI data closely matched their corresponding target ensembles (Supplementary Figure S9a and S9b). We further measured the similarity between the target ensemble and the XSI- and RDC-recovered ensembles using $\langle\Omega\rangle$ as described in ref. (7). $\langle\Omega\rangle$ is a measure of the differences in Euler angle values between two ensemble distributions, with a smaller $\langle\Omega\rangle$ indicating greater similarity between ensembles(7). We found the XSI method to be as effective as NMR-RDC in recovering these target ensembles, as judged by the smaller average $\langle\Omega\rangle$ values (Supplementary Figure S20).

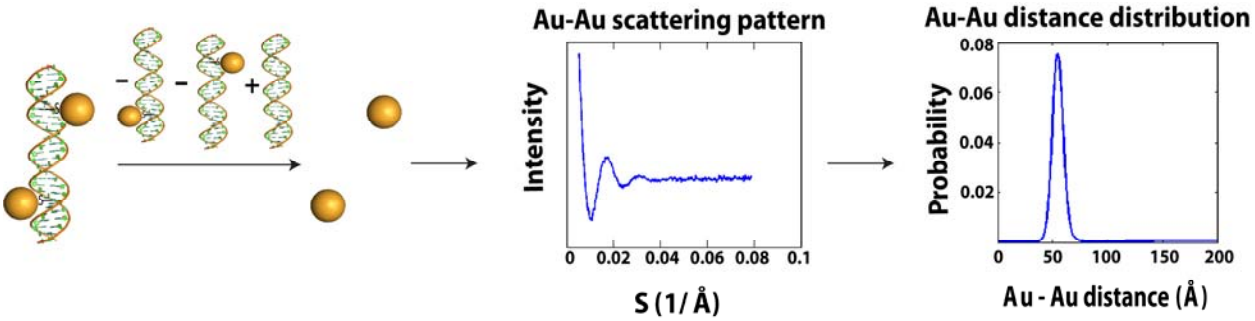


**Figure S1. General procedure for obtaining a probe-probe distance distributions from x-ray interferometry (reproduced from Figure 1 of ref. (2)).** (Left) DNA duplex labeled with a gold nanocrystal probe on each of the two strands of DNA. After subtracting the scattering signals involving the helix, as indicated by the species above the arrow, the pattern of scattering interference between the two gold probes is obtained (middle) and Fourier transformed to provide the probability distribution for the center-to-center distance between the probes (right). Data shown are for two gold probes separated by 15 base steps within a 26 base-pair duplex (2).

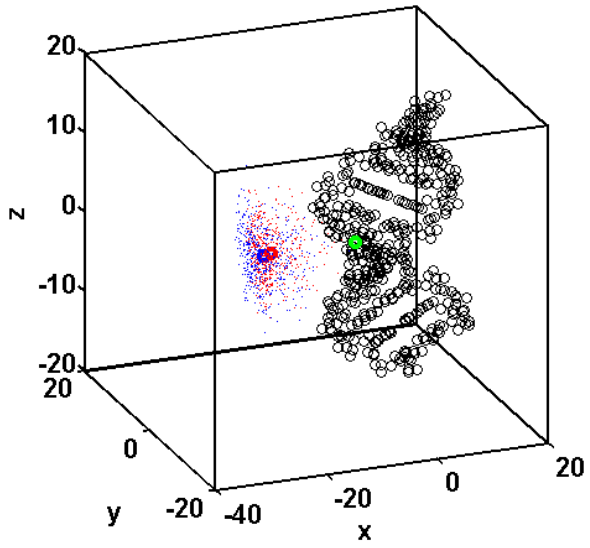


**Figure S2. The position of the internally labeled Au nanocrystals with respective to the RNA helix.** The Au position cloud represented by 500 random selected coordinates for the initial cloud (blue dots) and the data-optimized cloud (red dots). The centers of the two clouds are labeled with circles of the same color scheme. The attachment point (G-N2, red sphere in **b**) on the RNA helix (black) is shown by the green circle.

## Sequences

Duplex 1

```
     1    S5   S8   S11  S14            S23  S26   30
5′-GCUUGCUGUUGUCGUUCUUCUUGUCGUUGC-3′
3′-CGAACGACAACAGCAAGAAGAACAGCAACG-5′
        C6        C13  C17       C25
```

Duplex 2

```
     1     S6          S15   20
5′-GCUUGGAUCUUCACGAUGUC-3′
3′-CGAACCUAGAAGUGCUACAG-5′
          C9   C12
```

## Base steps investigated

Duplex 1

| | S5 | S8 | S11 | S14 | S23 | S26 |
|---|---|---|---|---|---|---|
| C6 | | | 5 | 8 | 17 | 20 |
| C13 | | | | | 10 | 13 |
| C17 | -12 | -9 | -6 | -3 | 6 | 9 |
| C25 | -20 | -17 | -14 | -11 | | |

Duplex 2

| | S6 | S15 |
|---|---|---|
| C9 | -3 | 6 |
| C12 | -6 | 3 |

**Figure S3. RNA sequences used to determine the solution ensemble of RNA double helix**. Duplexes 1 and 2 have distinct sequences. The tables at the bottom ('Base steps investigated') show the label positions on the top (or sequence, S) strand and the bottom (or complementary, C) strand; the number refers to the number of base steps separating the two Au labels.

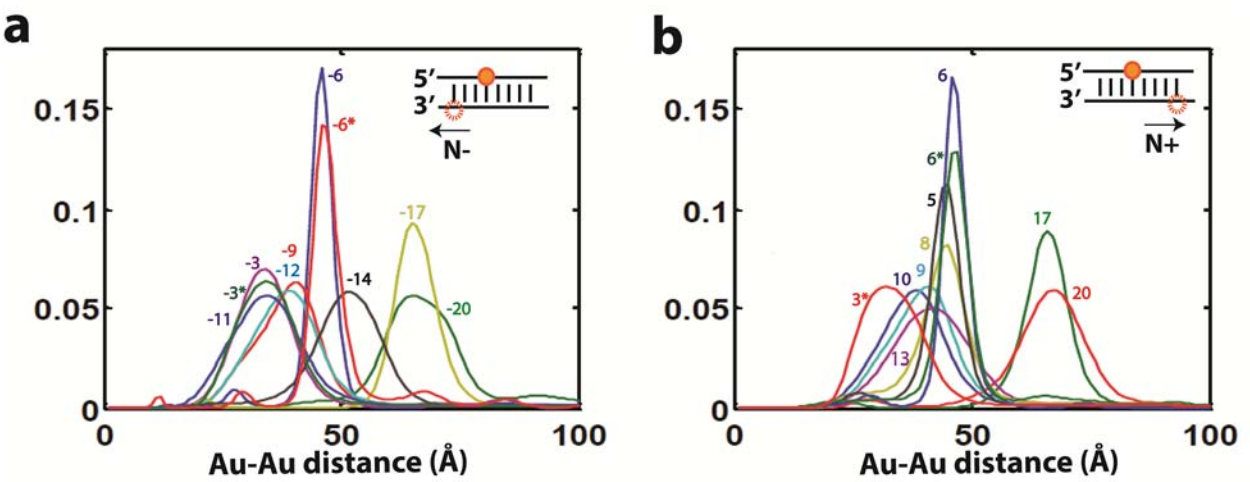


**Figure S4. Experimentally observed distance distributions for pairs of Au-nanocrystal probes labeled along a RNA duplex.** The Au-nanocrystal probes are separated by different negative (**a**) or positive (**b**) numbers of base steps, as indicated by differently colored numbers and lines, for Duplex 1 (no *) and Duplex 2 (with *); the sequences used are given in Supplementary Figure S3. Conditions: 10 mM sodium ascorbate, 70 mM Tris.HCl, pH 7.4, and 150 mM NaCl.

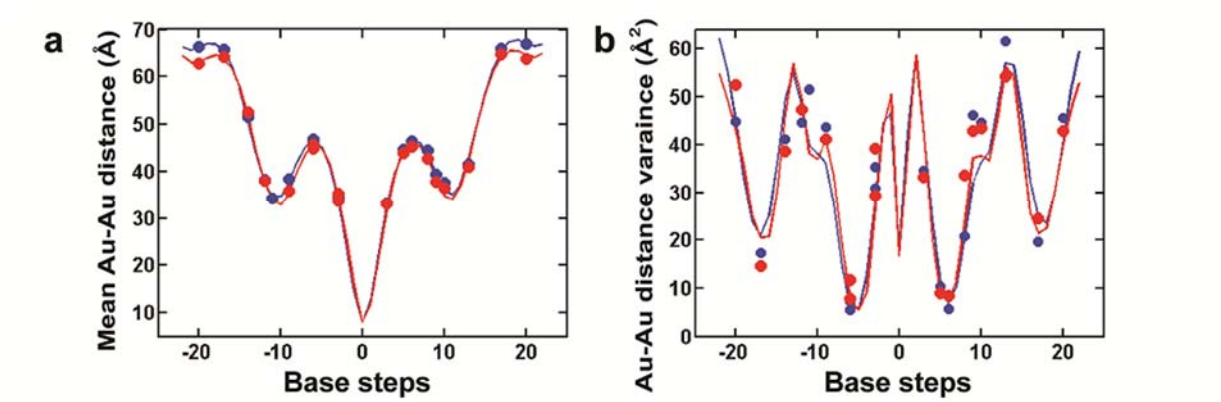


**Figure S5. Observed mean probe-probe distances, probe-probe distance variances, and model predictions for an RNA duplex.** The experimentally obtained mean center-to-center Au-Au distance (**a**, filled circles), Au-Au distance variance (**b**, filled circles), and best-fit model predictions (lines) for solution conditions with (red) and without (blue) 10 mM $MgCl_2$. See Supplementary Table S1 for detailed solution conditions and model parameters.

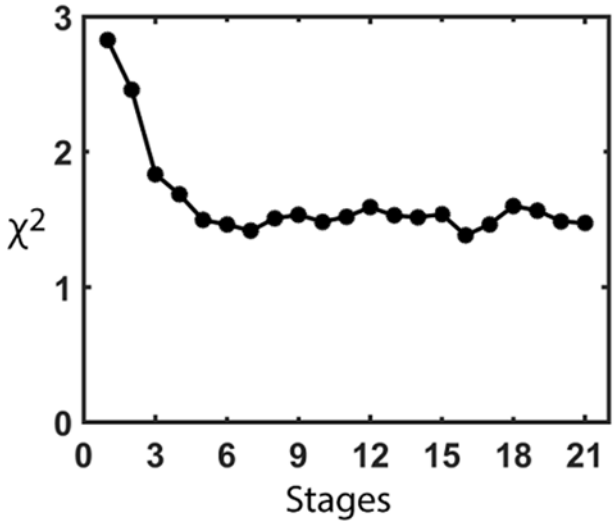


**Figure S6. Search for the optimum Au nanocrystal parameters and RNA helix parameters.** The search for the global minimum in the sum of the $\chi^2$ in mean Au-Au distance and variance converges after several stages of optimization, with variation of one group of parameters (out of three in total) in each stage (see Supplementary notes S4 for details).

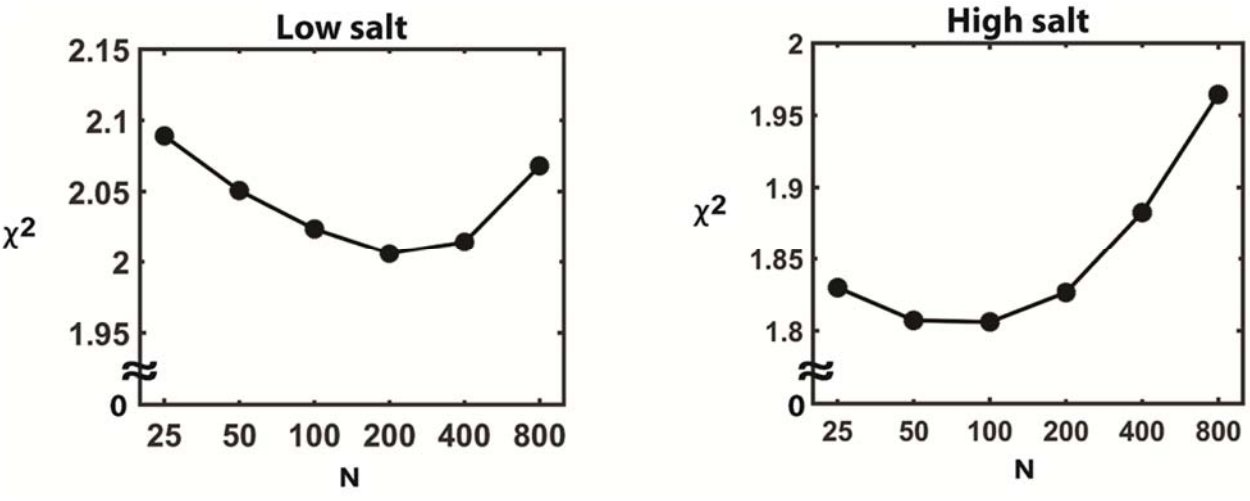


**Figure S7. Determination of the optimum N for empirical Bayesian Sampling (EBS).** The value of $\chi^2$ for the cross-validation predictions varies with N, the number of conformations to be included in each empirical Bayesian Sampling (EBS) step. See Methods *Estimation of the conformational ensemble using XSI data* for the definition of N in EBS. The optimum value of N is that with the smallest $\chi^2$. Salt conditions: left panel, 30 mM Tris.HCl pH 7.4, 10 mM sodium ascorbate and 10 mM NaCl; right panel, 70 mM Tris.HCl pH 7.4, 10 mM sodium ascorbate and 150 mM NaCl.

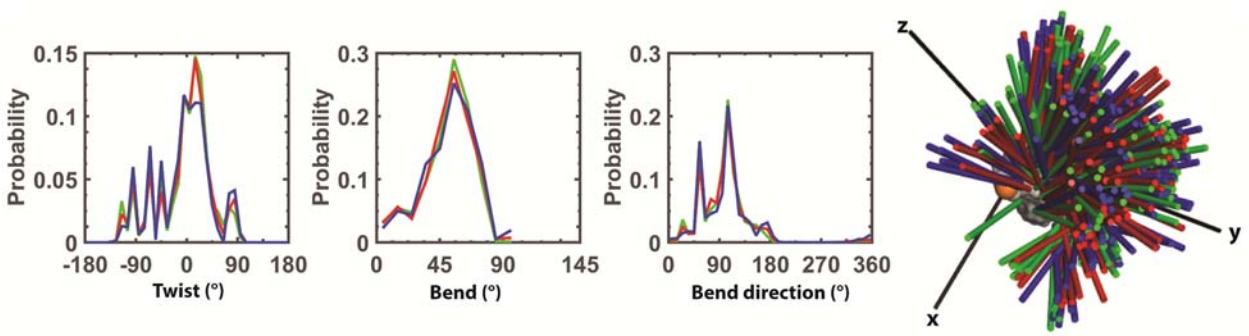


**Figure S8. Comparison of TAR RNA ensembles derived using different N values in empirical Bayesian Sampling (EBS).** See Methods *Estimation of the conformational ensemble using XSI data* for the definition of N in EBS. The N values used are 25 (green), 200 (red) and 1600 (blue). Helical twist, bending, and bending directions are $\alpha + \gamma$, $\beta$ and $\gamma$, respectively, where ($\alpha$, $\beta$, $\gamma$) are Euler angles defined according to the zyz convention(4) and are described in Figure 3a. The colored sticks represent the position of the long axes of the top helix (see Fig. 3a) and each stick represents 1% of the total population.

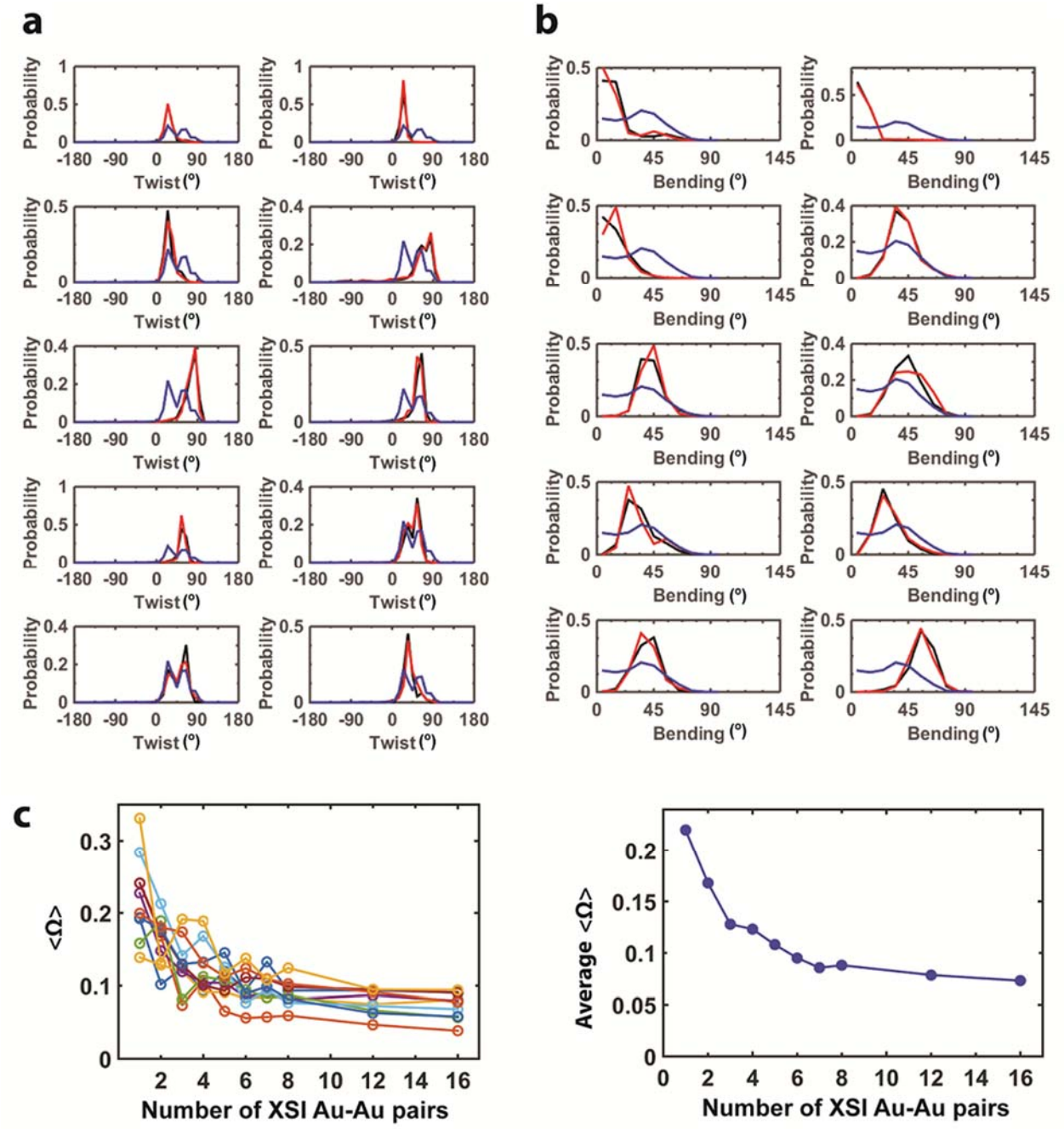


**Figure S9. Synthetic datasets to test the ability of XSI to determine ensembles.** Ten distinct target ensembles (black in the 10 panels of **a** and **b**) were used to generate 10 sets of synthetic XSI data, respectively (see Methods 'Comparison of the theoretical ensemble-solving capability of XSI and RDC using synthetic datasets'). Each of the ten synthetic XSI datasets was used to reweight the allowed space (blue in **a** and **b**) to obtain the corresponding XSI recovered ensemble (red in **a** and **b**). (See Supplementary Note S6 and Methods for a more detailed description of this test.) The ensembles are plotted as probability distributions of helical twist (**a**) and bending (**b**) angles. Helical twist and bending angles are α + γ and β, respectively, where (α, β, γ) are Euler angles defined according to the zyz

convention(4). (**c**) The differences between the ten target ensembles and their corresponding synthetic XSI recovered ensembles become smaller as the number of hypothetical Au pairs used in generating the synthetic XSI data increases from 1 to 16. The differences between ensembles are quantitated using <Ω>, a measure of the differences in Euler angle values between two ensembles in which a smaller <Ω> indicates greater similarity between ensembles(7). The ten sets of tests are plotted individually (left) and as an average (right). The number of hypothetical Au pairs in part **a** and **b** is eight, as was used experimentally in this study.

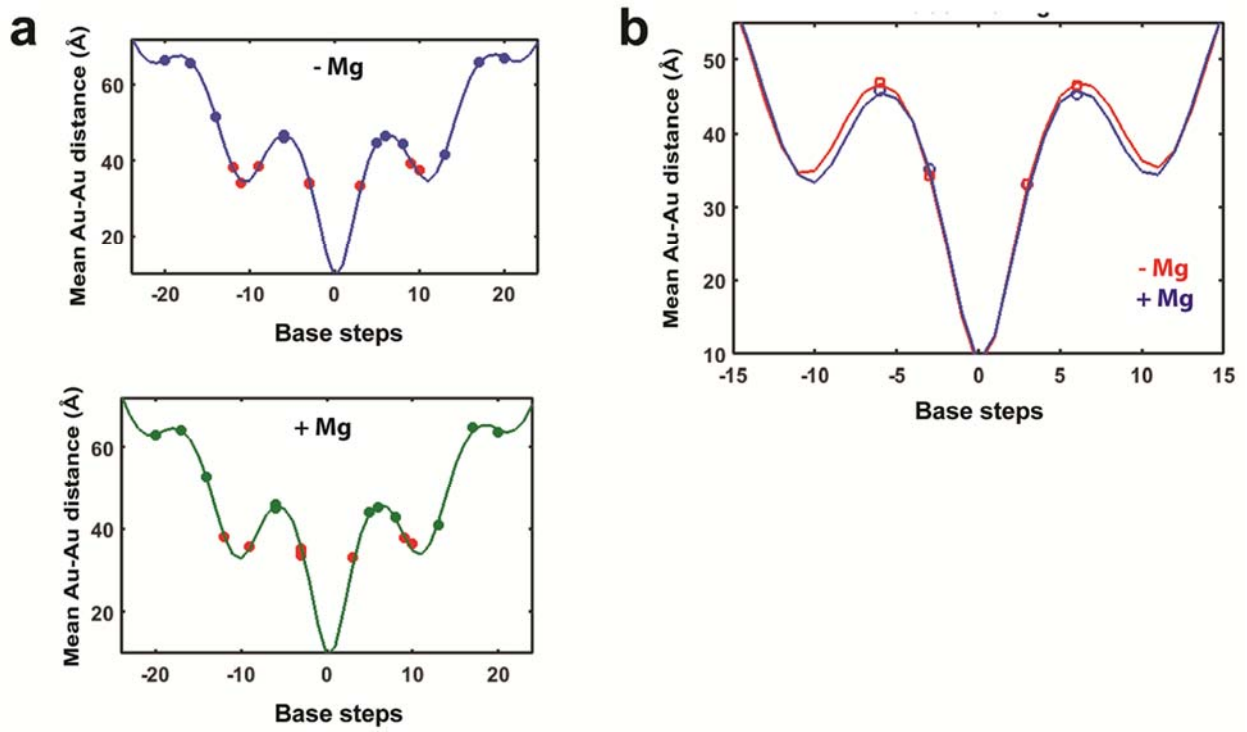


**Figure S10. Accurate prediction of the mean Au-Au distance using partial datasets provides evidence for an absence of structural perturbation from the Au nanocrystal labels. (a)** The measured mean Au-Au distances for Au pairs with small Au-Au separations (<40 Å, red dot) is accurately predicted using models (lines) generated from data for Au pairs with large Au-Au separations (>40 Å, blue and green dots) (top panel: average deviation: 0.2 ± 0.4 Å; bottom panel: 0.1 ± 0.6 Å). [Conditions: 10 mM sodium ascorbate, 70 mM Tris.HCl, pH 7.4, 150 mM NaCl, and 0 (top) or 10 (bottom) mM $MgCl_2$] (b) The measured mean Au-Au distances (circles) for duplex 2 (Supplementary Figure S3) can be accurately predicted using models (lines) generated from duplex 1 data (Supplementary Figure S3) (blue: with $Mg^{2+}$, average deviation: -0.1 ± 0.5 Å; red: without $Mg^{2+}$, average deviation: -0.3 ± 0.5 Å).

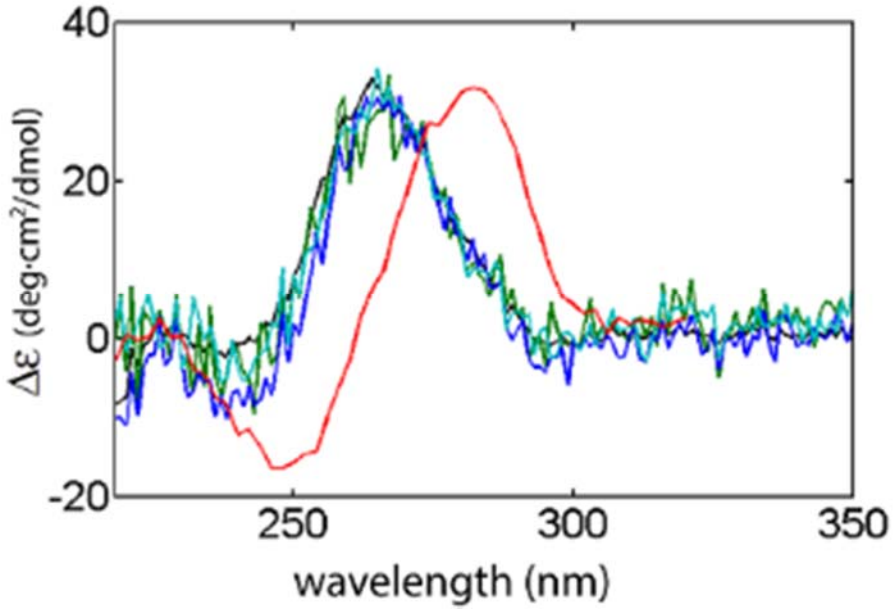


**Figure S11. Internal Au nanocrystal probes do not disturb RNA structure as observed by circular dichroism (CD).** Spectra of an unmodified duplex (black) and three double Au-labeled duplexes with base steps of -6 (green), 6 (blue) and -3 (cyan). The RNA duplexes have the sequence of duplex 2 in Supplementary Figure S3. The spectrum of a DNA duplex (red) is shown to illustrate the difference between A-form and B-form helices.

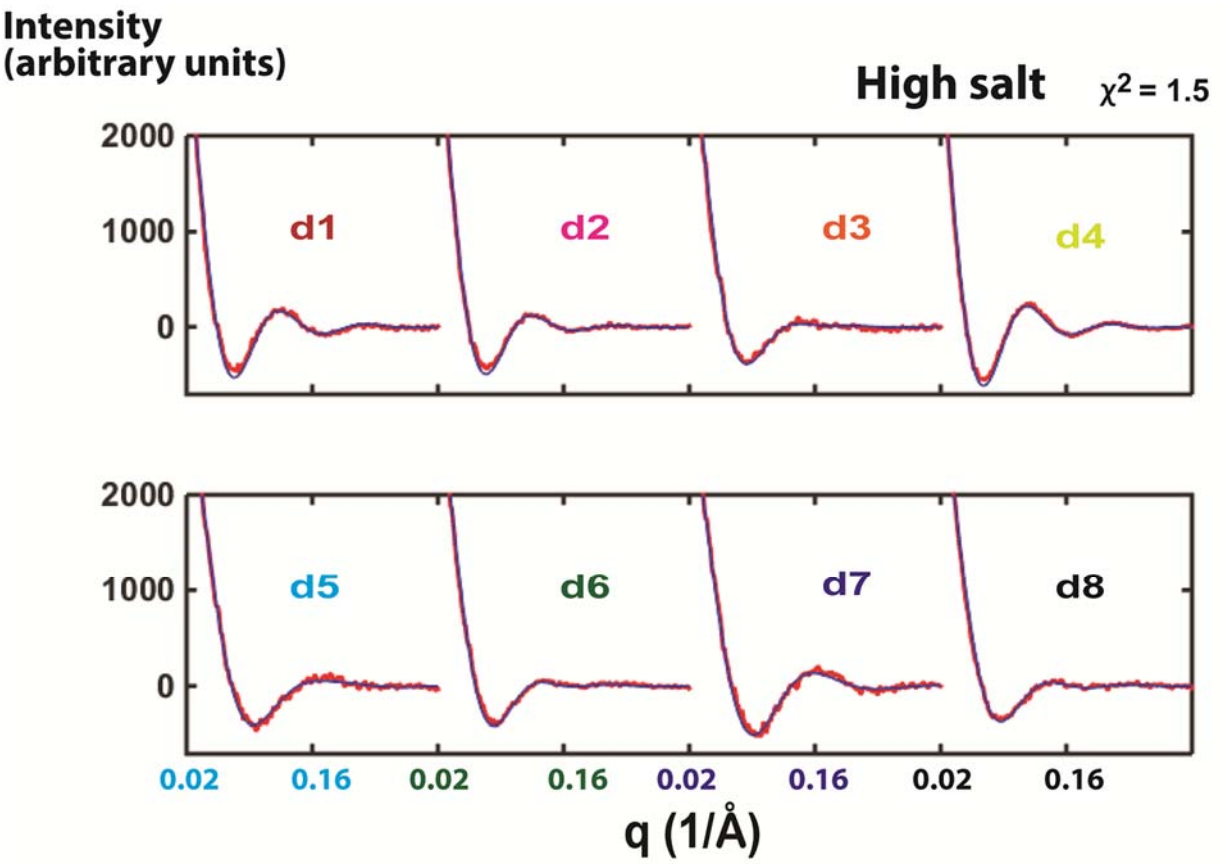


**Figure S12. XSI resolved TAR RNA ensemble.** The measured (red) and best-fit (blue) Au-Au scattering profiles under the solution condition of 10 mM sodium ascorbate, 70 mM Tris.HCl, pH 7.4, and 150 mM NaCl.

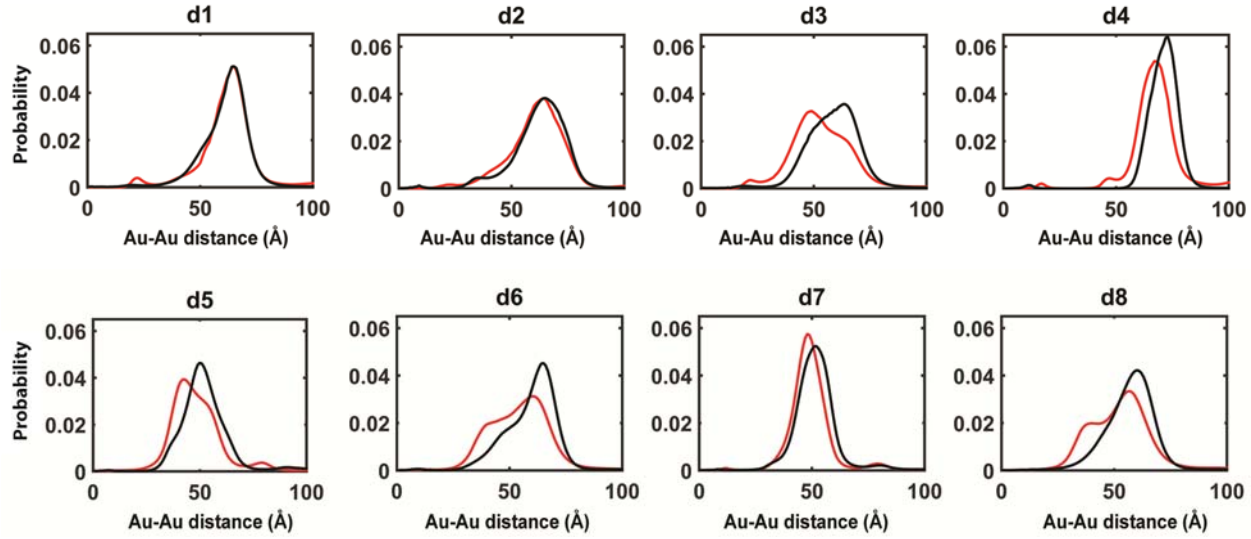


**Figure S13. The Au-Au distance distributions for TAR RNA.** The measured center-to-center Au-Au distance distributions for the constructs d1 to d8 in Figure 2b at low (black, 30 mM Tris.HCl pH 7.4, 10 mM sodium ascorbate) and high (red, 70 mM Tris.HCl pH 7.4, 10 mM sodium ascorbate and 150 mM NaCl) salt conditions.

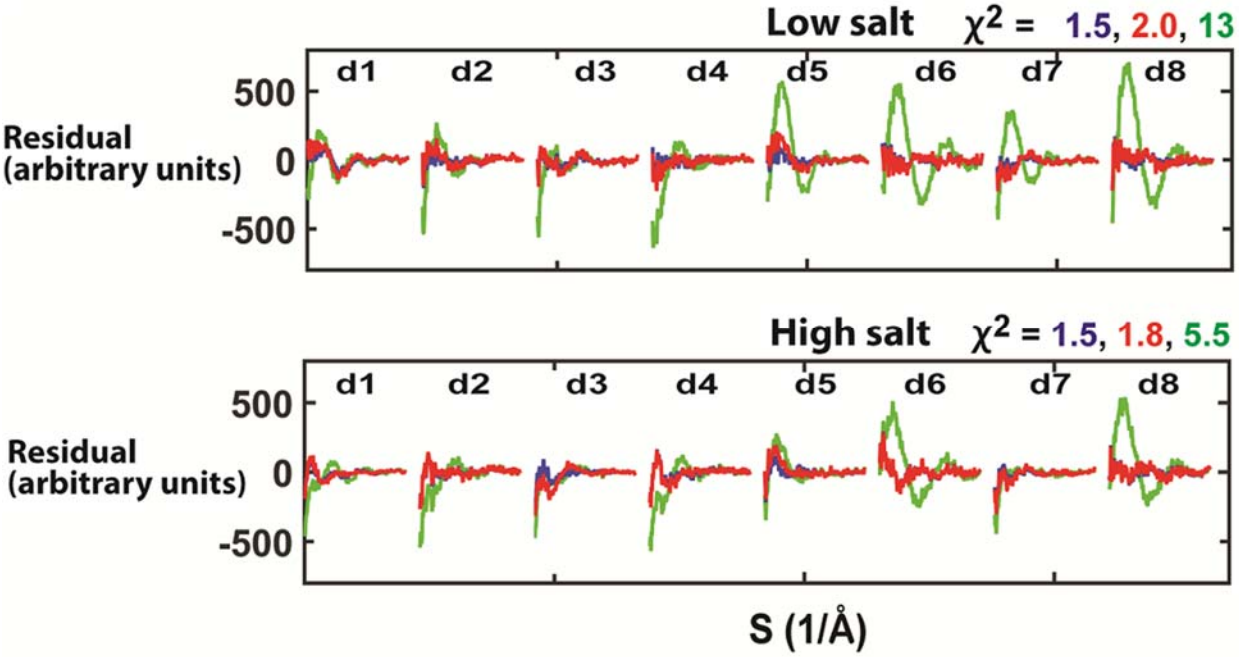


**Figure S14. Error analysis for the XSI-RNA TAR ensembles.** The fitting residuals for the profiles predicted from the best-fit ensemble (blue), the cross-validation predictions (red), and the MD pool (green) under two salt conditions: top panel, 30 mM Tris.HCl pH 7.4, 10 mM sodium ascorbate and 10 mM NaCl; bottom panel, 70 mM Tris.HCl pH 7.4, 10 mM sodium ascorbate and 150 mM NaCl.

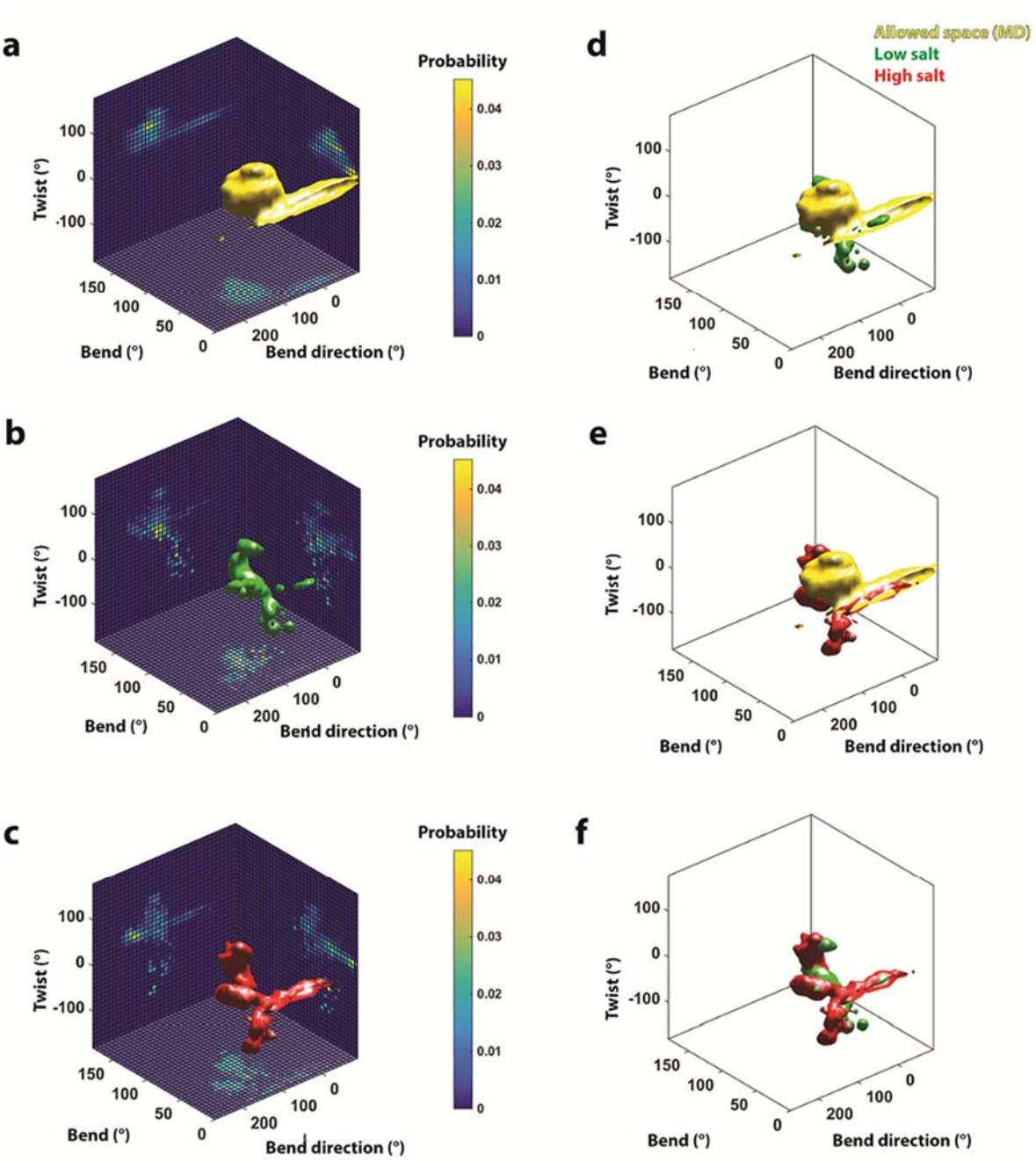


**Figure S15. The conformational space of the TAR RNA in Euler space**. The colored surfaces enclose 90% of the total population in the Euler space for the MD-generated conformational space (**a,** yellow in **d** and **e**) and the XSI-determined conformational space under the low (**b**, green in **d** and **f**) and the high (**c**, red in **e** and **f**) salt conditions. The probability in each set of two dimensions of Euler space (side panels in **a**-**c**) are indicated by the colors and colored scale bars. The color bar is unrelated to the color of the surface in the center of panel **a**-**f**. Euler angles are defined according to the zyz convention such that the three Euler angles [αβγ] represents a series of ordered rotation of α, β and γ around the z, y and z axis, respectively (4). The magnitude of bend, the bend direction and the overall twist are represented by β, γ, and α + γ, respectively. Low salt condition: 30 mM Tris-HCl, pH 7.4, 10 mM NaCl, and 10 mM sodium ascorbate; high salt condition: 70 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 10 mM sodium ascorbate.

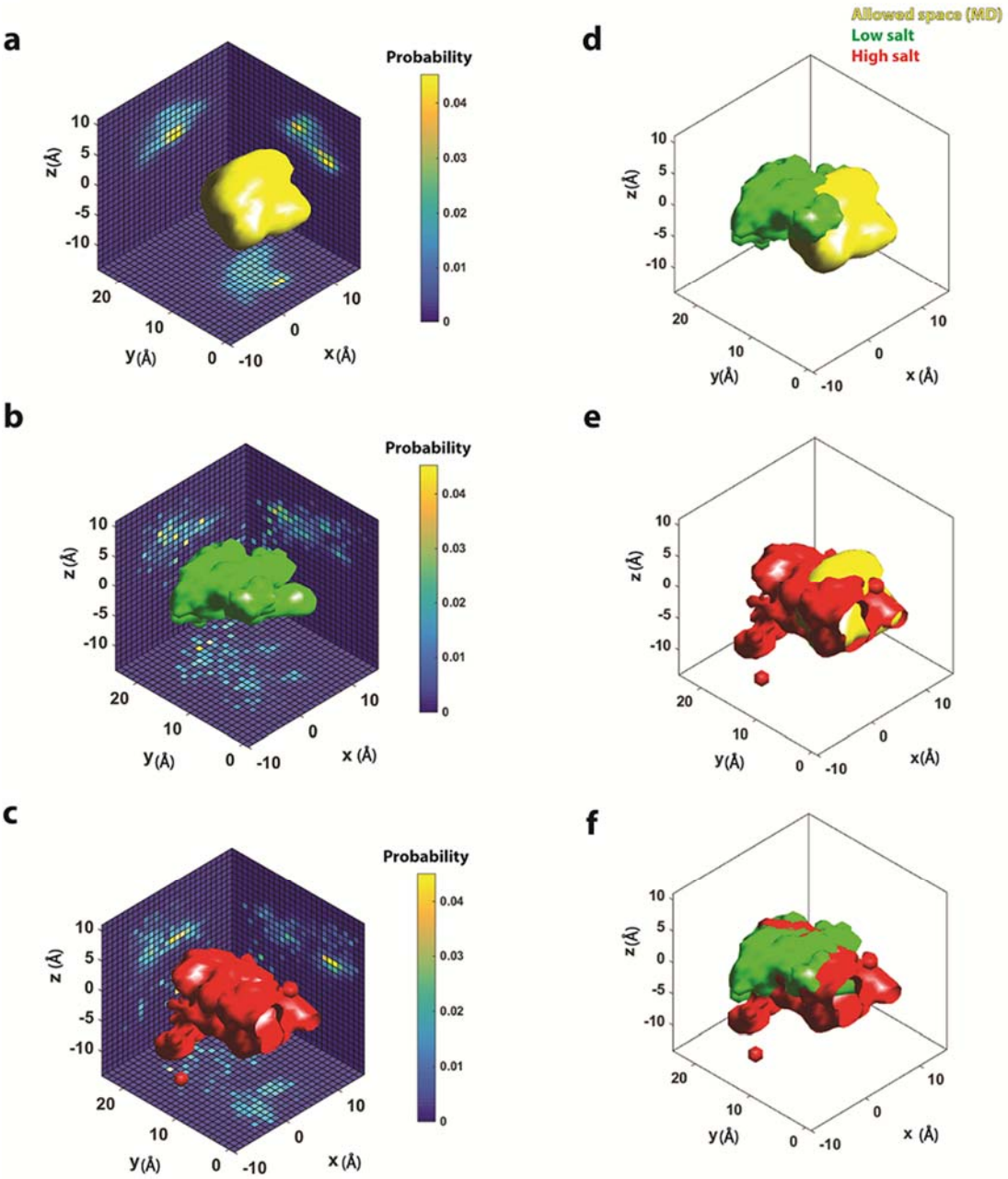


**Figure S16. The conformational space of the TAR RNA in translational space**. The colored surface encloses 90% of the total population in the translational space for the MD generated conformational space (**a,** yellow in **d** and **e**) and the XSI resolved conformational space under the low (**b**, green in **d** and **f**) and the high (**c**, red in **e** and **f**) salt conditions. The probability in each set of two dimensions of translational space (side panels in **a**-**c**) are indicated by the colors and colored scale bars. The color bar is unrelated to the color of the surface in the center of panel **a**-**f**. The xyz Cartesian coordinate system is defined as follows: the z axis is along the long axis of the bottom (5') helix and the positive z direction is the 5' to 3' direction of the sequence strand; the x axis points to O3 of the first nucleotide of the top helix on the sequence strand; and the direction of the y axis follows the right-hand rule. Low salt condition: 30 mM Tris-HCl, pH 7.4, 10 mM NaCl, and 10 mM sodium ascorbate; high salt condition: 70 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 10 mM sodium ascorbate.

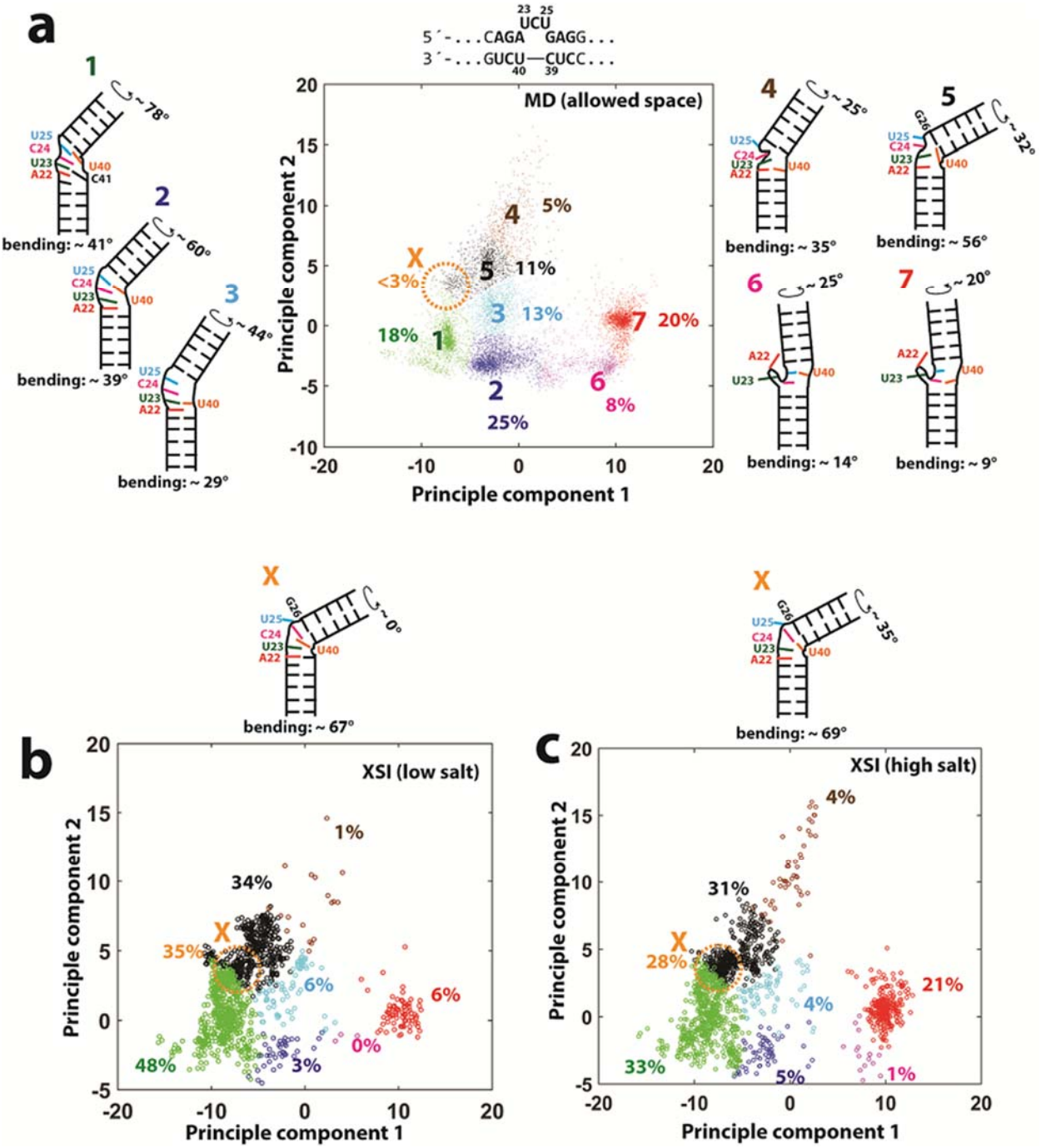


**Figure S17. Ensemble distributions of the local junction structures in TAR RNA.** To explore the distribution of local junctional structures in the MD generated conformational space(8), the distances between six pairs of junctional residues (A22-U40, U23-U40, C24-U40, U25-U40, G26-U40 and A22-C37) were calculated for each of the 10000 MD conformations. Principle component analysis was used to identify the principle components of the six distances in the MD space. As variations in the first two principle components or axes account for most of the variations in local junction structures (76%), these are plotted for the MD conformational space (**a**) and for the conformational space reweighted using XSI

data at low (**b**) and high salt (**c**) conditions. The MD space (**a**) is clustered into seven clusters (color coded in **a**) using K-means clustering and their representative structures are plotted and numbered with color-coded numbers in part **a**. The percentage population (colored coded % label) of each of the seven clusters varies under the low (**b**) and high salt (**c**) conditions. One region, X, is rarely populated in the MD space but is strongly enriched in the XSI ensembles (dotted orange circle in **a**-**c**). The representative conformations in this region (type X) are plotted on top of panel **b** and **c**, with the conformations above each being preferably populated in the respective salt condition.

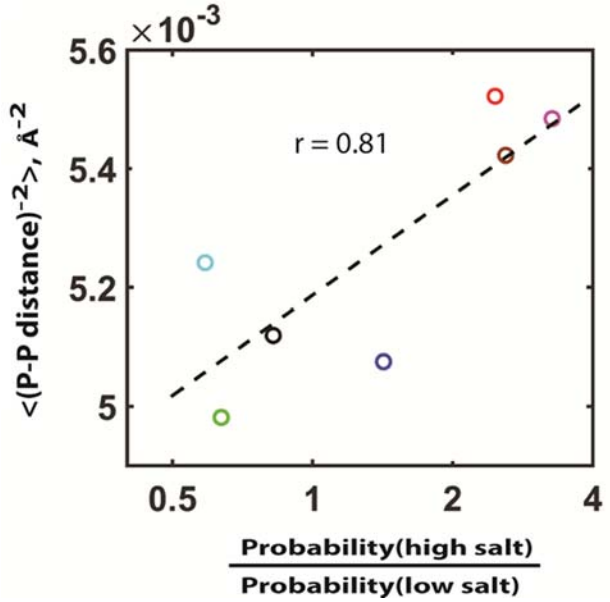


**Figure S18. The difference between the XSI ensembles obtained under different salt conditions are consistent with a simple electrostatic model.** The average in the inverse of the square of the phosphate-phosphate distances is expected to be proportional to the electrostatic energy, in the simplest model. An increase in the solution salt concentration is expected to weaken the electrostatic effect and thereby enrich conformations that have a higher electrostatic energy (i.e., are more destabilized). This model is consistent with the observed ensembles: the salt-dependent changes in the probability of the seven classes of TAR conformations (color coded as in Supplementary Figure S17) in the XSI ensembles in a manner that is positively correlated with the inverse of the square of their phosphate-phosphate distance.

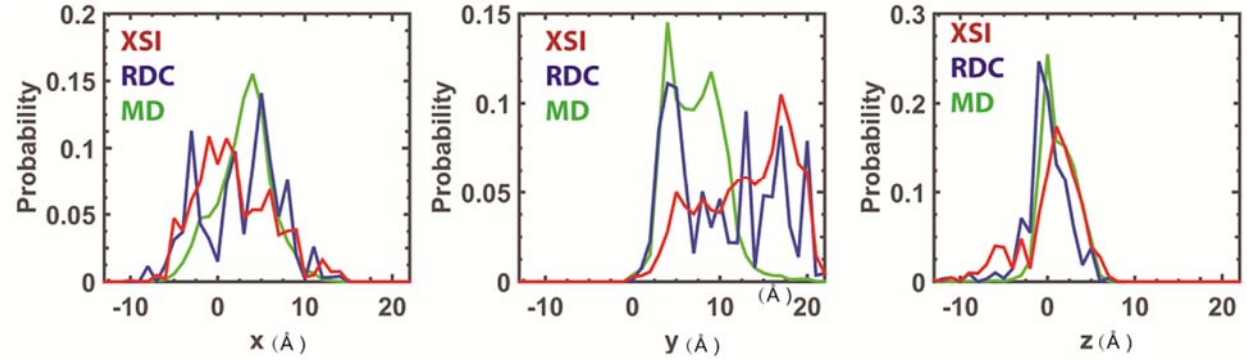


**Figure S19. Comparison of XSI- and RDC-derived TAR RNA ensembles in Cartesian space.** XSI (red) and reported RDC (blue) resolved TAR ensembles under similar low salt conditions (XSI: 10 mM sodium ascorbate and 30 mM Tris.HCl, pH 7.4, and 10 mM NaCl; RDC: 15 mM Na phosphate, pH 6.4, 25 mM NaCl, 0.1 mM EDTA) and the MD-derived ensemble (green). See Figure 4a for a comparison in Euler space.

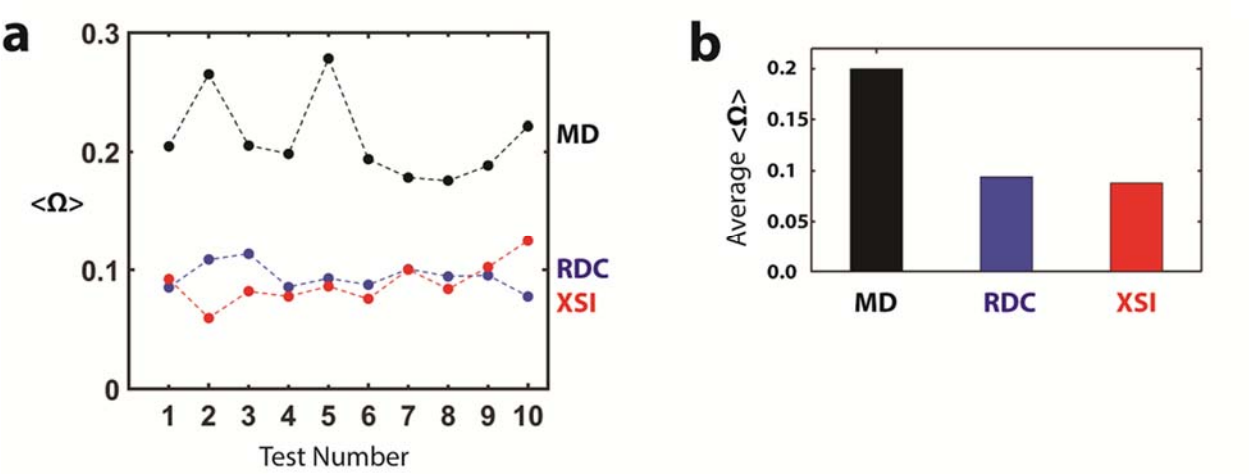


**Figure S20. Comparison of the abilities of XSI and NMR-RDC methods to determine ensembles using synthetic data.** (**a**) The difference between ten target ensembles and their corresponding ensembles derived from synthetic XSI (red) and synthetic RDC (blue) data (see Supplementary Note S6 and Material and Methods for details). The differences between ensembles were quantitated by <Ω>, which measures the differences in Euler angle distributions and a smaller <Ω> indicating greater similarity between ensembles(7). Also compared are the differences between the target ensembles and random MD ensemble solutions generated by randomly select 10% of the MD trajectory (black). (**b**) The average values for <Ω> from part a.

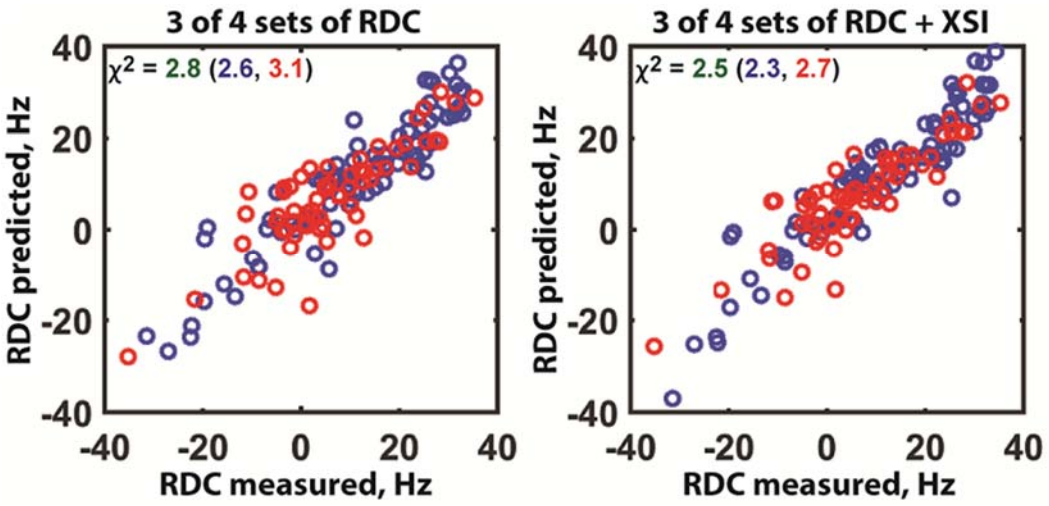


**Figure S21. Comparison of XSI- and RDC-derived TAR RNA ensembles.** Comparison of reported RDC values with RDC values predicted from three of the four sets of RDC measurements without (left) and with additional information from XSI data (right). The $\chi^2$ values were calculated as the average of $(RDC_{calculated} - RDC_{measured})^2 / RDC_{error}^2$ for all RDC values (green); RDC values of the CUC bulge and the two base pairs 5′ to the bulge (red); and the RDC values of the flanking helices (blue).

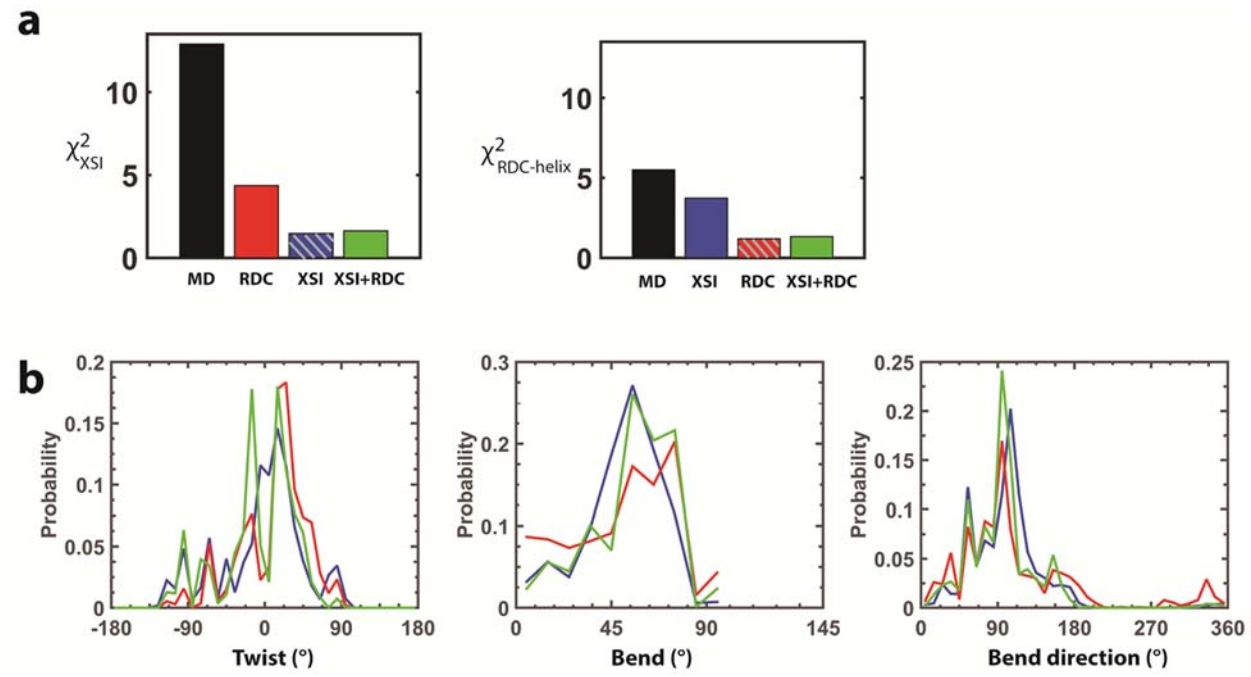


**Figure S22. Comparison of TAR RNA ensembles derived from XSI, RDC, and a combination of XSI and RDC data. (a)** The $\chi^2$ values for XSI profiles (left) and RDC of helical residues (right; see also Figure 4b & 4c) calculated from ensembles determined from MD (black), RDC data (red), XSI data (blue), and a combination of XSI and RDC data (green). The $\chi^2$ values obtained for the parent dataset is shown by the hatched lines. **(b)** TAR RNA ensembles derived using XSI (red), RDC (blue), and a combination of both XSI and RDC data (green). Twist, bend, and bend directions are $\alpha + \gamma$, $\beta$ and $\gamma$, respectively, where ($\alpha$, $\beta$, $\gamma$) are Euler angles defined according to the zyz convention and described in Figure 3a(4).

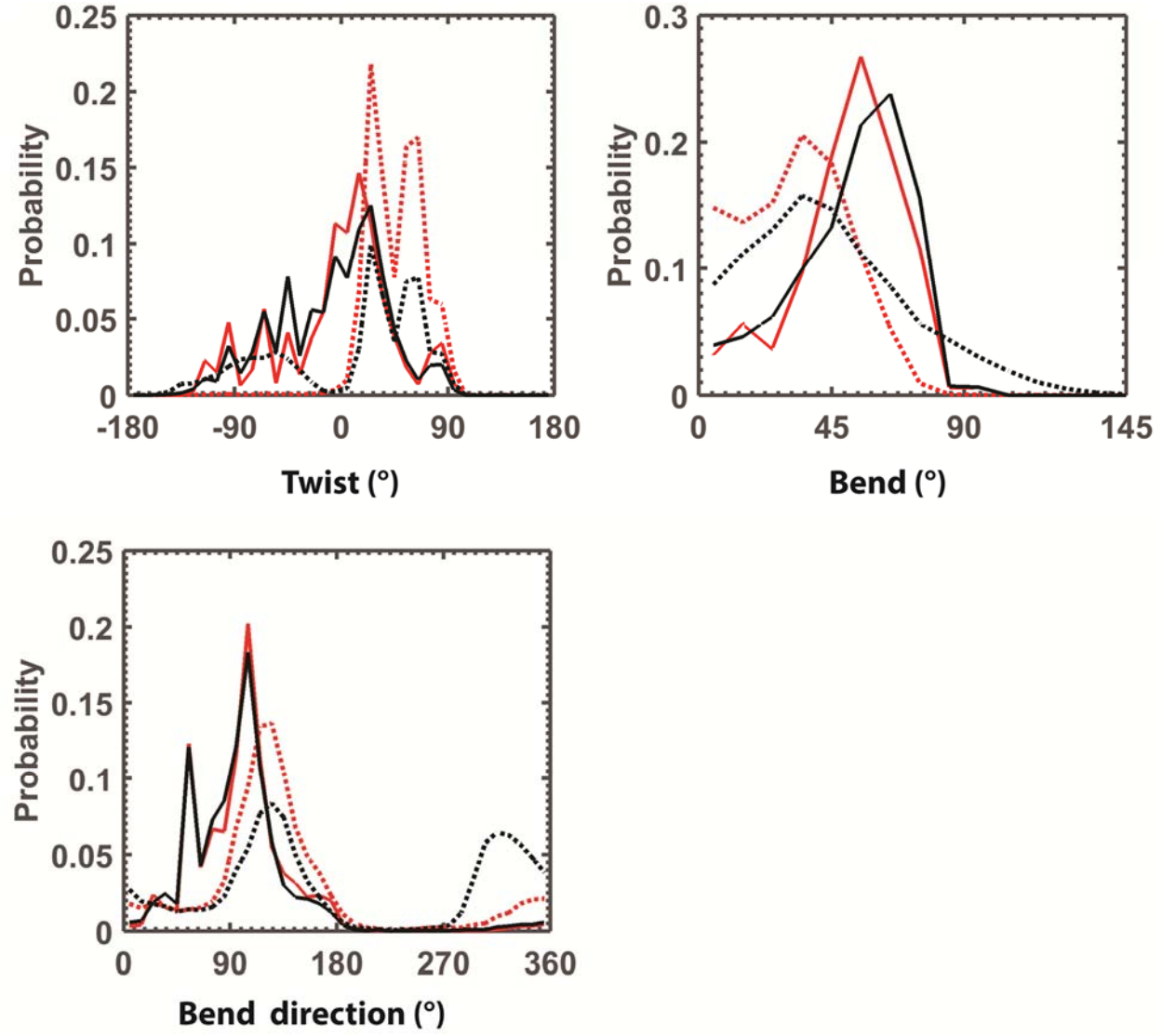


**Figure S23. Comparison of TAR RNA ensembles derived from two different sets of basis conformations.** The TAR RNA ensemble (red solid) derived using the 10,000 MD basis set conformations used herein (red dotted) is similar to the TAR RNA ensemble (black solid) derived using an extended set of 22,000 basis conformations (black dotted) that includes the 10,000 MD conformations plus 12,000 additional conformations that covers all topologically allowed conformational space of the TAR bulge. Twist, bend, and bend directions are $\alpha + \gamma$, $\beta$ and $\gamma$, respectively, where ($\alpha$, $\beta$, $\gamma$) are Euler angles defined according to the zyz convention and described in Figure 3a (4).

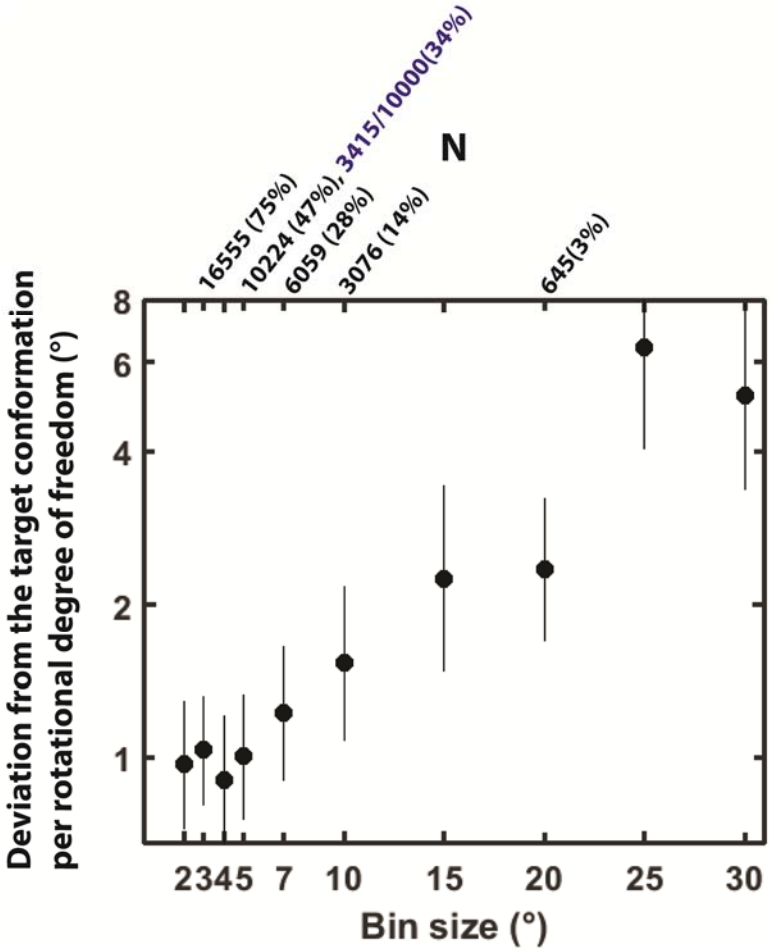


**Figure S24. The effect of the size of the basis set conformation pool on the ability to resolve a conformation.** The extended set of 22,000 basis conformations (black dotted in Figure S23) are binned with different size bins in Euler angle space and only one conformation is retained per bin (pruned) to generate a series of basis set conformational pools. Within each bin, a random conformation was selected. The number (N) and percentage (in parentheses) of conformations retained for each bin size is labeled on the top x-axis. Nine target ensembles that consist of just a single conformation were used to generate nine sets of corresponding synthetic XSI data, which were used to recover the corresponding ensembles and test how well XSI can resolve a single conformation and the effect of the size of the basis set conformation pool on this recovery. The target conformation was removed before the pruning process so that it was never contained within the pool. The deviation (y-axis) of the recovered ensemble from the target conformation is quantitated as the distance in Euler angle space between the weighted average of the resolved ensemble and the target conformation, averaged over the three rotational degree of freedoms, and is plotted as the mean (dots) and the error of the mean (lines) over the nine synthetic tests.

**Table S1. Probe and helical parameters obtained from optimizing the mean Au-Au distances of the Au probes against predictions from a linear elastic rod RNA model using parameters from literature or from re-parameterization against experimental data**

| $MgCl_2$[1] | Probe parameters[2] | | | |
|---|---|---|---|---|
| | $D$ (Å) | $2\theta_0$ (°) | $2axial_0$ (Å) | $\varepsilon_{Au}$ |
| 0 mM | 17.7 ± 0.3 | -4.2 ± 2.5 | -0.2 ± 0.4 | 0.06 ± 0.02 |
| 10 mM | 17.4 ± 0.3 | -7.8 ± 1.9 | -0.5 ± 0.5 | 0.06 ± 0.02 |
| | **RNA helical parameters (average per base step)**[3] | | | |
| | Twist (°) | Tilt (°) | Roll (°) | |
| 0 mM | 30.9 ± 0.4 | -2.3 ± 0.1 | 5.6 ± 2.9 | |
| 10 mM | 31.0 ± 0.4 | -4.2 ± 2.3 | 7.2 ± 0.9 | |
| | **RNA elastic parameters** | | | |
| | B (bending persistence length, nm) | | C (twisting persistence length, nm) | |
| 0 mM | 66 ± 9 | | 42 ± 19 | |
| 10 mM | 87 ± 29 | | 26 ± 16 | |

Parameters in this table were obtained by comparing observed mean Au-Au distance and Au-Au distance variance with predictions from a linear elastic rod model of RNA (see Supplementary Note S4 for details) parameterized for helical parameters and RNA mechanical parameters against experimental data. The best-fit parameters were determined by a numeral search using MATLAB (see Supplementary Note S4 for details) and the errors reported are one standard deviation.

[1.] Common experimental conditions are: 150 mM NaCl, 10 mM sodium ascorbate, 70 mM Tris, pH 7.4.

[2.] The probe parameters D, $\theta_0$, and $axial_0$ define the average position of the Au probe and are defined as previously described(2,3). Briefly, in the coordinate of the base-pair (Figure 1b), D is the radial displacement of the probe from the z-axis, $\theta_0$ is the azimuthal angular rotation of the probe away from the short axis of the base pair, and $axial_0$ is the displacement along the z-axis of the probe from the base pair. $\varepsilon_{Au}$ is a parameter related to linker rigidity; when $\varepsilon_{Au}$ equals to zero, the linker is flexible and the rigidity of the linker increases as $\varepsilon_{Au}$ becomes larger (See Supplementary Note S3 for detail).

[3.] The average shift = 0.0 Å, slide = -1.38 Å and rise = 3.30 Å was not varied and was set to equal to the literature values(5) calculated from the crystal structure database of RNA.

[4]. $Mg^{2+}$ slightly increases the degree of bending at base steps, increases bending rigidity, and decreases twisting rigidity.

**Table S2. Comparison of XSI-derived RNA helix models and literature models.**

| | Helical Parameters | | | | |
|---|---|---|---|---|---|
| | This work | | Literature models | | |
| | XSI (- Mg) | XSI (+ Mg) | RNA fiber diffraction* | RNA crystal structure database(5) | NMR(9) |
| **Twist (°)** | 30.9 ± 0.4 | 31.0 ± 0.4 | 30.5 | 32.4 ± 1.7 | 31.6 ± 1.6 |
| **Tilt (°)** | -2.3 ± 0.1 | -4.2 ± 2.3 | -0.1 | 0.0 ± 0.5 | 0.1 ± 2.4 |
| **Roll (°)** | 5.6 ± 2.9 | 7.2 ± 0.9 | 12.6 | 7.9 ± 2.1 | 10.0 ± 7.6 |
| **Shift (Å)** | 0.0** | 0.0** | 0.0 | 0.0 ± 0.1 | 0.0 ± 0.2 |
| **Slide (Å)** | -1.4** | -1.4** | -1.3 | -1.4 ± 0.2 | -1.3 ± 0.2 |
| **Rise (Å)** | 3.3** | 3.3** | 3.3 | 3.3 ± 0.2 | 3.3 ± 0.4 |

| | Elastic parameters | | | |
|---|---|---|---|---|
| | This work | | Literature values*** | |
| | XSI (- Mg) | XSI (+ Mg) | Magnetic tweezers | AFM |
| **Bending persistence length** | 66 ± 9 | 87 ± 29 | 59(10); 57(11) | 60(10) |
| **Twisting persistence length** | 42 ± 19 | 26 ± 16 | <50(11) | N/A |

*Average values from different RNA fiber diffraction studies, as reported in ref. (12).
**These values were set to equal to the values in the last column.
***Under solution condition of 100-150 mM monovalent salt with no $Mg^{2+}$.

**Table S3. RNA thermal stability with and without internally labeled Au nanocrystals.**

| | Position of Labeling* | $T_m$ (°C) | | |
|---|---|---|---|---|
| Unmodified (U) | N/A | 77.7 ± 0.5 | | |
| | | | $\Delta T_m = (A - U)$ (°C) | |
| Single labeled (A) | S6 | 75.1 ± 0.6 | -2.6 ± 0.8 | |
| | S15 | 74.8 ± 0.5 | -3.1 ± 0.7 | |
| | C9 | 75.8 ± 0.6 | -1.9 ± 0.8 | |
| | C12 | 75.5 ± 0.4 | -2.2 ± 0.7 | |
| | | | | $\Delta\Delta T_m =$ $(A_1A_2\text{-U})\text{-}(A_1\text{-U})\text{-}(A_2\text{-U})$ (°C) |
| Double labeled ($A_1A_2$) | S6 and C9 | 73.3 ± 1.0 | | 0.2 ± 1.5 |
| | S15 and C12 | 74.2 ± 1.3 | | 1.8 ± 1.6 |

Melting temperatures for unmodified (U), single Au labeled (A) and double labeled ($A_1A_2$) RNA duplexes. All duplexes share a common base sequence (duplex 2 in Supplementary Figure S3). The samples were approximately 1 μM concentration in 0.15 M NaCl and 70 mM Tris·HCl, pH 7.4, and the reported $T_m$ values are adjusted to 1 μM. The reported errors are standard deviation from multiple measurements.

*See duplex 2 in Supplementary Figure S3.

## References


1. Shi, X., Huang, L., Lilley, D.M., Harbury, P.B. and Herschlag, D. (2016) The solution structural ensembles of RNA kink-turn motifs and their protein complexes. *Nat. Chem. Biol.*, **12**, 146-152.
2. Shi, X., Herschlag, D. and Harbury, P.A. (2013) Structural ensemble and microscopic elasticity of freely diffusing DNA by direct measurement of fluctuations. *Proc. Natl. Acad. Sci. USA*, **110**, E1444-1451.
3. Mathew-Fenn, R.S., Das, R. and Harbury, P.A. (2008) Remeasuring the double helix. *Science*, **322**, 446-449.
4. Shi, X., Beauchamp, K.A., Harbury, P.B. and Herschlag, D. (2014) From a structural average to the conformational ensemble of a DNA bulge. *Proc. Natl. Acad. Sci. USA*, **111**, E1473-1480.
5. Faustino, I., Perez, A. and Orozco, M. (2010) Toward a consensus view of duplex RNA flexibility. *Biophys. J.*, **99**, 1876-1885.
6. Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108-5121.
7. Yang, S., Salmon, L. and A-Hashimi, H.M. (2014) Measuring similarity between dynamic ensembles of biomolecules. *Nat. Methods*, **11**, 552-554.
8. Salmon, L., Bascom, G., Andricioaei, I. and Al-Hashimi, H.M. (2013) A general method for constructing atomic-resolution RNA enembles using NMR residual dipolar couplings: the basis for interhelical motions revealed. *J. Am. Chem. Soc.*, **135**, 5457-5466.
9. Tolbert, B.S., Miyazaki, Y., Barton, S., Kinde, B., Starck, P., Singh, R., Bax, A., Case, D.A. and Summers, M.F. (2010) Major groove width variations in RNA structures determined by NMR and impact of C-13 residual chemical shift anisotropy and H-1-C-13 residual dipolar coupling on refinement. *J. Biomol. NMR*, **47**, 205-219.
10. Herrero-Galan, E., Fuentes-Perez, M.E., Carrasco, C., Valpuesta, J.M., Carrascosa, J.L., Moreno-Herrero, F. and Arias-Gonzalez, J.R. (2013) Mechanical Identities of RNA and DNA Double Helices Unveiled at the Single-Molecule Level. *J. Am. Chem. Soc.*, **135**, 122-131.
11. Lipfert, J., Skinner, G.M., Keegstra, J.M., Hensgens, T., Jager, T., Dulin, D., Kober, M., Yu, Z.B., Donkers, S.P., Chou, F.C. *et al.* (2014) Double-stranded RNA under force and torque: Similarities to and striking differences from double-stranded DNA. *P Natl Acad Sci USA*, **111**, 15408-15413.
12. Zheng, G.H., Lu, X.J. and Olson, W.K. (2009) Web 3DNA-a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.*, **37**, W240-W246.