

# Biophysical characterisation of human LincRNA-p21 sense and antisense Alu inverted repeats

Michael H. D'Souza<sup>1</sup>, Tyler Mrozowich<sup>1</sup>, Maulik D. Badmalia<sup>1</sup>, Mitchell Geeraert<sup>1</sup>, Angela Frederickson<sup>1</sup>, Amy Henrickson<sup>1</sup>, Borries Demeler<sup>1,3,4</sup>, Michael T. Wolfinger<sup>5,7</sup> and Trushar R. Patel<sup>1,2,6,\*</sup>

<sup>1</sup>Department of Chemistry and Biochemistry, Alberta RNA Research and Training Institute, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada, <sup>2</sup>Department of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 1N4, Canada, <sup>3</sup>Department of Chemistry and Biochemistry, University of Montana, Missoula, MT 59812, USA, <sup>4</sup>NorthWest Biophysics Consortium, University of Lethbridge, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada, <sup>5</sup>Bioinformatics and Computational Biology, Faculty of Computer Science, Währingerstrasse 29, 1090 Vienna, Austria, <sup>6</sup>Li Ka Shing Institute of Virology and Discovery Lab, University of Alberta, Edmonton, AB T6G 2E1, Canada and <sup>7</sup>Department of Theoretical Chemistry, University of Vienna, Währingerstrasse 17, 1090 Vienna, Austria

Received December 08, 2021; Revised April 26, 2022; Editorial Decision April 26, 2022; Accepted May 09, 2022

## ABSTRACT

Human Long Intergenic Noncoding RNA-p21 (LincRNA-p21) is a regulatory noncoding RNA that plays an important role in promoting apoptosis. LincRNA-p21 is also critical in down-regulating many p53 target genes through its interaction with a p53 repressive complex. The interaction between LincRNA-p21 and the repressive complex is likely dependent on the RNA tertiary structure. Previous studies have determined the two-dimensional secondary structures of the sense and antisense human LincRNA-p21 *AluSx1* IRs using SHAPE. However, there were no insights into its three-dimensional structure. Therefore, we *in vitro* transcribed the sense and antisense regions of LincRNA-p21 *AluSx1* Inverted Repeats (IRs) and performed analytical ultracentrifugation, size exclusion chromatography, light scattering, and small angle X-ray scattering (SAXS) studies. Based on these studies, we determined low-resolution, three-dimensional structures of sense and antisense LincRNA-p21. By adapting previously known two-dimensional information, we calculated their sense and antisense high-resolution models and determined that they agree with the low-resolution structures determined using SAXS. Thus, our integrated approach provides insights into the structure of LincRNA-p21 Alu IRs. Our study also offers a viable pipeline for combining the secondary structure information with biophysical

and computational studies to obtain high-resolution atomistic models for long noncoding RNAs.

## INTRODUCTION

The tumour suppressor protein p53 is an important transcription factor that regulates a variety of cellular processes, including cell-cycle control, DNA repair, apoptosis, senescence, and cellular stress responses through the activation and repression of target genes (1,2). Despite playing a critical role in the DNA damage response, p53's genome is frequently mutated in cancer cells, exposing a vulnerability in cell cycle regulation (3,4). Nevertheless, when DNA damage occurs, p53 upregulates the expression of genes involved in the cell cycle arrest and DNA repair processes, which leads to cell survival, but also facilitates the initiation of apoptosis for cancerous cells (5). The p53 pathway itself is composed of a network of genes, regulatory proteins, and their transcriptional products which can help respond to intrinsic and extrinsic stress signals (6). These networks enable the regulation of p53 and can be additionally modulated by long noncoding RNAs (LncRNAs) which have been shown to act in a regulatory role within the p53 pathway. Often, the transcription of LncRNA genes are the targets of p53 itself (7–10).

LncRNAs are noncoding RNA molecules devoid of an open reading frame and are generally around 200–100,000 nucleotides (nts). They also do not retain any significant protein-coding capabilities and are therefore generally not expressed (11–13). LncRNAs were previously thought to have no biological function but have been identified to regulate biological processes by altering gene expression and signal pathways (11). Consequently, LncRNAs play a role

\*To whom correspondence should be addressed. Email: [Trushar.patel@uleth.ca](mailto:Trushar.patel@uleth.ca)

in the regulation of gene expression and appear poised to affect the progression of cancers. Long intergenic noncoding RNA-p21 (LincRNA-p21) is found to be a transcriptional repressor in the p53 pathway, playing a role in triggering cellular apoptosis (14). LincRNAs are also capped, spliced, and polyadenylated due to being RNA polymerase II transcripts (12). Under stress conditions including DNA damage, p53 activates the transcription of LincRNA-p21 which accumulates in the nucleus and associates with the heterogeneous nuclear ribonuclear protein K (hnRNP-K) (15). The hnRNP-K contains RNA recognition motifs Arg-Gly-Gly repeats or hnRNP-K homology (KH) domains and whose role is important for nucleic acid metabolism and transcription (16,17). The hnRNP-K is integral in the induction of apoptosis since it will combine with the p53 promoted and transcribed LincRNA-p21 which will then act to repress p53 target genes resulting in apoptosis (18). LincRNA-p21 is thus required to help direct hnRNP-K to bind to the promoters of the target repressed genes (17). Additionally, hnRNP-K was observed to be a transcriptional coactivator of p53, enabling gene expression in response to DNA damage (16).

Human LincRNA-p21 exhibits two isoforms which further contains the presence of Alu repeats that which influence the function of the RNA (15). Alu elements are particularly important because they are highly conserved among primates and fold to produce independent domains. These repeated DNA sequences comprise upwards of 60% of the human genome and can be divided into several classes including micro-satellites (repeat sequences greater than 7 bp), mini satellites (basic repeats of 7 bp or less), or telomeres. These interspersed repeated DNA sequences are further divided into two classes: Short interspersed elements (SINES), and long interspersed elements (LINES) (19). Alu SINES themselves are repetitive elements present in multiple copies of the genomes they reside in and are named because the family of repeats contains a recognition site for the restriction enzyme *AluI* (20,21). Full-length Alu elements are roughly 300 bp long and are frequently located in the 3'-untranslated regions of genes and their intergenic genomic regions and continue to be the most abundant mobile or transposable element in the entirety of the human genome. Determining the structural-dependent role of LincRNA-p21 Alu elements will have an impact on elucidating their overall function and responsibilities within the cell.

Many studies using molecular and computational structural biology seek to identify LincRNA secondary and tertiary structures, and whether said structures have an impact on their function (22,23). This also includes the application of RNA secondary structure prediction techniques such as selective 2'-hydroxyl acylation analysed by primer extension (SHAPE) (24). Doing so is important because it conceptualises structures present on LincRNA-p21 and can elucidate potential specific interactions within the p53 and hnRNP-K pathways. Previous studies have investigated the secondary structure of LincRNA-p21 Alu Inverted Repeats (IRs) and identifying important functional regions that are involved in LincRNA-p21 nuclear localisation and its subsequent transcriptional factor interactions (15). They identified that the two isoforms of LincRNA-p21 contained Alu IRs that retained integral secondary structures that can fold

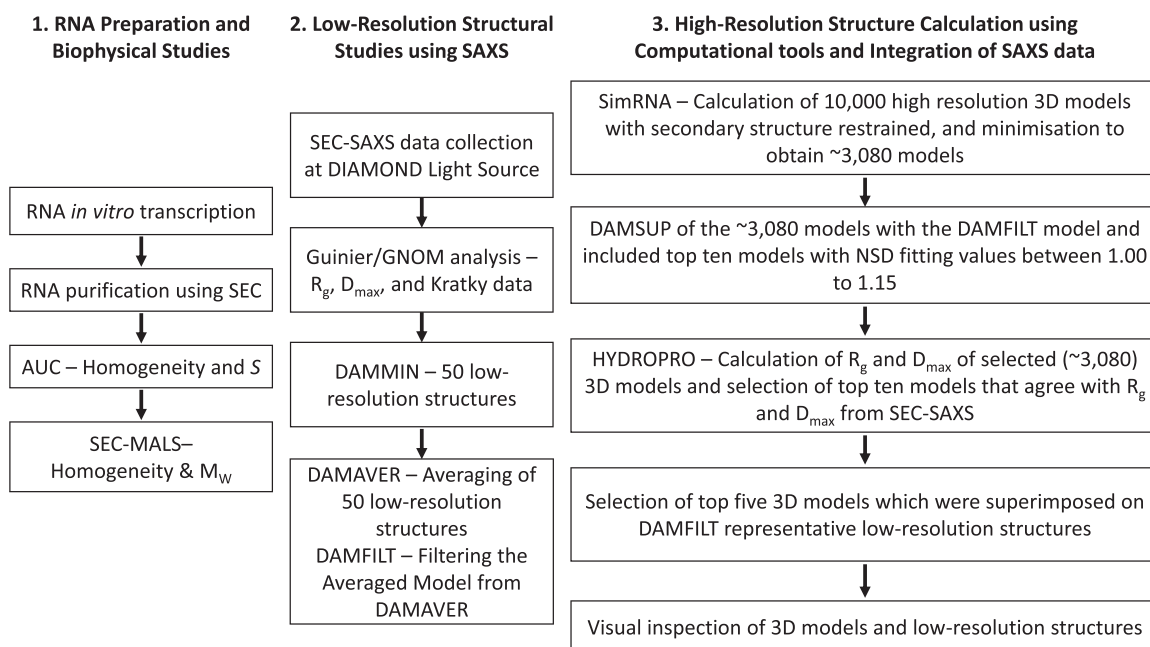
into independent domains. These structures were suggested to be conserved in primates and contribute towards the regulation of cellular localisation of LincRNA-p21 during the cellular stress response. LincRNA structure determination has been challenged by several limitations, including the absence of unique LincRNA secondary structure databases, the inherent complexity in RNA folding, preparation of sufficient amounts of homogenous RNAs, the inability of large RNAs to crystallize, and computational approaches to converge the secondary structure information to tertiary structures (23). RNA structures involving localised or long-range interactions such as stem-loops or pseudoknots, respectively, have challenged computational prediction methods (25). The development of *ab initio*, sequence-based secondary structures is equally difficult due to the insufficient secondary structure information and sequences for comparative analysis. Computational methods can also be resource-demanding and have a tendency of becoming unreliable with longer sequences (26). Consequently, there are no experimental, high-resolution, three-dimensional structures of full-length LincRNA motifs which make their representation and investigation difficult (23).

Incidentally, the intent of this study is to investigate the overall structure of the sense and antisense LincRNA-p21 *AluSx1* Inverted Repeats by employing small angle X-ray scattering (SAXS) and computational modelling to develop their three-dimensional structures (27). We have employed analytical ultracentrifugation (AUC) and size exclusion chromatography coupled to multiangle light scattering (SEC-MALS) instruments to biophysically characterise transcribed and purified *AluSx1* RNAs. AUC experiments revealed that *AluSx1* RNAs were present as monomeric, full-length transcripts under denaturing conditions, while SEC-MALS characterised their Molecular Weight ( $M_w$ ). By combining chemically probed secondary structure information proposed by Chillón and Pyle, 2016, and SimRNA computational modelling, several three-dimensional, high-resolution models can be calculated and fitted to SAXS determined structures. We determined that human LincRNA-p21 isoform LIsoE2's *AluSx1* RNA adopts an asymmetrical, and extended structure in near-physiological buffer conditions. The importance being that no previous three-dimensional structure determination has been performed on LincRNA-p21 *AluSx1* IRs which can improve on the understanding of its interactions with hnRNP-K. We describe a workflow utilising SAXS and SimRNA computational modelling to produce three-dimensional, high-resolution models devised from two-dimensional structures determined *via* SHAPE and other secondary structure probing techniques.

## METHODS

### Sense and antisense LincRNA-p21 *AluSx1* plasmid preparation

A flowchart of the procedure is outlined in Figure 1. LincRNA-p21 *AluSx1* transcripts for the sense and antisense (taken from the TP53COR1 gene located on Chr6:36,663,392–36,667,296 (GRCh38/hg38), Chr6:36,631,169–36,635,073 (GRCh37/hg19)) were designed from the hLincRNA-p21 LIsoE2 isoform sequences



**Figure 1.** Organisational Flowchart for the Purification and Characterisation of Sense and Antisense LincRNA-p21 *AluSx1* RNA. The determination of LincRNA three-dimensional, low-resolution structures overlaid by high-resolution, atomistic models was conducted in three phases: RNA preparation and biophysical studies to determine sample homogeneity and sample properties; low-resolution structure determination by SAXS; and high-resolution modelling using SimRNA, with constraints imposed by HYDROPRO. All methods are further described below.

presented in from Chillón and Pyle (14,15). *AluSx1* IRs are bidirectionally transcribed from the total TP53COR1 gene, producing sense and antisense designated transcripts (15). Thus, RNA constructs used in this experiment are represented below:

>LincRNA-p21 *AluSx1* Sense RNA Sequence (307nt) | TP53COR1\_LIsoE2\_ *AluSx1*\_P

5'-AGCUGGGCGUGGUGGCUCACGCCUGUAA  
UCCCACCACUUUGGGAGGCCGAGGCAGGCG  
GAUCACUUGAGGUCAGGAGUCCAAGACCAGCC  
UGGCCAACAAAGGCGAAACCCUGUCUCUACUAA  
AAUACAAAAACUAGCUGGGCGUAGUGGUG  
GGCACCUGUAAUCCAGCUACUCGGGAGGCUG  
AGACAGGACAAUCGCUUGGACUCCGGAGGCAG  
AGGUUGCAGUGAGCUGGGAUUCGUGCCACUA  
CACUCCAGUCUGGGCGACAGAGCAAGACUCUG  
CAUAAAAAAAAAAAAAAAAAGAAAGAGUAAUAA-3'

>LincRNA-p21 *AluSx1* Antisense RNA Sequence (280nt) | TP53COR1\_LIsoE2\_ *AluSx1*\_P

5'-GCAGAGGAGGAAUGGAAUCAUUCUUUUU  
UUUUUUUUGGAGACGGAGUCUCACUCUGU  
UGCUCAGGCUGGAGUGUAGUGGUGCGAACU  
UGGCUCACUCGAGCCUCCACUCCAGGCUCA  
AGCAAUUCUCCUGCCUCAGCCUCCGAGUAGC  
UGGGAUUACAGGUGUCUGCUAUCACACCCAGC  
UAAAGUUUUUAUUAUUUUUAGUAGAAAUGGA  
GUUUCACCAUGUUGGACAGGCUGGUCUCGA  
ACUCCUGACCUCAGGUGAUCCACCCGCCUCAG  
CCUC-3'

This human LincRNA-p21 isoform LIsoE2 *AluSx1* RNAs will be referred to as sense and antisense *AluSx1* RNAs throughout.

The plasmids were synthesised commercially, each sequence was flanked by a T7 RNA polymerase promoter sequence at the 5'-end and an *XbaI* restriction endonuclease cut-site sequence at the 3'-end. To increase RNA yield, two additional Gs were added to the 3'-end of the T7 promoter region which is reflected in the theoretical  $M_w$  (Table 1) (28). LincRNA-p21 sequences were inserted into Genewiz pUC-57-KAN plasmids (Azena Life Sciences, USA). Plasmids were transformed and cultured in *E. coli* NEB $\alpha$  (NEB, Canada) competent cells and were purified using NEB Monarch Miniprep Kits (NEB, Canada) as per manufacturer's protocol.

#### ***In vitro* transcriptions of LincRNA-p21 sense and antisense *AluSx1* inverted repeats and RNA purification**

RNA transcripts were prepared using run-off *in vitro* transcriptions (IVT) as prepared previously (29,30). Briefly, concentrated plasmid samples were digested by *XbaI* restriction endonuclease (NEB, Canada). 1 mL *in vitro* transcription reactions were performed using laboratory purified in-house T7 RNA polymerase and commercial RiboLock RNase inhibitor (Thermo Fisher Scientific, USA) (31). Linearised plasmids were additionally incubated with 10% DMSO and 0.1% Triton X-100 to increase RNA transcript yields (32). Sense and antisense *AluSx1* RNA were purified by SEC using a Superdex 200 Increase GL 10/300 (Global Life Science Solutions USA LLC, Marlborough, MA, USA) and purification buffer (PB) (10 mM Bis-tris pH 5.0, 100 mM NaCl, 15 mM KCl 15 mM MgCl<sub>2</sub>, 10% glycerol) with an ÄKTA pure FPLC (Global Life Science Solutions USA LLC, Marlborough,

MA, USA) (33). SEC peak fractions were assessed for purity by urea-polyacrylamide gel electrophoresis (Urea-PAGE) and sedimentation velocity analytical ultracentrifugation (SV-AUC) in 6M urea. Urea-PAGE (10%) was run at room temperature, 300V, for 40 min in 1x TBE (Tris-Borate-EDTA) buffer, followed by staining with Sybr<sup>TM</sup> Safe (ThermoFisher Scientific, Saint-Laurant, QC, Canada) and visualisation. Pure fractions were pooled and concentrated by ethanol precipitation, with resuspension in HEPES Folding Buffer (HFB) (50mM HEPES, 150 mM NaCl, 15 mM MgCl<sub>2</sub>, 3% Glycerol, pH 7.4) for SAXS submission.

### **Multiangle light scattering (MALS), and analytical ultracentrifugation (AUC) studies of LincRNA-p21 *AluSx1* sense and antisense inverted repeats**

SEC purified LincRNA-p21 *AluSx1* RNAs were subjected to an additional SEC purification by a Superdex 200 Increase 10/300 GL column in HFB and analysed directly by an in tandem DAWN Multiangle Light Scatterer (MALS) with Optilab Refractive Index System (Wyatt Technology, USA) to determine the  $M_w$  as per Wyatt Technologies guidelines (34). Samples were eluted at a 0.5 mL/min flowrate and measured using 18 multiangle detectors, including a UV A<sub>260</sub> and A<sub>280</sub>, and a refractive index (RI) detector. MALS measurements were taken using a helium-neon red laser (632.8 nm) at 25°C. For data analysis, the refractive index increment ( $dn/dc$ ) was adjusted to 0.1721 mL/g for sense and antisense *AluSx1* RNA samples (35–37). Data were analysed using the ASTRA v9 software package (Wyatt Technology, USA) (38,39).

Purified sense and antisense *AluSx1* RNA were measured by SV-AUC under denaturing conditions in 6M urea to ascertain purity and composition of the transcript. Both samples were measured in two-channel centrepieces and spun at 25,000 rpm for 6 hours at 20°C. Denaturing 6M urea buffer density (1.0899 g/mL) and viscosity (1.3896 cP) were estimated with Ultrascan and used to convert observed sedimentation and diffusion coefficients to standard conditions (water at 20°C). Data were collected in intensity mode at 260 nm and processed using the UltraScan III Software (40). SV-AUC data were processed as described in (41). Briefly, systematic noise contributions and boundary conditions (meniscus and bottom of the cell position) samples were processed with the two-dimensional spectrum analysis (42). Data was further refined by genetic algorithm analysis to achieve parsimonious regularisation (43). The final step included a genetic algorithm-Monte Carlo (GA-MC) analysis, that was performed with 50 iterations to obtain 95% confidence intervals for the determined parameters (Table 1) (44). AUC data analysis was performed on the XSEDE high-performance computing infrastructure using Expanse and Bridges-2 at the San Diego and Pittsburgh supercomputing centres, respectively. The final model produced very low residual mean square deviations (RMSD) of 0.00139 at 0.438 OD<sub>260</sub> for sense and 0.00177 at 0.71 OD<sub>260</sub> for antisense LincRNA-p21 *AluSx1*. All fits produced random residuals, which, together with the low RMSD is evidence for excellent convergence.

### **Small angle X-Ray scattering (SAXS) analysis of LincRNA-p21 *AluSx1* sense and antisense**

SAXS data for sense and antisense *AluSx1* RNA samples were collected at 2.5 mg/mL. Samples were run at Diamond Light Source Ltd. synchrotron (Didcot, Oxfordshire, UK) on the B21 SAXS beamline, with a high-performance liquid chromatography (HPLC) system attached upstream to ensure sample monodispersity (45). A specialised flow cell was employed in conjunction with an inline Agilent 1200 HPLC system (Agilent Technologies, Stockport, UK); sense and antisense *AluSx1* RNA samples were injected onto a Shodex KW403-4F (Showa Denko America Inc., New York, NY, USA) size exclusion column pre-equilibrated with HFB. The flow rate of the column was maintained at 0.160 mL/minute with eluted samples being exposed to X-rays with 3 second exposure time and 600 frames.

Analysis of scattering data was carried out using the AT-SAS suite (46). Using Chromixs, the buffer contribution was subtracted from the sample peak (47). A Guinier analyses ( $q^2$  vs.  $\ln(I(q))$ ) was performed on each data set to obtain the radius of gyration ( $R_g$ ) and to determine the sample's quality (48). A dimensionless Kratky plot ( $qR_g$  vs  $qR_g^2 \cdot I(q)/I(0)$ ) was generated to evaluate folding of RNA molecules (49). A paired-distance distribution function (P(r) analysis was performed using GNOM to obtain real-space  $R_g$  and the maximum particle dimension ( $D_{max}$ ) of the sample (50,51). Employing the information derived from the P(r) plot, a total of fifty sense and antisense *AluSx1* RNA models were generated using DAMMIN (52). These models were then averaged using DAMAVER and then filtered using DAMFILT to produce a single representative model of each of the RNAs (52,53).

### **Sense and antisense LincRNA-p21 *AluSx1* RNA tertiary structure determination**

Using the secondary structure information from Chillón and Pyle, 2016, sense and antisense *AluSx1* tertiary structures were calculated using SimRNA v3.20 (27). SimRNA v3.20 is a Monte Carlo sampler that operates on a coarse-grained model of RNA structure. SimRNA employs a five-bead system per nucleotide, as well as an empirically derived knowledge-based potential. A total of 20 million SimRNA iterations in replica exchange mode were performed for both sense and antisense *AluSx1* RNAs. SimRNA clustering was then performed within one percent of all trajectories with the lowest energy. A RMSD cut-off of five was applied to filter 3080 clusters of similar structures for both sense and antisense *AluSx1* RNAs.

### **High-resolution structural modelling of LincRNA-p21 *AluSx1* sense and antisense**

The representative cluster models containing 3080 computationally generated high-resolution models for both the sense and antisense *AluSx1* were separately assessed by HYDROPRO to generate hydrodynamic properties for each model (54). Running conditions for HYDROPRO involved buffer properties for HFB as determined by UltraScan III:

a viscosity of 1.10068 cP; and buffer density of 1.014 g/cm<sup>3</sup> (40). The theoretical  $M_w$  of 99.418 kDa and 89.543 kDa were applied to the HYDROPRO parameters of sense and antisense *AluSx1* RNA, respectively. Models were superimposed onto the SAXS DAMFILT structures and fitted using DAMSUP. Models exhibiting an NSD (normalized spatial discrepancy) value of 1.00 to 1.15 which indicates close fitting were further selected to represent the high-resolution, atomistic RNA model (53). Models exhibiting similar HYDROPRO determined  $R_g$  and  $D_{max}$  were further selected for and formed the top ten models of interest. The top ten models were energy minimised using an additional step involving QRNAS, which employed a subset of the AMBER force field to achieve energy minimisation of the structures generated from coarse-grained three-dimensional modelling (55). 20,000 QRNAS MD iterations were performed from the original SimRNA full-atom reconstructed high-resolution models that best-fit the averaged, filtered low-resolution, three-dimensional structure obtained from DAMFILT. Subsequently, five best fit models were superimposed on SAXS structures and represented using PymOL (53,56).

## RESULTS

### Purification of LincRNA-p21 sense and antisense *AluSx1* inverted repeats

Both sense and antisense *AluSx1* RNAs were purified using SEC, eluting principally around ~10.0 – 12.5 mL at a flowrate of 0.5 mL/min on the Superdex 200 Increase GL 10/300 (Figure 2A). The left peak indicates plasmid excluded from subsequent analysis while the right peak represents the RNA of interest. Figure 2B depicts the 10% Urea PAGE gel for the RNA fractions indicating that both RNAs migrated closely with similar length and, around the ~300bp marker. Sense LincRNA-p21 *AluSx1* generally produced closely eluting bands (around ~300bp). SV-AUC experiments were conducted using the single fractions collected at 11.0 mL and 11.5 mL for sense and antisense *AluSx1* RNA respectively (indicated by the right shoulder, blue inset, Figure 2A).

### Biophysical characterisation of LincRNA-p21 sense and antisense *AluSx1* inverted repeats

Sedimentation and diffusion coefficients and their 95% confidence intervals resulting from the GA-MC analyses are listed in Table 1. Together with sequence based molar masses, SV-AUC results can be used to derive partial specific volumes and anisotropies for the RNA measurements. Since both RNA molecules were measured in the same urea buffer, it is reasonable to assume that the partial specific volume is similar for both molecules. Sedimentation experiments were performed in 6M urea to denature the molecule and disrupt hydrogen bonding within double-stranded RNA regions of the molecule. Results shown in Figure 2C and D indicate that both samples contain one major species with similar sedimentation coefficients, and molar masses in agreement with molar masses predicted from sequence when using a partial specific volume of 0.516 mL/g. This result is consistent with a monomeric and homogeneous full-length transcript of sense (84% of total

**Table 1.** Solution Properties of Sense and Antisense LincRNA-p21 *AluSx1*

Sample	Sense LincRNA-p21 <i>AluSx1</i>	Antisense LincRNA-p21 <i>AluSx1</i>
$M_w$ Theoretical (kDa)	99.418	89.543
$M_w$ AUC (kDa) <sup>+</sup>	94.770	92.561
$M_w$ SEC-MALS (kDa) <sup>∇</sup>	99.24 ± 0.01	94.52 ± 3.71
Sedimentation Coefficient, $s_{20,w}$ (10 <sup>-13</sup> s) <sup>+</sup>	5.56 ± 0.25	5.53 ± 0.05
Diffusion Coefficient, $D_{20,w}$ (10 <sup>-7</sup> cm <sup>2</sup> /s) <sup>+</sup>	2.95 ± 0.59	3.00 ± 0.03
Frictional Ratio, $f/f_0$ <sup>+</sup>	2.71	2.68
$R_h$ (Å) <sup>+</sup>	72.7	71.3
$q.R_g$ range <sup>#</sup>	0.43–1.29	0.42–1.25
$R_g$ (Å) <sup>#</sup>	60.87 ± 0.85	59.07 ± 0.15
$I(0)$ <sup>Δ</sup>	0.01 ± 9.90 × 10 <sup>-5</sup>	0.07 ± 8.14 × 10 <sup>-5</sup>
$R_g$ (Å) <sup>Δ</sup>	61.71 ± 0.31	58.37 ± 0.07
$D_{max}$ (Å) <sup>Δ</sup>	185.0	180.7
$\chi^2$ *	~1.148	~1.084
NSD *	1.080 ± 0.024	1.005 ± 0.022

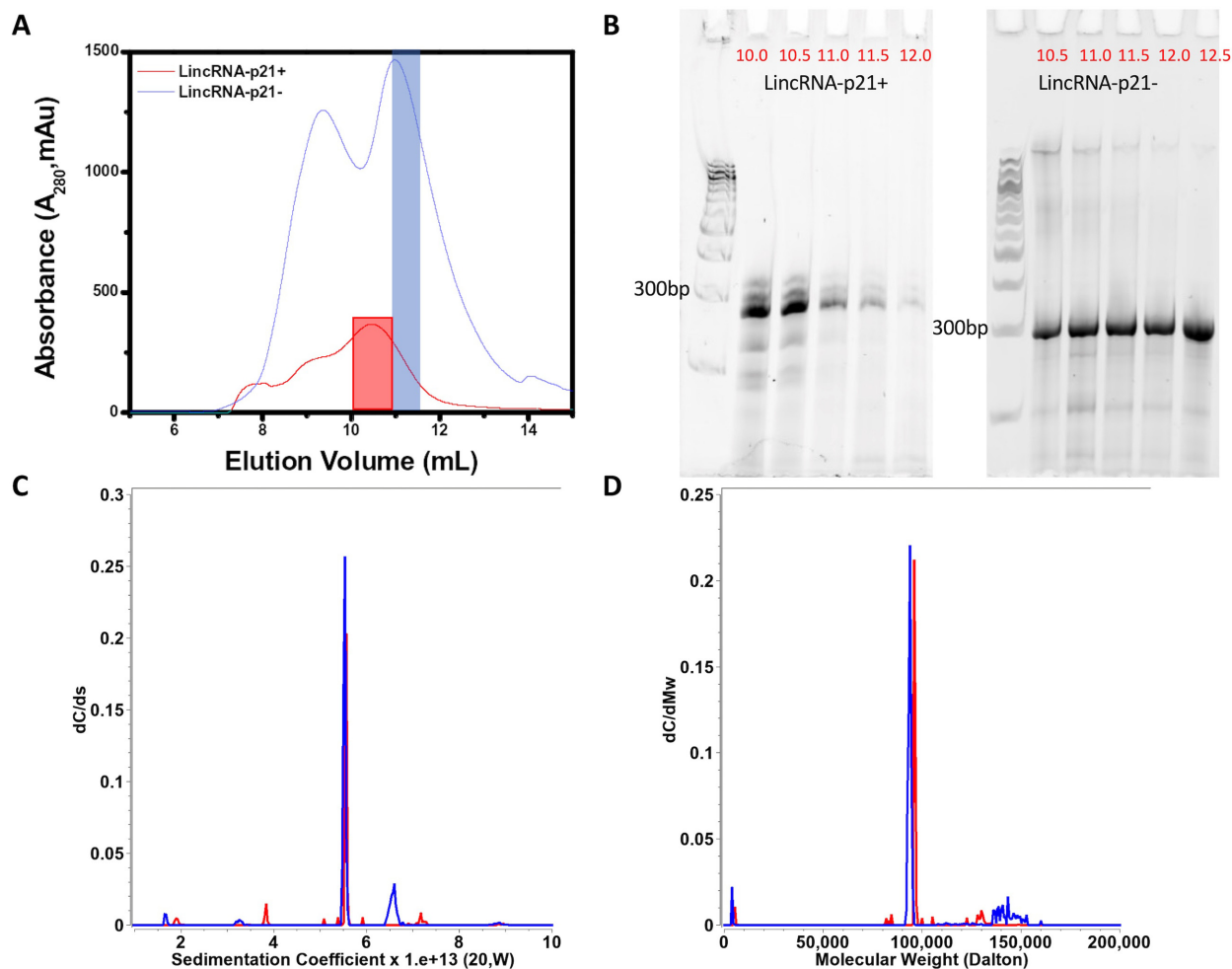
The  $M_w$  of the sense and antisense LincRNA-p21 *Alu* Repeat RNA were calculated using the nucleotide sequences provided by Chillón and Pyle, 2016. <sup>+</sup> Molar masses and frictional ratios determined by AUC assume a partial specific volume of 0.516 mL/g and refer to conditions where the RNA is denatured by 6M urea. <sup>+</sup> are within 95% confidence intervals. Data points <sup>∇</sup> were determined from SEC-MALS experiments. Data points <sup>#</sup> were derived from the Guinier analysis. Data points <sup>Δ</sup> were determined using P(r) analysis using the GNOM program. Data points \* were derived from DAMMIN and DAMAVER analysis. Terms: Hydrodynamic Radius ( $R_h$ ); Radius of Gyration ( $R_g$ ); Maximum Particle Dimension ( $D_{max}$ ); Normalised Spatial Discrepancy (NSD).

concentration) and antisense (75% of total concentration) LincRNA-p21 *AluSx1* RNA. Frictional ratios and hydrodynamic radii derived for both molecules indicate a high anisotropy for both molecules, consistent with an unfolded and extended molecule.

SEC-MALS analysis was conducted to determine the  $M_w$  of the sense and antisense *AluSx1* RNA SAXS. LincRNA-p21 *AluSx1* RNAs were purified again on a Superdex 200 Increase GL 10/300 SEC column which produced peaks eluting between 11 – 12.5 mL (Figure 3A).  $M_w$  values of sense and antisense *AluSx1* reported from SEC-MALS are slightly higher than the molar masses calculated from their sequences except for sense LincRNA-p21 *AluSx1* which exhibited less than 0.2% difference from the theoretical  $M_w$  at 99.24 ± 0.01 kDa (Table 1).  $M_w$  uniformity throughout Figure 3B and C indicates that both RNAs are monomeric. SEC-MALS results further confirm the homogeneous composition of sense and antisense *AluSx1* RNA determined in SV-AUC. RNA degradation or shorter transcripts can be excluded since no significant smaller fragments were detected. Hydrodynamic parameters derived from SEC-MALS, and SV-AUC are summarised in Table 1.

### Low-Resolution structural studies of LincRNA-p21 sense and antisense *AluSx1* inverted repeats

SAXS is a powerful method that can represent the overall solution shape of biomolecules under physiologically relevant conditions. Using SEC-SAXS, which can separate different species according to their size before being applied to the SAXS measuring cell, provides confidence in the



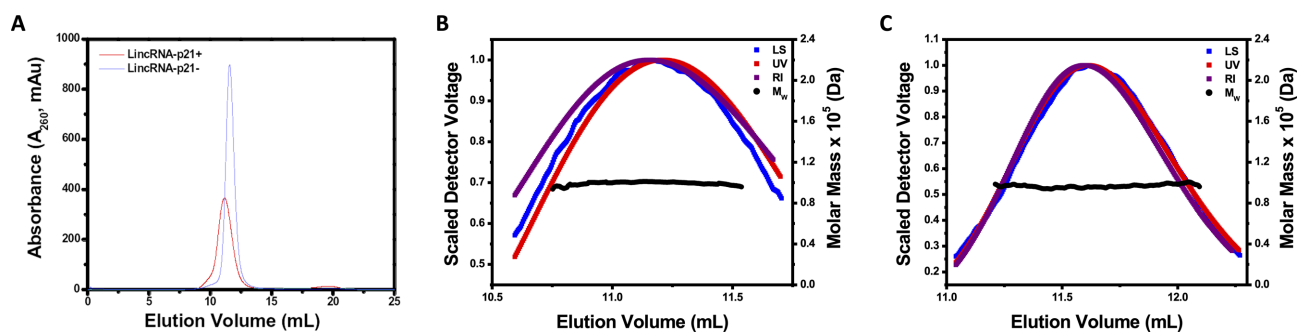
**Figure 2.** Purification of Sense and Antisense *in vitro* Transcribed LincRNA-p21 *AluSx1* RNA. (A) depicts the size exclusion chromatogram of the sense and antisense *AluSx1* RNA elution profile using the Superdex 200 Increase GL 10/300 column. SEC-MALS and SV-AUC experiments were performed with the fractions highlighted in red (sense) and blue (antisense). (B) shows the 10% urea PAGE gel used to ascertain the sense and antisense LincRNA-p21 RNA purity extracted using 0.5 mL fractions (volumes in red) using an ÄKTA Pure FPLC through a Superdex 200 Increase GL 10/300 SEC column. Fractions collected at 11.0 mL and 11.5 mL for sense and antisense *AluSx1* purifications were consolidated and used for SAXS and SV-AUC experiments. A Quick-Load® Purple 100 bp DNA Ladder (NEB, Canada) was used for the 10% urea PAGE gels in lanes 1 and 7 of each gel. (C) dC/ds sedimentation coefficient distributions for sense (Red) and anti-sense (Blue) under 6M urea denaturing conditions. (D) same as (C), except transformed to molar mass distributions assuming a partial specific volume of 0.516 mL/g.

monodispersity of purified samples (57–60). The resulting datasets were merged and presented in Figure 4A depicting the scattering intensity relative to angle for sense and antisense *AluSx1* RNA. A Guinier analysis ( $I(q)$  vs.  $(q^2)$ ) represented by Figure 4B displays the LincRNA-p21 *AluSx1* RNA samples' purity (48). The Guinier analysis determined the Guinier  $R_g$  from the low- $q$  region as being  $60.87 \pm 0.87$  Å for sense LincRNA-p21 *AluSx1* and  $59.07 \pm 0.15$  Å for antisense LincRNA-p21 *AluSx1*. Intensity data from Figure 4A was transformed to produce a dimensionless Kratky plot (Figure 4C) to determine the LincRNA-p21 *AluSx1* RNAs' conformations in solution (61). The dimensionless Kratky plot for both the sense and antisense *AluSx1* shows a levelled-plateau which suggests them as being folded and extended in solution (62).

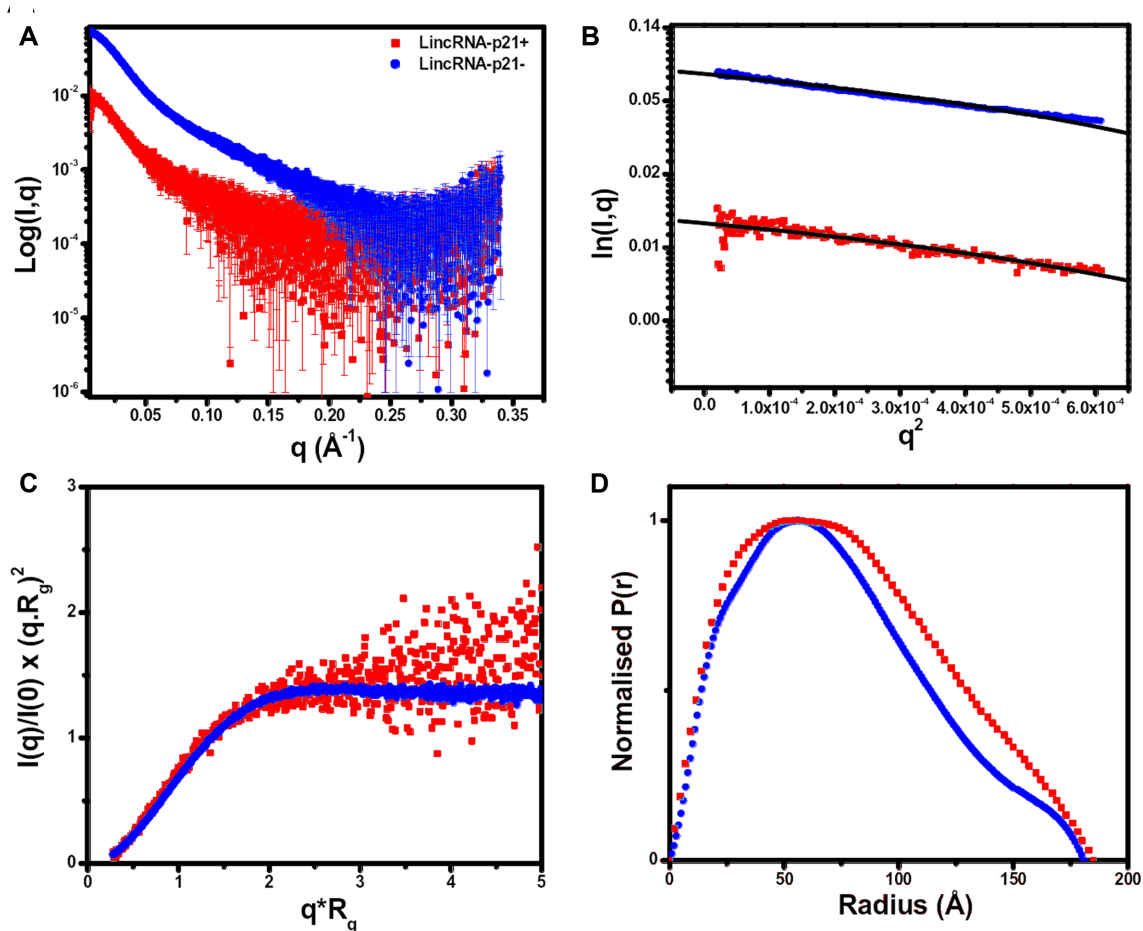
Figure 4D represents the  $P(r)$  plot which was derived from indirect Fourier transformations to convert the reciprocal-space information of the intensity data in Fig-

ure 4A to real-space electron pair distance distribution data (63). Using the  $P(r)$  plot, sense LincRNA-p21 *AluSx1* presented a real-space  $R_g$  of  $61.71 \pm 0.31$  Å and a  $D_{max}$  of 185.0 Å, while the antisense LincRNA-p21 *AluSx1* presented a real-space  $R_g$  of  $58.37 \pm 0.07$  Å and  $D_{max}$  of 180.7 Å. DAMMIN was performed to obtain low-resolution structures for the sense and antisense *AluSx1* RNA. Fifty models were calculated for each sense and antisense *AluSx1* RNAs which demonstrated favourable agreement as indicated by the  $X^2$  values ( $\sim 1.148$  for sense LincRNA-p21 *AluSx1* and  $\sim 1.084$  for antisense LincRNA-p21 *AluSx1*). DAMFILT and DAMAVER were performed to filter and averaged the models. The NSDs were estimated to be  $1.080 \pm 0.024$  and  $1.005 \pm 0.022$  for sense and antisense respectively (Table 1) (56).

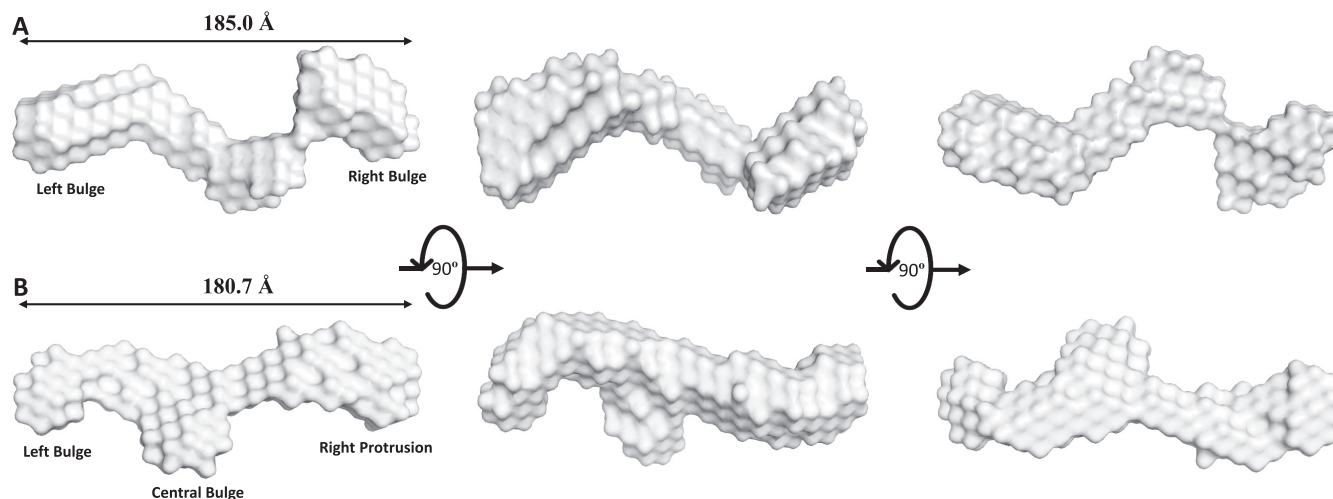
We identified two, single representative SAXS envelopes illustrated by Figure 5. The averaged, single-representative SAXS envelope of sense LincRNA-p21 *AluSx1* is gener-



**Figure 3.** Molecular Weight Determination of Sense and Antisense LincRNA-p21 *AluSx1* RNA using SEC-MALS. (A) Portrays the elution curve from the Superdex 200 Increase GL 10/300 SEC of sense (Red) and antisense (Blue) *AluSx1* RNAs. (B) Demonstrates the absolute molecular weight distribution across the elution peak of sense LincRNA-p21 *AluSx1* RNA's elution profile, and light scattering (blue), UV (red), and RI (purple) scattering. (C) Portrays the absolute molecular weight distribution across the elution peak the results fitting of antisense LincRNA-p21 *AluSx1* RNA's elution profile, and light scattering (Blue), UV (Red), and RI (Purple) scattering.



**Figure 4.** Small Angle X-Ray Scattering (SAXS) Characterisation of Sense (red) and Antisense (blue) LincRNA-p21 *AluSx1* RNA. (A) merged scattering data of sense and antisense *AluSx1* RNA depicting the scattering intensity ( $\log I(q)$ ) vs. scattering angle ( $q = 4\pi\sin\theta/\lambda$ ). (B) Guinier plots allowing for the determination of  $R_g$  from the low-angle region data and representing the homogeneity of samples. (C) Dimensionless Kratky plots ( $I(q)/I(0) \times (qR_g)^2$  vs.  $qR_g$ ) of sense and antisense *AluSx1* RNA depicting the elongated, tube-like structures because of the non-Gaussian, levelled-plateau shape of the curve. (D) Normalised pair distance distribution plots for sense and antisense *AluSx1* RNA which permits the determination of  $R_g$  derived from the SAXS dataset and including each molecule's  $D_{max}$ .



**Figure 5.** Low-Resolution Structures of Sense (A, Grey) and Antisense (B, Pale Cyan) LincRNA-p21 *AluSx1* Inverted Repeats Determined using SAXS. (A) The averaged DAMAVER SAXS low-resolution structure of sense LincRNA-p21 *AluSx1* RNA, taking on an elongated, asymmetrical, and extended structure with maximum length of 185.0 Å. Key features include a left and right Bulge. (B) The averaged DAMAVER SAXS low-resolution structure of antisense LincRNA-p21 *AluSx1* RNA, adopting an elongated, asymmetrical, and extended structure with maximum length of 180.7 Å. Key features include a left bulge, central bulge, and a right protrusion. Dimensions are represented by the  $D_{\max}$  obtained from the  $P(r)$  analysis. Models are rotated along their x-axis by  $90^\circ$  as represented by the inset.

ally extended, adopting a non-spherical, nonglobular surface model (Figure 5A). The SAXS envelope is additionally asymmetrical in its rotation along its x- and y-axes, exhibiting two prominent bulges that are primarily located on its ends. Figure 5B shows the antisense LincRNA-p21 *AluSx1* SAXS envelope which is similarly elongated and asymmetrical. Antisense LincRNA-p21 *AluSx1* though has three prominent bulges, two located centrally, while the third distally protrudes outwards from the centre.

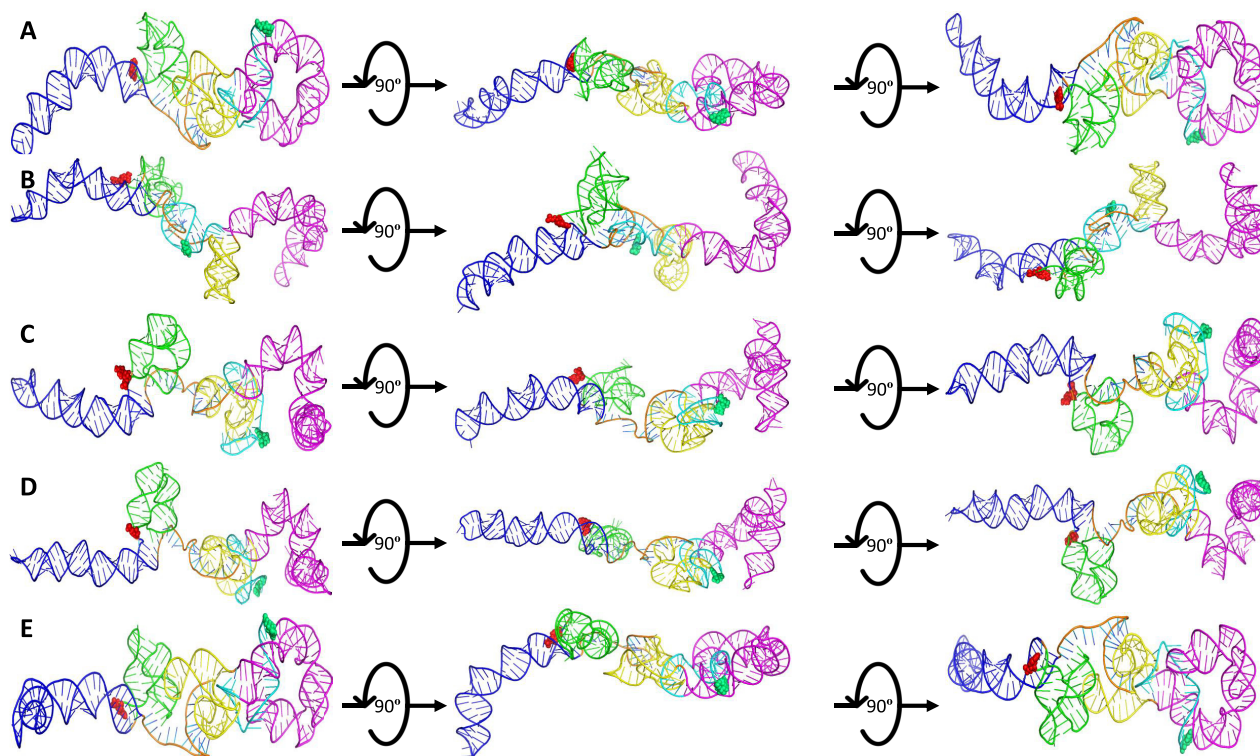
#### High-resolution atomistic models of LincRNA-p21 *AluSx1* sense and antisense inverted repeats

Using SimRNA v3.20, and the secondary structure constraints for both RNAs based on previous studies, we calculated 10,000 clusters of high-resolution, atomistic models for each RNA (15). These models were further refined through energy minimisation steps to remove models that did not satisfy constraints such as defined atom distances, bond lengths and angles. Subsequently, we obtained 3080 high-resolution models that can be superimposed on the DAMFILT SAXS envelopes using DAMSUP. Sense LincRNA-p21 *AluSx1* models when superimposed produced ten models that had an NSD range from 1.053 to 1.094, while for the antisense RNA, the top ten models retained an NSD range of 1.113 to 1.175. We further applied a selection process using the real-space  $R_g$  and  $D_{\max}$  values determined from SAXS for each molecule. We employed the program HYDROPRO to calculate biophysical properties such as  $R_g$  and  $D_{\max}$  from the 3080 high-resolution models, as performed previously (64). Top ten models were further reduced to five using the HYDROPRO properties to achieve models that were in close approximations to SAXS determined  $R_g$  and  $D_{\max}$ .

Both the top five high-resolution, high-fidelity sense and antisense *AluSx1* models are represented by Figures 6 and

7 respectively. Figure 6 depicts the sense LincRNA-p21 *AluSx1* RNAs that closely fit with the SAXS envelopes generated in Figure 5A. Previous chemically probed secondary structure predictions identified three major secondary structures: a left and right arm, and a 3'-three-way junction which have been modelled using SimRNA and represented in Figure 6. High-fidelity sense LincRNA-p21 *AluSx1* characteristically exhibits the right arm (Magenta) that curls into the central RNA body while the left arm (Blue) extends outwards. Multiple models depict variance in the right arm's position, appearing to adopt multiple conformations that curl, but rarely extend outwards, towards the central RNA body. The right arm's stem-loop on its head is additionally compacted and has either bridged with the main RNA body against the 3'-three-way junction or curls outwards. We have identified a consistent 3'-adenyl tail (Cyan) that consistently wraps around the right arm's base or the connection with the 3'-three-way junctions (Yellow). A 5'-junction (Green) is also a present feature identified but not named by the previous study, but consistently appears to project outwards, perpendicularly from the RNA's x-axis (Figure 6). A flexible, and generally unnamed region – the single-stranded linker (Orange), is presented centrally between the two arms, adopting no specific structure. An animated representation of Figure 6B is attached in Supplementary Movie S1.

Figure 7 similarly presents the high-fidelity, high-resolution structures of antisense LincRNA-p21 *AluSx1* which were modelled using previous chemically probed secondary structure predictions using SimRNA v3.20 (15). They exhibit the left and right arms, and the 5'-three-way junctions. Both the left (Blue) and right (Magenta) arms of antisense LincRNA-p21 *AluSx1* project laterally in line with the x-axis. The identified 5'-uridyl tail (Cyan) consistently wraps around the left arm. Both the 3'-three-way junctions, and the identified 5'-junction are compacted cen-



**Figure 6.** The SimRNA High-Resolution, High-Fidelity Models of Sense LincRNA-p21 *AluSx1* RNA. Figure 6 presents the high-resolution, high-fidelity sense models that have good fitting with their SAXS envelope as demonstrated by low NSD values. (A) represents model 514; (B) represents model 1036; (C) represents model 1476; (D) represents model 1677; and (E) represents model 1794 which exhibit chemically probed secondary structures: left arm (Blue); 5'-junction (Green); three-way junction (Yellow), right arm (Magenta), and the 3'-adenyl tail (Cyan). Terminal nucleotides are displayed as: 5' nt (Red, Sphere Modelled) and 3' nt (Lime Green, Sphere Modelled). Models are rotated along their x-axis by 90° as indicated by the inset. A flexible, single-stranded linker sequence is represented centrally (Orange).

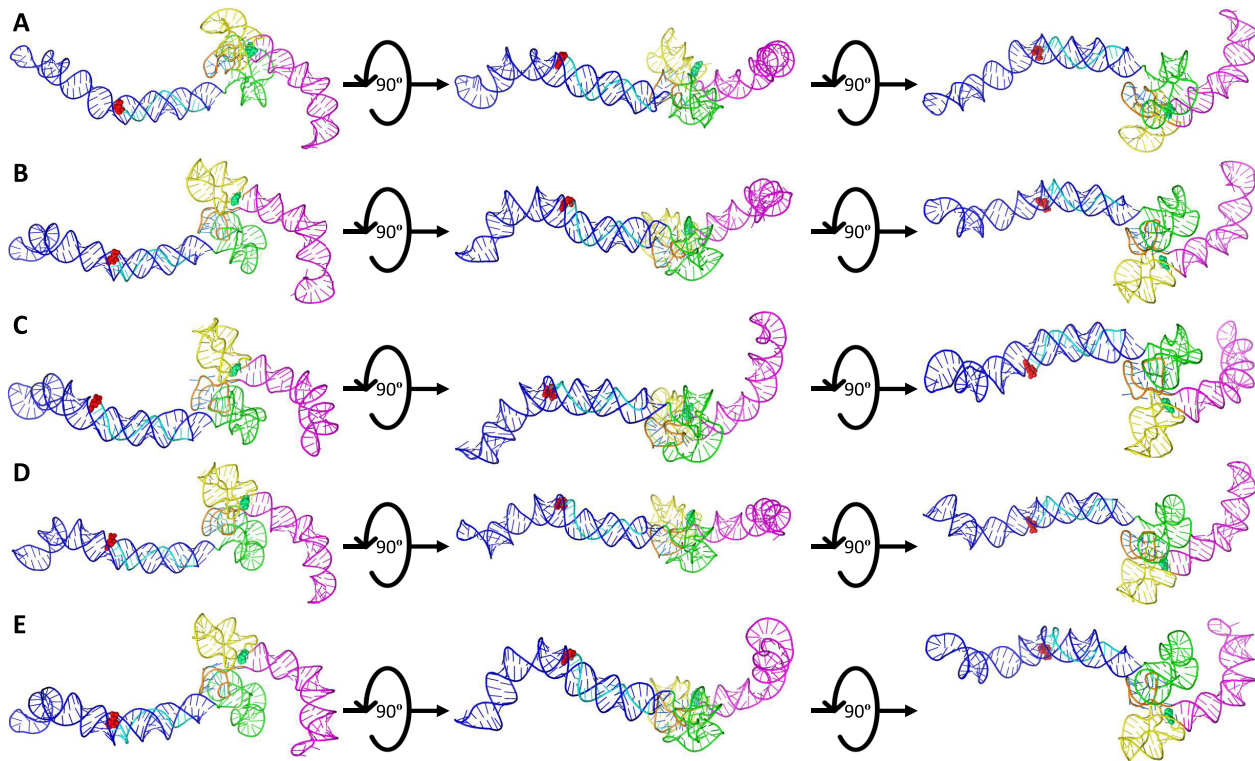
trally, with regions that project perpendicularly from the RNA body's x-axis. The right arm of antisense LincRNA-p21 *AluSx1* does not retain the characteristic stem-loop head that sense LincRNA-p21 *AluSx1* has. An animated representation of Figure 7C is attached in Supplementary Movie S2.

After performing DAMSUP, the high-fidelity, high-resolution models were visually inspected in terms of their alignment with the low-resolution SAXS envelope from DAMFILT and were represented in Figures 8 and 9 for sense and antisense *AluSx1* RNA respectively. Figure 8 displays a general agreement with the sense SAXS envelope overlaid by their high-fidelity, high resolution SimRNA models. Both the left and right arms when superimposed fit neatly within the protruding two bulges of the SAXS envelope. Minor disagreement occurs for the right arm's stem-loop head which appears to be excluded from the SAXS structure, which is similar for the tip of the Left arm. Both the 5'-junction and 3'-three-way junction exhibits considerable overlap with the SAXS low-resolution structure when superimposed. Figure 9 shows a general agreement with the antisense SimRNA models with their respective SAXS low-resolution structures. The left arm, 5'-junction, and 3'-three-way junctions are secondary structures that exhibit the highest relative overlap with the SAXS envelope and occur within the central bulge and right-most protrusion. The right arm (Magenta) depicts relatively lower agree-

ment, with its tip and core regions exposed and externalised from the SAXS envelope. This is primarily confined to the left bulge, however, which indicates a lack of adequate fitting. Two animated representations of one of the sense (Figure 8B) and antisense (Figure 9C) models are attached in Supplementary Movie S3 and S4 respectively for demonstrating the lowest NSD values 1.073 and 1.156 respectively of the five selected models.

## DISCUSSION

RNA secondary structure is especially important in defining an RNA molecule's roles and functions (65,66). These structures can be studied using a variety of techniques including secondary structure probing methods such as SHAPE, Nuclear Magnetic Resonance (NMR) spectroscopy, comparative sequence alignment and analysis, or three-dimensional structure predictions software for conceptualising higher order complexes (67–71). Identifying functional structural elements of RNA is especially important when observing the noncoding elements of the genome, whose significant contribution of noncoding RNA (ncRNA) play important regulatory roles in complex organisms. LincRNA-p21 is one such important regulatory lincRNA that is directly targeted by p53 in response to DNA damage (14). Being a transcriptional repressor in the p53 pathway, the process of understanding LincRNA-p21's sec-



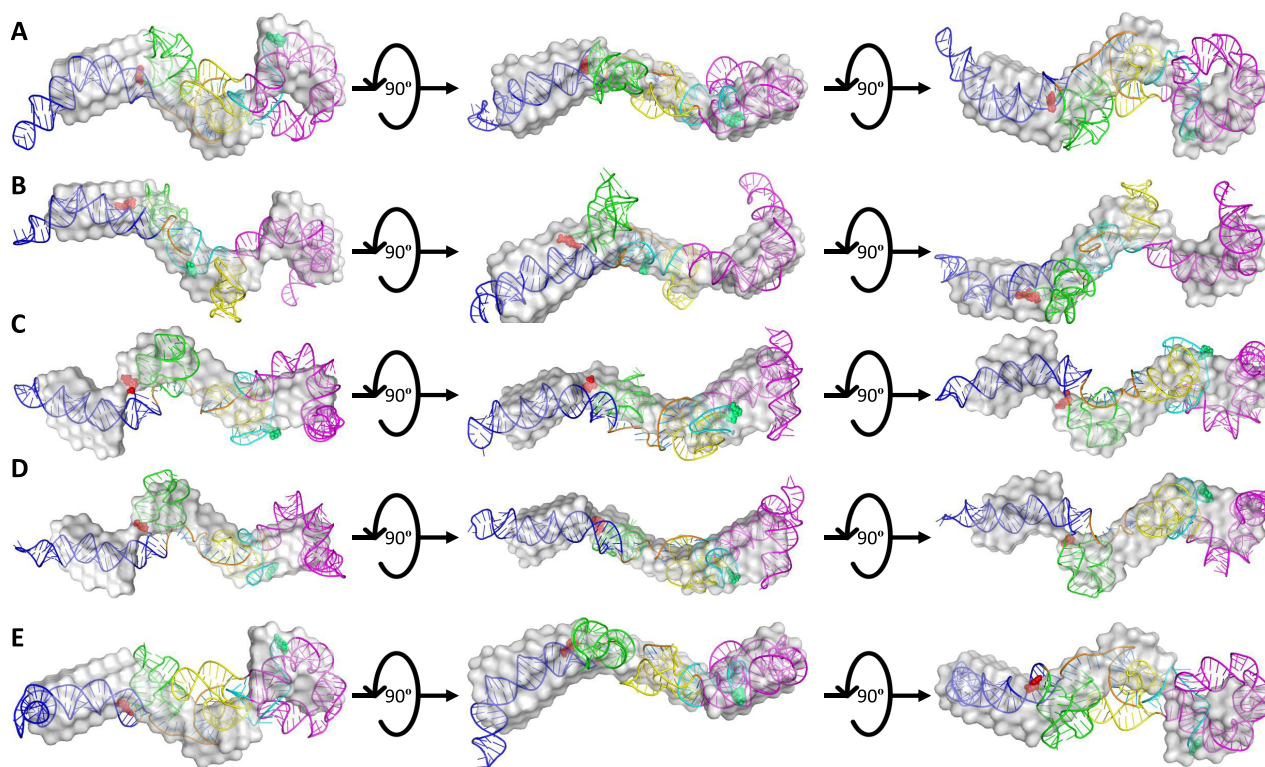
**Figure 7.** The SimRNA High-Resolution, High-Fidelity Models of Antisense LincRNA-p21 *AluSx1* RNA. Figure 7 presents the high-resolution, high-fidelity sense models that have good fitting with their SAXS envelope as demonstrated by low NSD values. (A) represents model 66; (B) represents model 974; (C) represents model 1013; (D) represents model 1074; and (E) represents model 1417 which exhibit chemically probed secondary structures: left arm (Blue); 5'-junction (Green); three-way junction (Yellow), right arm (Magenta), and the 5'-uridylyl tail (Cyan). Terminal nucleotides are displayed as: 5' nt (Red, Sphere Modelled) and 3' nt (Lime Green, Sphere Modelled). Models are rotated along their x-axis by 90° as indicated by the inset. A flexible, single-stranded linker sequence is represented centrally (Orange).

ondary structure is a first step evaluation towards how it binds to and modulates hnRNP-K localisation which is a process that ultimately triggers apoptosis in DNA damaged or cancerous cells.

The primary intent of our work is to obtain the high-resolution details of LincRNA-p21 *AluSx1* RNA by combining the previously determined secondary structure of the Inverted Repeat transposable elements determined by Chillón and Pyle with SAXS structure determination and computational modelling (15). The previous study observed that the two isoforms of hLincRNA-p21 contain IR *Alu* elements, with the sense *AluSx1* located at positions 2589–2895 and the antisense *AluSx1* located at positions 1351–1651 of the TP53COR1 gene (15). Since these RNAs are transcribed bidirectionally, observation of the secondary structure details of both the sense and antisense *AluSx1* was studied. Figure 1 outlines our multifaceted process for biophysically characterising molecules using SEC-SAXS, SEC-MALS, SV-AUC, and SimRNA from chemically probed secondary structure determinations. SEC-MALS and AUC act as orthogonal quality control and validation techniques when combined with PAGE (72).

We initially started by transcribing sense and antisense *AluSx1* RNAs, which were purified to homogeneity as indicated by Figure 2A. Both RNA displays multi-modal peaks with a primary peak representative of the monomeric RNA species. Fractions along the primary peak were taken and

analysed by 6M urea PAGE, which showed relatively pure and pronounced RNA bands around 300 bp indicating that most of the species is the primary RNA of interest (Figure 2B). Subsequent urea PAGES were performed, and each resulted in an inconclusive answer to RNA homogeneity. We therefore turned to SV-AUC to determine if the RNA samples were heterogeneous as indicated by urea-PAGE, or homogenous enough for further characterisation. As indicated in Figure 2C, D, both sense and antisense *AluSx1* RNA adopted a single, major species in solution when denatured by 6M urea. Any minor species had negligible concentrations indicated by the minor noise contributions. The SV-AUC experiments showed no evidence for RNA degradation or aggregation. The  $M_w$  values of sense and antisense *AluSx1* RNAs determined by AUC were less than 5% different (Table 1). AUC experiments were carried out in the presence of 6M urea as a denaturant, so we needed to assess each RNA in non-denaturing conditions. We therefore utilised SEC-MALS to determine absolute molecular weight in solution (34,73). Each RNA eluted as a singular, tight, and gaussian distribution elution profile evident in Figure 3A. Comparatively to AUC results, the RNA also exists monomerically in HFB as indicated by their uniform  $M_w$  demonstrated in Figures 3B/C. The sense RNA molecular weight resulted in  $99.24 \pm 0.01$ , virtually identical to the theoretical molecular weight of 99.42 kDa. Antisense RNA also resulted in a similar molecular weight of



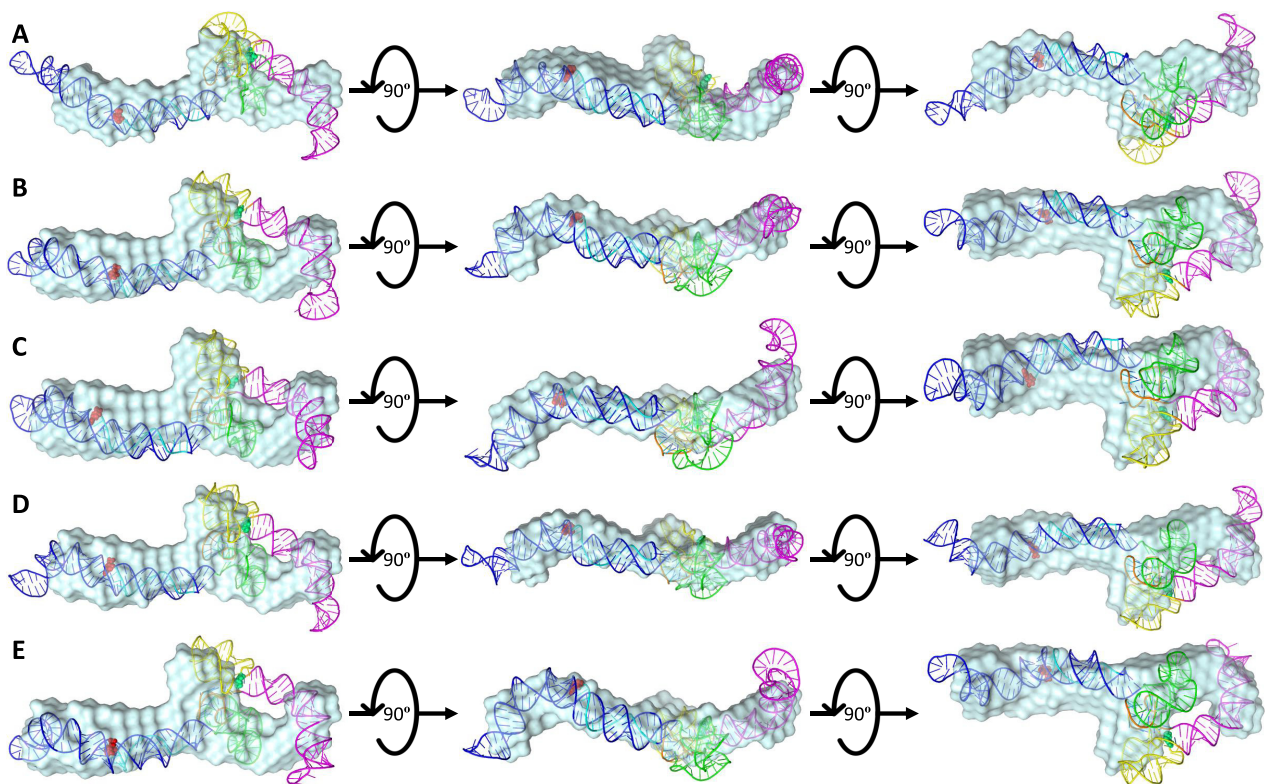
**Figure 8.** Superimposed Overlays of Sense SAXS Envelopes with their High-Resolution, High-Fidelity SimRNA Models. Figure 8 represents the combined overlays of the sense LincRNA-p21 *AluSx1* RNA envelope with their top high-resolution, SimRNA models: (A) 514; (B) 1036; (C) 1476; (D) 1677; and (E) 1794. Overall, SAXS envelopes indicate a general agreement with computationally generated structures, showing high overlap of the extended molecule with the major secondary structures identified using chemical probing techniques.

$94.52 \pm 3.71$  to the theoretical value of 89.54 kDa. These results suggest the RNA was acceptable to undergo SEC-SAXS and three-dimensional structure determination since SAXS requires highly homogenous samples.

SAXS is an ideal technique to determine low-resolution, three-dimensional structures of molecules in their native state and was used to investigate the solution conformations of both RNAs (52,74,75). SAXS data is derived from the elastic scattering and detection of high-energy photons from the electrons of an irradiated sample which can be used to examine the sample's electron density and overall structure (74). The process of SAXS involves a solubilised sample exposed to a collimated monochromatic X-ray beam which scatters as it comes in contact with the sample's molecules in question (75). The resulting scattered X-rays can be measured based upon their intensity and altered angle versus incident and scattered beams. The solvent is measured independently to be subtracted as background noise and to remove interfering signals. Data analysis follows *ab initio* shape determination which takes the scattering pattern and represents them by finite volume elements (beads). Monte-Carlo type algorithms are employed to fit the experimental data at first using random configurations. DAMMIN and DAMMIF are *ab initio* programs that can be utilised to develop compact and interconnected bead models that illustrate the overall structure of the RNA in question (52,76). SAXS sampling is relatively easy and merely includes the purification of a sample (using SEC or filtering) and its sub-

sequent exposure to an X-ray source. Such samples have the benefit of being observed in their near-physiological conditions without having to undergo artificial adjustments such as complex labelling, crystallisation, or cryogenic freezing. SAXS, however, has some limitations, namely its lower resolution compared to Cryo-Electron Microscopy or X-ray Crystallography causing ambiguity in discerning structural elements from scattering profiles (77). Additionally, X-rays can damage samples but can be minimised with buffer protectants like glycerol (78).

Overall, scattering intensities were acceptable even in low-angle regions as shown in Figure 4A. The linear regression of the sense and antisense *AluSx1* samples in Figure 4B portrays the intensities within the defined low- $q^2$  range as being linear with the absence of upward curves, illustrating monodispersity and the absence of attractive or repulsive interactions between scatterers (62,79,80). Guiner  $R_g$  approximation resulted in  $60.87 \pm 0.85$  and  $59.07 \pm 0.15$  Å for sense and antisense *AluSx1* RNAs, respectively. These  $R_g$  approximations are consistent with an elongated RNA molecule (81,82). Furthermore, the relative foldedness of sense and antisense RNAs can be deduced from the dimensionless Kratky plot in Figure 4C. The Kratky plot depicts both sense and antisense RNAs as being elongated but relatively folded, like other ncRNA (29,83). Additionally, the Guiner  $R_g$  and the real-space  $R_g$  are close with less than 1.5% difference. Figure 4D represents both RNA's distance distribution functions and are non-Gaussian, further



**Figure 9.** Superimposed Overlays of Antisense SAXS Envelopes with their High-Resolution, High-Fidelity SimRNA Models. Figure 9 represents the combined overlays of the antisense LincRNA-p21 *AluSx1* RNA envelope with their top high-resolution, SimRNA models: (A) 66; (B) 974; (C) 1013; (D) 1074; and (E) 1417. Overall, SAXS envelopes indicate a general agreement with computationally generated structures, showing high overlap of the extended molecule with the major secondary structures identified using chemical probing techniques. However, there is a slight overhang present with the right arm (Magenta) which has an area excluded from overlapping with the SAXS envelope.

consistent with extended molecules. Globular molecules will generate a Gaussian-like  $P(r)$  distributions which is not demonstrated in Figure 4D further justifying its extended shape (63). Using the  $P(r)$  distribution, molecules can also be described on their overall shape and symmetry to confirm solution folding acquired from the dimensionless Kratky plot (63). Generally, globular molecules will display a bell-shaped curve with a maximum at approximately  $D_{\max}/2$  while elongated molecules retain non-Gaussian, asymmetrical distributions with a maximum at smaller distances which appear as shoulders. For elongated molecules, this distribution will correspond to the radius of the cross section which will generally be illustrated by a tailing of the profile at larger distances (29,84). This is present for both sense and antisense *AluSx1* RNA whereby their respective  $D_{\max}/2$  do not demonstrably produce an even bell-curved but rather lead into right-leaning tails reinforcing the overall elongated shape established by Figure 4C. Sense and antisense adopt maximum distance towards 185.0 Å and 180.7 Å, respectively. Both RNAs result in similar  $D_{\max}$  measurements which is expected of two RNA of similar lengths, with the sense RNA being larger than the antisense RNA. These size difference can be attributed both to the RNAs' folding properties, and the 307 nt vs 280 nt lengths of the sense and antisense sequences as derived from Chillón and Pyle (15).

SAXS analysis of both the sense and antisense RNAs provided structures with noticeably consistent features, including a left and right bulge for the sense RNA and a left and central bulge, and right protrusion for the antisense RNA (Figure 5). Use of the standard HFB was important since human LincRNA-p21 folds at near physiological concentrations at around 5 mM  $MgCl_2$  with maximum compaction at 15 mM  $MgCl_2$  (15). These features are observed to overlap with predicted features seen in the secondary structure analysis by Chillion and Pyle (15). These features, while they can be seen in the SAXS structures, were arbitrarily based on structure orientation because directionality cannot be determined from these models. Therefore, we needed to not only computationally derive higher resolution models for clarity, but also directionality.

When analysing the high-resolution computational models derived from the secondary structure determinations, a straitened selection process was needed to screen for models that match the experimentally determined low-resolution structures. Therefore, HYDROPRO was employed to compute the hydrodynamic properties of sense and antisense *AluSx1* rigid macromolecules from their atomic-level structure (85–88). HYDROPRO's calculation comprise the basic hydrodynamic properties including the translational diffusion coefficient, sedimentation coefficient, intrinsic viscosity, and relaxation times, and can additionally provide the

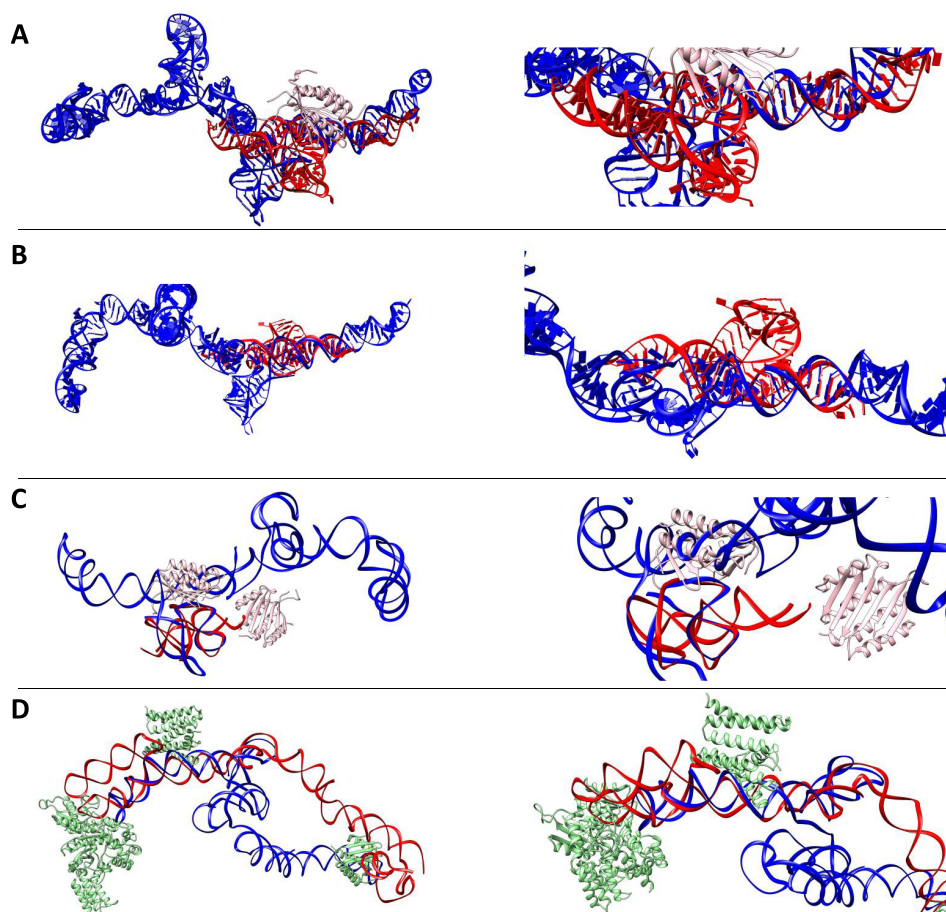
radius of gyration (54). Incidentally, HYDROPRO can be used as another orthogonal selection process against computationally generated models that don't fit the solutions scattering data. It was employed to minimise the ~3080 models which could subsequently be fit using DAMSUP based on the solution scattering range of  $D_{\max}$  and  $R_g$ . Consequently, observing HYDROPRO merely as a fitting tool narrowed the selection of potential models to five.

As detailed in Figures 6 and 7, sense and antisense *AluSx1* RNA folds into a double-stranded RNA molecule that appear to be consistent with Chillón and Pyle's secondary structure determinations. No previous attempts at secondary structure determination of *AluSx1* has been performed before Chillón and Pyle, although alternative techniques such as *in vivo* click (ic) SHAPE has been suggested as a technique to study the secondary structure of LncRNAs natively outside of the *in vitro* techniques applied (89–91). Human LincRNA-p21 is itself a linear, single exon lncRNA which contains IR Alu repeats such as *AluSx1* (15). Both the sense and antisense *AluSx1* RNAs were determined to comprise a left and right Arm with both arms connected by a single-stranded region. The 5' domain of each arm is characterised by a central three-way junction while the 3' domain is characterised by a long stem-loop. These subsequently form independent structural domains which contribute to a variety of core functions such as human LincRNA-p21's nuclear localisation following cell stress and DNA damage events (14). They suggested that the 5'-end of LincRNA-p21 interacts with hnRNP-K which could also regulate its nuclear localisation in conjunction with its *AluSx1* inverted repeat elements.

Our combined high- and low-resolution, three-dimensional structure of sense LincRNA-p21 *AluSx1* RNA (Figure 8) exhibits what appears to be two dsRNA arms that fold helically into the otherwise elongated, but radially compact or cylindrical model (Supplementary Movies S1 and S3). When overlaid with the SAXS envelope, considerable overlap is present between the high-fidelity, high-resolution models (Supplementary Movies S2 and S4). Both the 5'-junction and 3'-three-way junctions are tightly localised in the centre of the RNA body, while the left and right arms branch outwards. This likely forms a binding pocket or coordination site for hnRNP-K's one of three KH RNA/DNA binding domains which are involved in eliciting transcription regulation in the nucleus (92–94). These domains are responsible for nucleic acid binding and regulation of multiple biological processes including interaction with LincRNA-p21 for cancer suppression (94–97). The third KH domain of hnRNP-K KH3 has been shown to bind to nucleic acids as an isolated domain, although, exhibiting lower affinity than the full-length protein (98). These previous studies have observed that the KH3 domains bind to TCCC or CCCC-rich as well as CCTC C/T-rich motifs (98,99). Two TCCC motifs are present in the sense LincRNA-p21 *AluSx1* RNA in both the 5'-junction and 3'-three-way junctions. Antisense LincRNA-p21 *AluSx1* RNA is represented as a dsRNA molecule with both the 5'-junction and 3'-three-way junctions being compressed around the central bulge of the SAXS envelope (Figure 9). The RNA molecule itself is mostly extended with the left and right arms projecting

outward which likely contributes towards interacting with the hnRNP-K RNA binding domains. Specifically, sequences enriched with Alu repeats that contained three stretches of at least six pyrimidines (C/T) and matching the consensus sequence RCCTCCC (R = A/G) derived from a SINE-derived nuclear RNA Localisation (SIRLOIN) element is found to interact with hnRNP-K to increase lncRNA nuclear enrichment (93,100). Three are apparent within the antisense *AluSx1*, with two on each junction and one on the left arm suggesting potential interaction sites with hnRNP-K's three KH RNA-binding domains. This has been shown to be reliant on triplets of C/T regions (101). The previous study has identified that hnRNP-K binds to the sense, not the antisense, of LincRNA-p21 and has identified a 780 nts region on the 5'-end of the LincRNA-p21 that is required for its interaction with hnRNP-K (14). Additional studies also involved deletions of Alu IRs from LincRNA-p21 which have identified that full or partial deletions caused a 2~fold decrease in the localisation of LincRNA-p21-hnRNP-K complexes (15). This was statistically significant, although hnRNP-K localisation in the nucleus still occurred to a lesser extent which is hypothesised to be the result of the LincRNA-p21's 5'-end interacting with hnRNP-K (15). Employing our three-dimensional structures as a starting framework, future studies can investigate the effects of deleting the arms or introducing point mutations which can also include additional computational techniques such as molecular docking software to visualise the LincRNA-p21-hnRNP-K interaction (102). While these mutations and deletions may not change the structure dramatically, our structure clearly demonstrates the presence of RCCTCC consensus sequences that provide easy access to hnRNP-K interaction. Mutating said sequence might affect the interfaces as targets for structure-function interactions with hnRNP-K.

The flexibility and tendency towards being unfolded for single-stranded regions creates difficulty in visualising an RNA structure absolutely which would require additional SAXS optimisation methods to represent (108,109). Despite this, SAXS still remains the primary choice for visualising lncRNA, which can be further supplemented by atomic force microscopy (AFM) since they are too flexible for X-ray crystallography (110). Consequently, the impact of ssRNA regions being flexible is limited to the 5'-adenyl/3'-uridyl tails and central linker, with the former generally wrapping around its most proximate arm. Despite this, Figures 6 and 7 show multiple high-resolution models which were available within the constraints set by the solution scattering data. Observing Supplementary Movie S5, portraying the superimposed sense *AluSx1* models, we identified that the five models generally overlapped at the left arm, the 5'-junctions, and 3'-three-way junctions exhibiting minor conformational differences that manifested largely at flex points in the left arm. Major differences occurred in the right arm whose inward folding deviated between each individual models. This lowered confidence into an absolute, single-representative model for the sense *AluSx1* structure. Supplementary Movie S6 similarly shows the superimposed antisense *AluSx1* models, however, they exhibited clear visible overlap and uniformity throughout the four major secondary structure elements. This indicates mini-



**Figure 10.** Superimposed Overlays of Alu Crystal Fragments with High-Resolution LincRNA-p21 Sense and Antisense Models. Left panels depict the overall structures and their overlays; right panels present the same but magnified view. (A) represents the crystal structure of archaeal (*P. horikoshii*) Alu RNA (Red, 4UYK) bound to the Signal Recognition Particle (SRP, 9 kDa) protein (Pink) and overlaid with the high-resolution 1036 model of sense LincRNA-p21 *AluSx1* (Blue) (103) (see Supplementary Movie S7). (B) presents the crystal structure of Alu RNA derived from *B. subtilis* (Red, 4WFL), overlaid with the high-resolution 1036 model of sense LincRNA-p21 *AluSx1* (Blue) (104) (see Supplementary Movie S8). (C) shows the overlap of the high-resolution LincRNA-p21 *AluSx1* 1036 model (Blue) with the high-resolution structure of human Alu RNA fragment (Red, 5AOX) bound to the SRP protein (Pink) (105) (see Supplementary Movie S9). (D) presents the high-resolution antisense LincRNA-p21 *AluSx1* 1013 model (Red) overlaid with the crystal structure of canine 7S RNA (Blue, 4UE5) bound to SRP protein (Green) (106) (see Supplementary Movie S10). Each Alu crystal fragment generally depicts strong alignment and conservation of secondary structure with the sense LincRNA-p21 *AluSx1* model. RNAalign models were visualised by UCSF Chimera (107).

mal conformational differences, with the models being relatively indistinguishable from each other suggesting that all five models are equally likely given their close  $R_g$  and  $D_{max}$  parameters to the solution scattering data. Agreement of solution structure and computational models for sense LincRNA-p21 *AluSx1* is evident by high overlaps and tight fitting. However, tangible exclusions are present for antisense LincRNA-p21 *AluSx1* RNA caused by a lack of overlap with the right arm and the tip of the left arm. This can potentially be attributed to the difference in length between the solution scattering sequence (280 nts) and the secondary structure prediction sequence (301 nts). The difference can be explained by construct synthesis whose sequence lacked 41 nts at the 3'-end but encompassed the remaining 80% of the secondary structure sequence predicted which still comprises the core and majority of the SAXS structure. Nevertheless, both structures detail the presence of the previously identified secondary structures and their orientations. De-

spite this, the best fitting models that uniquely overlapped with their solution scattering three-dimensional structures are Figures 8B and 9C for sense and antisense *AluSx1* due to their low NSD values. Altogether, the models are not an ideal, true fits which would constitute a zero-NSD value and is largely the result of systematic differences (56,111). The discernible difference in antisense *AluSx1* is accounted by the deleted 3'-end while overall differences can be accounted to SAXS bead to atom approximations within SUPCOMB itself.

Previous studies on truncated domains of Alu elements show a substantial conserved structure despite varying sequence composition and origin (112). Bearing this in mind we decided to compare our structure with already reported structures of other Alu elements. We selected crystal structures 5AOX (112), 4UYK (103) and 4WFL (104) and aligned with the representative structure 1036 for sense LincRNA-p21 *AluSx1* using the program RNAalign (Fig-

ure 10, Supplementary Movies S7–S10) (113). This program aligns two RNA molecules based on their three-dimensional structure and not sequence. An ideal RMSD of 1 isn't possible given the lack of a complete crystal structure, however, the stretches of RNA present from the crystal structures strongly align with the sense 1036 model identified to have low NSD scores with respect to their SAXS envelope. A primary observation made during the overlay process was that all Alu elements had a Y-shaped structure, comprised of three helices centred around a junction. Additionally, two of the three RNA structures (4UYK and 5AOX) are in complex with the RNA binding protein, Signal Recognition Particle (SRP), and it can be clearly observed that the binding of an interacting protein takes place in the grooves generated between helices of RNA (Figure 10A and C, Supplementary Movies S7 and S9). These observations gave us confidence that the structure of LincRNA-p21 sense identified through SAXS and *in silico* techniques was similar to independently transcribed Alu elements. A similar analysis was done for the high-resolution LincRNA-p21 antisense 1013 model, where we compared it with the PDB structure 4UE5 (Figure 10D, Supplementary Movie S10). In this case, we observed that only the partial antisense LincRNA-p21 *AluSx1* structure resembles the reported structures in Figure 9. We could conclude that using our technique one can reliably reconstruct the high-resolution, three-dimensional structure of LincRNA by employing SHAPE secondary structure information, SAXS data, and SimRNA (Figure 1).

Applying both solution scattering techniques and coarse-grained computational modelling reveals that LincRNA-p21 Alu Inverted Repeats do not adopt completely stable, single-representative structures due in part to the conformational flexibility present in their respective DAMMIN and SimRNA models. Both the sense and antisense *AluSx1* RNAs adopt multiple conformations, however, they closely approximate into a generally similar, single-representative structure with mild conformational and structural differences when averaged. The main shape – one that is elongated, asymmetrical, with regions that encompass the main left and right arms, and three-way junction – is uniform and preserved throughout solution scattering and computational fitting techniques. The wholesale and complete LincRNA-p21 three-dimensional structure determination from SAXS isn't entirely necessary, since combining SAXS structures from fragments and piecing them together using the divide and conquer approach has shown conservation in overall structure as indicated in previous protein and RNA studies (109,114–116). Biological structure-function relationships have been determined from this starting methodology (117–120). Consequently, applying a combination of SAXS and computationally generated tertiary structure models concertedly determined appropriate representations of the LincRNA-p21 *AluSx1* Inverted Repeats.

## CONCLUSION

We have demonstrated that chemical probing techniques involved in RNA secondary structure predictions such as SHAPE can be combined using a multifaceted biophysical

approach involving SAXS, AUC, SEC-MALS, and computational RNA modelling. Although generally averaged, the three-dimensional structures of sense and antisense LincRNA-p21 *AluSx1* IRs provided from Supplementary Movie S3 and S4 respectively represent a current understanding of their overall structure. Naturally, higher resolution structure details can be achieved using NMR and single particle Cryo-Electron Microscopy methods, employing similar computational techniques to acquire as low as 3.7 Å resolution on very small (~100 nt) RNA (71,121,122). This would be to expand on structure correlation between crystal or solution scattering methods and computational ones to overcome the heterogeneous and flexible nature of RNAs. Nevertheless, three-dimensional models are important in confirming secondary structure motifs that are predicted through probing techniques. By expanding the structure to include three-dimensional native state folding, essential regulatory RNAs involved in apoptosis and tumour suppression in cancer cells can be effectively visualised to identify functional domains and potential RNA-Protein binding regions.

## DATA AVAILABILITY

SAXS data and fits have been deposited at SASBDB under the accession codes of 3663 and 3664 for sense and antisense LincRNA-p21 *AluSx1* RNAs, respectively.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank all the B21 beamline scientists at DIAMOND for their continual support. Molecular graphics and analyses was performed using UCSF Chimera.

*Author contributions:* M.H.D., M.G., and A.F. prepared RNAs and purified them. M.H.D., M.D.B., and T.R.P. processed SAXS data. T.M. and M.H.D. collected and processed SEC-MALS data. A.H. collected AUC data. A.H., B.D., and M.H.D. processed AUC data. M.T.W. performed computational RNA modelling. M.T.W., M.H.D., and T.R.P. analysed computational models. M.H.D. developed initial drafts of manuscript. All authors reviewed and edited manuscript.

## FUNDING

M.H.D and M.G. are supported by the NSERC Discovery grant awarded to T.R.P. M.H.D is supported by the South Alberta Light Horse Regimental Foundation, The King's Own Calgary Regiment (RCAC) Funds Foundation, and the Canadian Armed Forces Individual Learning Plan and Self-Development Program. M.D.B. is supported by Alberta Innovates grant to T.R.P. T.M. is supported by a Natural Sciences and Engineering Research Council (NSERC) PGS-D award. A.H. is supported by the NSERC CGS-D award. T.R.P. is a Canada Research Chair in RNA and Protein Biophysics. Infrastructure support was provided by the Canada Foundation for Innovation (CFI 37155 to T.R.P.

and CFI-41008 to T.R.P. and B.D.) and NSERC RTI (RTI 2020–00090 to T.R.P. and B.D.) Grants. This work was supported by the Canada 150 Research Chairs program (B.D., C150-2017-00015), the Canada Foundation for Innovation (B.D., CFI-37589), the National Institutes of Health (B.D., 1R01GM120600) and the Canadian Natural Science and Engineering Research Council (B.D., DG-RGPIN-2019–05637). UltraScan supercomputer calculations were supported through NSF/XSEDE grant TG-MCB070039N (to B.D.). The SAXS data collection was supported by DIAMOND Synchrotron (SM26855), UK.

*Conflict of interest statement.* None declared.

## REFERENCES

- Rufini, A., Tucci, P., Celardo, I. and Melino, G. (2013) Senescence and aging: the critical roles of p53. *Oncogene*, **32**, 5129–5143.
- Amaral, J.D., Xavier, J.M., Steer, C.J. and Rodrigues, C.M. (2010) The role of p53 in apoptosis. *Discov. Med.*, **9**, 145–152.
- Treviño, V., Martínez-Ledesma, E. and Tamez-Peña, J. (2017) Identification of outcome-related driver mutations in cancer using conditional co-occurrence distributions. *Sci. Rep.*, **7**, 43350.
- Catana, C.-S., Gulei, D. and Berindan - Neagoe, I. (2017) New insights into the role of non-coding RNAs as transcriptional targets of p53. *Molecular Life*, **1**, 43–49.
- Hafner, A., Bulyk, M.L., Jambhekar, A. and Lahav, G. (2019) The multiple mechanisms that regulate p53 activity and cell fate. *Nat. Rev. Mol. Cell Biol.*, **20**, 199–210.
- Levine, A.J. (2020) p53: 800 million years of evolution and 40 years of discovery. *Nat. Rev. Cancer*, **20**, 471–480.
- Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–227.
- Jin, S., Yang, X., Li, J., Yang, W., Ma, H. and Zhang, Z. (2019) p53-targeted lincRNA-p21 acts as a tumor suppressor by inhibiting JAK2/STAT3 signaling pathways in head and neck squamous cell carcinoma. *Mol. Cancer*, **18**, 38.
- Wang, X., Ruan, Y., Wang, X., Zhao, W., Jiang, Q., Jiang, C., Zhao, Y., Xu, Y., Sun, F., Zhu, Y. *et al.* (2017) Long intragenic non-coding RNA lincRNA-p21 suppresses development of human prostate cancer. *Cell Prolif.*, **50**, e123118.
- Tran, U.M., Rajarajacholan, U., Soh, J., Kim, T.S., Thalappilly, S., Sensen, C.W. and Riabowol, K. (2015) LincRNA-p21 acts as a mediator of ING1b-induced apoptosis. *Cell Death. Dis.*, **6**, e1668.
- Chen, S., Liang, H., Yang, H., Zhou, K., Xu, L., Liu, J., Lai, B., Song, L., Luo, H., Peng, J. *et al.* (2017) LincRNA-p21: function and mechanism in cancer. *Med. Oncol.*, **34**, 98.
- Tang, S.S., Zheng, B.Y. and Xiong, X.D. (2015) LincRNA-p21: implications in human diseases. *Int. J. Mol. Sci.*, **16**, 18732–18740.
- Kesheh, M.M., Mahmoudvand, S. and Shokri, S. (2021) Long noncoding RNAs in respiratory viruses: a review. *Rev. Med. Virol.*, **32**, e2275.
- Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M. *et al.* (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, **142**, 409–419.
- Chillón, I. and Pyle, A.M. (2016) Inverted repeat alu elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function. *Nucleic Acids Res.*, **44**, 9462–9471.
- Moumen, A., Masterson, P., O'Connor, M.J. and Jackson, S.P. (2005) hnRNP K: an HDM2 target and transcriptional coactivator of p53 in response to DNA damage. *Cell*, **123**, 1065–1078.
- Sun, X., Haider Ali, M.S.S. and Moran, M. (2017) The role of interactions of long non-coding RNAs and heterogeneous nuclear ribonucleoproteins in regulating cellular functions. *Biochem. J.*, **474**, 2925–2935.
- Baldassarre, A. and Masotti, A. (2012) Long non-coding RNAs and p53 regulation. *Int. J. Mol. Sci.*, **13**, 16708–16717.
- Arcot, S.S., Wang, Z., Weber, J.L., Deininger, P.L. and Batzer, M.A. (1995) Alu repeats: a source for the genesis of primate microsatellites. *Genomics*, **29**, 136–144.
- Batzer, M.A. and Deininger, P.L. (2002) Alu repeats and human genomic diversity. *Nat. Rev. Genet.*, **3**, 370–379.
- Tajaddod, M., Tanzer, A., Licht, K., Wolfinger, M.T., Badelt, S., Huber, F., Pusch, O., Schopoff, S., Janisiw, M., Hofacker, I. *et al.* (2016) Transcriptome-wide effects of inverted SINEs on gene expression and their impact on RNA polymerase II activity. *Genome Biol.*, **17**, 220.
- Novikova, I.V., Hennelly, S.P. and Sanbonmatsu, K.Y. (2012) Sizing up long non-coding RNAs: do lncRNAs have secondary and tertiary structure? *Bioarchitecture*, **2**, 189–199.
- Chillón, I. and Marcia, M. (2020) The molecular structure of long non-coding RNAs: emerging patterns and functional implications. *Crit. Rev. Biochem. Mol. Biol.*, **55**, 662–690.
- Lusvardi, S., Sztuba-Solinska, J., Purzycka, K.J., Rausch, J.W. and Le Grice, S.F. (2013) RNA secondary structure prediction using high-throughput SHAPE. *J. Vis. Exp.*, **75**, e50243.
- Leontis, N.B., Stombaugh, J. and Westhof, E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
- Markham, N.R. and Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.*, **453**, 3–31.
- Boniecki, M.J., Lach, G., Dawson, W.K., Tomala, K., Lukasz, P., Soltysinski, T., Rother, K.M. and Bujnicki, J.M. (2016) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.*, **44**, e63.
- Conrad, T., Plumbom, I., Alcobendas, M., Vidal, R. and Sauer, S. (2020) Maximizing transcription of nucleic acids with efficient T7 promoters. *Commun. Biol.*, **3**, 439.
- Mrozowich, T., Henrickson, A., Demeler, B. and Patel, T.R. (2020) Nanoscale structure determination of murray valley encephalitis and powassan virus non-coding RNAs. *Viruses*, **12**, 190.
- Chillón, I., Marcia, M., Legiewicz, M., Liu, F., Somarowthu, S. and Pyle, A.M. (2015) Native purification and analysis of long RNAs. *Methods Enzymol.*, **558**, 3–37.
- Beckert, B. and Masquida, B. (2011) Synthesis of RNA by in vitro transcription. *Methods Mol. Biol.*, **703**, 29–41.
- Chen, Z. and Zhang, Y. (2005) Dimethyl sulfoxide targets phage RNA polymerases to promote transcription. *Biochem. Biophys. Res. Commun.*, **333**, 664–670.
- McKenna, S.A., Kim, I., Puglisi, E.V., Lindhout, D.A., Aitken, C.E., Marshall, R.A. and Puglisi, J.D. (2007) Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nat. Protoc.*, **2**, 3270–3277.
- Some, D., Amartely, H., Tsadok, A. and Lebediker, M. (2019) Characterization of proteins by size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS). *J. Vis. Exp.*, **148**, e59615.
- Pam Wang, R.A., Chen, Michelle and Legaspi, Kristine. (2020) In: *ANI1616: SEC-MALS Method for Characterizing mRNA Biophysical Attributes*. Wyatt Technologies Moderna Therapeutics.
- Pam Wang, R.A., Chen, M. and Legaspi, K. (2020) SEC-MALS method for characterizing mRNA biophysical attributes. *Wyatt Technologies Moderna Therapeutics*. pp. 1–4.
- Patel, T.R., Chojnowski, G., Astha, Koul, A., McKenna, S.A. and Bujnicki, J.M. (2017) Structural studies of RNA-protein complexes: a hybrid approach involving hydrodynamics, scattering, and computational methods. *Methods*, **118-119**, 146–162.
- Wyatt, P.J. (2014) Measurement of special nanoparticle structures by light scattering. *Anal. Chem.*, **86**, 7171–7183.
- Wyatt, P.J. (2018) Measuring nanoparticles in the size range to 2000 nm. *J. Nanopart. Res.*, **20**, 322–322.
- Demeler, B. and Gorbet, G.E. (2016) In: Uchiyama, S., Arisaka, F., Stafford, W.F. and Laue, T. (eds). *Analytical Ultracentrifugation: Instrumentation, Software, and Applications*. Springer Japan, Tokyo, pp. 119–143.
- Demeler, B. (2010) Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. *Curr. Protoc. Protein. Sci.*, **Chapter 7**, Unit 7.13.
- Brookes, E., Cao, W. and Demeler, B. (2010) A two-dimensional spectrum analysis for sedimentation velocity experiments of

- mixtures with heterogeneity in molecular weight and shape. *Eur. Biophys. J.*, **39**, 405–414.
43. Brookes, E.H. and Demeler, B. (2007) In: *Proceedings of the 9th annual conference on Genetic and evolutionary computation*. Association for Computing Machinery, London, England, pp. 361–368.
  44. Demeler, B. and Brookes, E. (2008) Monte carlo analysis of sedimentation experiments. *Colloid and Polymer Science*, **286**, 129–137.
  45. Cowieson, N.P., Edwards-Gayle, C.J.C., Inoue, K., Khunti, N.S., Douth, J., Williams, E., Daniels, S., Preece, G., Krumpa, N.A., Sutter, J.P. *et al.* (2020) Beamline B21: high-throughput small-angle X-ray scattering at diamond light source. *J. Synchrotron Radiat.*, **27**, 1438–1446.
  46. Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D.T., Gruzinov, A., Borges, C. *et al.* (2021) ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **54**, 343–355.
  47. Panjkovich, A. and Svergun, D.I. (2018) CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics*, **34**, 1944–1946.
  48. Putnam, C.D. (2016) Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data. *J. Appl. Crystallogr.*, **49**, 1412–1419.
  49. Burke, J.E. and Butcher, S.E. (2012) Nucleic acid structure characterization by small angle X-ray scattering (SAXS). *Curr. Protoc. Nucleic Acid Chem.*, **Chapter 7**, Unit7.18–Unit7.18.
  50. Semenyuk, A.V. and Svergun, D.I. (1991) GNOM— a program package for small-angle scattering data processing. *J. Appl. Crystallogr.*, **24**, 537–540.
  51. Svergun, D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 495–503.
  52. Svergun, D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J.*, **76**, 2879–2886.
  53. Volkov, V.V. and Svergun, D.I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **36**, 860–864.
  54. Ortega, A., Amorós, D. and García de la Torre, J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J.*, **101**, 892–898.
  55. Stasiewicz, J., Mukherjee, S., Nithin, C. and Bujnicki, J.M. (2019) QRNAS: software tool for refinement of nucleic acid structures. *BMC Struct. Biol.*, **19**, 5.
  56. Kozin, M.B. and Svergun, D.I. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.*, **34**, 33–41.
  57. Brosey, C.A. and Tainer, J.A. (2019) Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Curr. Opin. Struct. Biol.*, **58**, 197–213.
  58. Pérez, J. and Vachette, P. (2017) A successful combination: coupling SE-HPLC with SAXS. *Adv. Exp. Med. Biol.*, **1009**, 183–199.
  59. O'Brien, D.P., Brier, S., Ladant, D., Durand, D., Chenal, A. and Vachette, P. (2018) SEC-SAXS and HDX-MS: a powerful combination. The case of the calcium-binding domain of a bacterial toxin. *Biotechnol. Appl. Biochem.*, **65**, 62–68.
  60. Gräwert, M., Da Vela, S., Gräwert, T., Molodenskiy, D., Blanchet, C., Svergun, D. and Jeffries, C. (2020) Adding size exclusion chromatography (SEC) and light scattering (LS) devices to obtain high-quality small angle X-Ray scattering (SAXS) data. *Crystals*, **10**, 975.
  61. Rambo, R.P. and Tainer, J.A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the porod-debye law. *Biopolymers*, **95**, 559–571.
  62. Putnam, C.D., Hammel, M., Hura, G.L. and Tainer, J.A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev. Biophys.*, **40**, 191–285.
  63. Kikhney, A.G. and Svergun, D.I. (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.*, **589**, 2570–2577.
  64. Vadlamani, G., Thomas, M.D., Patel, T.R., Donald, L.J., Reeve, T.M., Stetefeld, J., Standing, K.G., Vocadlo, D.J. and Mark, B.L. (2015) The  $\beta$ -lactamase gene regulator AmpR is a tetramer that recognizes and binds the D-Ala-D-Ala motif of its repressor UDP-N-acetylmuramic acid (MurNAc)-pentapeptide. *J. Biol. Chem.*, **290**, 2630–2643.
  65. Vandivier, L.E., Anderson, S.J., Foley, S.W. and Gregory, B.D. (2016) The conservation and function of RNA secondary structure in plants. *Annu. Rev. Plant Biol.*, **67**, 463–488.
  66. Chełkowska-Pauszek, A., Kosiński, J.G., Marciniak, K., Wysocka, M., Bąkowska-Zywicka, K. and Żywicki, M. (2021) The role of RNA secondary structure in regulation of gene expression in bacteria. *Int. J. Mol. Sci.*, **22**, 7845.
  67. Lorenz, R., Wolfinger, M.T., Tanzer, A. and Hofacker, I.L. (2016) Predicting RNA secondary structures from sequence and probing data. *Methods*, **103**, 86–98.
  68. Chen, J.L., Bellaousov, S. and Turner, D.H. (2016) RNA secondary structure determination by NMR. *Methods Mol. Biol.*, **1490**, 177–186.
  69. Kenyon, J., Prestwood, L. and Lever, A. (2014) Current perspectives on RNA secondary structure probing. *Biochem. Soc. Trans.*, **42**, 1251–1255.
  70. Tahiri, F., Du, T.T.V. and Boucheham, A. (2017) In silico prediction of RNA secondary structure. *Methods Mol. Biol.*, **1543**, 145–168.
  71. Barnwal, R.P., Yang, F. and Varani, G. (2017) Applications of NMR to structure determination of RNAs large and small. *Arch. Biochem. Biophys.*, **628**, 42–56.
  72. Patel, B.A., Gospodarek, A., Larkin, M., Kenrick, S.A., Haverick, M.A., Tugcu, N., Brower, M.A. and Richardson, D.D. (2018) Multi-angle light scattering as a process analytical technology measuring real-time molecular weight for downstream process control. *MAbs*, **10**, 945–950.
  73. Stetefeld, J., McKenna, S.A. and Patel, T.R. (2016) Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.*, **8**, 409–427.
  74. Gräwert, T.W. and Svergun, D.I. (2020) Structural modeling using solution small-angle X-ray scattering (SAXS). *J. Mol. Biol.*, **432**, 3078–3092.
  75. Svergun, D.I. (2010) Small-angle X-ray and neutron scattering as a tool for structural systems biology. *Biol. Chem.*, **391**, 737–743.
  76. Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **42**, 342–346.
  77. Jeffries, C.M., Ilavsky, J., Martel, A., Hinrichs, S., Meyer, A., Pedersen, J.S., Sokolova, A.V. and Svergun, D.I. (2021) Small-angle X-ray and neutron scattering. *Nat. Rev. Methods Primers*, **1**, 70.
  78. Rambo, R.P. and Tainer, J.A. (2010) Improving small-angle X-ray scattering data for structural analyses of the RNA world. *RNA*, **16**, 638–646.
  79. Grant, T.D., Luft, J.R., Carter, L.G., Matsui, T., Weiss, T.M., Martel, A. and Snell, E.H. (2015) The accurate assessment of small-angle X-ray scattering data. *Acta Crystallogr D Biol. Crystallogr.*, **71**, 45–56.
  80. Cantara, W.A., Olson, E.D. and Musier-Forsyth, K. (2017) Analysis of RNA structure using small-angle X-ray scattering. *Methods*, **113**, 46–55.
  81. Zhang, Y., Zhang, Y., Liu, Z.Y., Cheng, M.L., Ma, J., Wang, Y., Qin, C.F. and Fang, X. (2019) Long non-coding subgenomic flavivirus RNAs have extended 3D structures and are flexible in solution. *EMBO Rep.*, **20**, e47016.
  82. Cantero-Camacho, A., Fan, L., Wang, Y.X. and Gallego, J. (2017) Three-dimensional structure of the 3'X-tail of hepatitis c virus RNA in monomeric and dimeric states. *RNA*, **23**, 1465–1476.
  83. Nelson, C.R., Mrozowich, T., Park, S.M., D'Souza, S., Henrickson, A., Vigar, J.R.J., Wieden, H.J., Owens, R.J., Demeler, B. and Patel, T.R. (2020) Human DDX17 unwinds rift valley fever virus non-coding RNAs. *Int. J. Mol. Sci.*, **22**, 54.
  84. Choi, K.H. and Morais, M. (2014) Use of small-angle X-ray scattering to investigate the structure and function of dengue virus NS3 and NS5. *Methods in Molecular Biology (Clifton, N.J.)*, **1138**, 241–252.

85. Carrasco, B. and García de la Torre, J. (1999) Hydrodynamic properties of rigid particles: comparison of different modeling and computational procedures. *Biophys J.*, **76**, 3044–3057.
86. Fernandes, M.X., Ortega, A., López Martínez, M.C. and García de la Torre, J. (2002) Calculation of hydrodynamic properties of small nucleic acids from their atomic structure. *Nucleic Acids Res.*, **30**, 1782–1788.
87. García De La Torre, J., Huertas, M.L. and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J.*, **78**, 719–730.
88. Garcia de la Torre, J.G. and Bloomfield, V.A. (1981) Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications. *Q Rev. Biophys.*, **14**, 81–139.
89. Feyder, M. and Goff, L.A. (2016) Investigating long noncoding RNAs using animal models. *J. Clin. Invest.*, **126**, 2783–2791.
90. Sahu, A., Singhal, U. and Chinnaiyan, A.M. (2015) Long noncoding RNAs in cancer: from function to translation. *Trends Cancer*, **1**, 93–109.
91. Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T. *et al.* (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, **519**, 486–490.
92. Bomsztyk, K., Denisenko, O. and Ostrowski, J. (2004) hnRNP K: one protein multiple processes. *Bioessays*, **26**, 629–638.
93. Xu, Y., Wu, W., Han, Q., Wang, Y., Li, C., Zhang, P. and Xu, H. (2019) New insights into the interplay between non-coding RNAs and RNA-Binding protein HnRNPK in regulating cellular functions. *Cells*, **8**, 62.
94. Makeyev, A.V. and Liebhaber, S.A. (2002) The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. *RNA*, **8**, 265–278.
95. Baber, J.L., Libutti, D., Levens, D. and Tjandra, N. (1999) High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein k, a c-myc transcription factor. *J. Mol. Biol.*, **289**, 949–962.
96. Nazarov, I.B., Bakhmet, E.I. and Tomilin, A.N. (2019) KH-Domain poly(c)-binding proteins as versatile regulators of multiple biological processes. *Biochemistry (Mosc)*, **84**, 205–219.
97. Wang, Z., Qiu, H., He, J., Liu, L., Xue, W., Fox, A., Tickner, J. and Xu, J. (2020) The emerging roles of hnRNPK. *J. Cell Physiol.*, **235**, 1995–2008.
98. Backe, P.H., Messias, A.C., Ravelli, R.B., Sattler, M. and Cusack, S. (2005) X-ray crystallographic and NMR studies of the third KH domain of hnRNP k in complex with single-stranded nucleic acids. *Structure*, **13**, 1055–1067.
99. Yao, J., Tu, Y., Shen, C., Zhou, Q., Xiao, H., Jia, D. and Sun, Q. (2021) Nuclear import receptors and hnRNPK mediates nuclear import and stress granule localization of SIRLOIN. *Cell Mol. Life Sci.*, **78**, 7617–7633.
100. Lubelsky, Y. and Ulitsky, I. (2018) Sequences enriched in alu repeats drive nuclear localization of long RNAs in human cells. *Nature*, **555**, 107–111.
101. Paziewska, A., Wyrwicz, L.S., Bujnicki, J.M., Bomsztyk, K. and Ostrowski, J. (2004) Cooperative binding of the hnRNP k three KH domains to mRNA targets. *FEBS Lett.*, **577**, 134–140.
102. Philip, M., Chen, T. and Tyagi, S. (2021) A survey of current resources to study lncRNA-Protein interactions. *Non-Coding RNA*, **7**, 33.
103. Bousset, L., Mary, C., Brooks, M.A., Scherrer, A., Strub, K. and Cusack, S. (2014) Crystal structure of a signal recognition particle alu domain in the elongation arrest conformation. *RNA*, **20**, 1955–1962.
104. Kempf, G., Wild, K. and Sinning, I. (2014) Structure of the complete bacterial SRP alu domain. *Nucleic Acids Res.*, **42**, 12284–12294.
105. Ahl, V., Keller, H., Schmidt, S. and Weichenrieder, O. (2015) Retrotransposition and crystal structure of an alu RNP in the ribosome-stalling conformation. *Mol. Cell*, **60**, 715–727.
106. Beckert, B., Kedrov, A., Sohmen, D., Kempf, G., Wild, K., Sinning, I., Stahlberg, H., Wilson, D.N. and Beckmann, R. (2015) Translational arrest by a prokaryotic signal recognition particle is mediated by RNA interactions. *Nat. Struct. Mol. Biol.*, **22**, 767–773.
107. Petersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
108. Meisburger, S.P., Sutton, J.L., Chen, H., Pabit, S.A., Kirmizialtin, S., Elber, R. and Pollack, L. (2013) Polyelectrolyte properties of single stranded DNA measured using SAXS and single-molecule FRET: beyond the wormlike chain model. *Biopolymers*, **99**, 1032–1045.
109. Plumridge, A., Meisburger, S.P. and Pollack, L. (2017) Visualizing single-stranded nucleic acids in solution. *Nucleic Acids Res.*, **45**, e66.
110. Sanbonmatsu, K. (2021) Getting to the bottom of lncRNA mechanism: structure-function relationships. *Mamm Genome*, 1–11.
111. Konarev, P.V., Petoukhov, M.V. and Svergun, D.I. (2016) Rapid automated superposition of shapes and macromolecular models using spherical harmonics. *J. Appl. Crystallogr.*, **49**, 953–960.
112. Soni, K., Kempf, G., Manalastas-Cantos, K., Hendricks, A., Flemming, D., Guizzetti, J., Simon, B., Frischknecht, F., Svergun, D.I., Wild, K. *et al.* (2021) Structural analysis of the SRP alu domain from plasmodium falciparum reveals a non-canonical open conformation. *Commun. Biol.*, **4**, 600.
113. Gong, S., Zhang, C. and Zhang, Y. (2019) RNA-align: quick and accurate alignment of RNA 3D structures based on size-independent TM-scoreRNA. *Bioinformatics*, **35**, 4459–4461.
114. Grandin, M., Meier, M., Delcros, J.G., Nikodemus, D., Reuten, R., Patel, T.R., Goldschneider, D., Orriss, G., Krahn, N., Boussouar, A. *et al.* (2016) Structural decoding of the netrin-1/unc5 interaction and its therapeutic implications in cancers. *Cancer Cell*, **29**, 173–185.
115. Reuten, R., Patel, T.R., McDougall, M., Rama, N., Nikodemus, D., Gilbert, B., Delcros, J.G., Prein, C., Meier, M., Metzger, S. *et al.* (2016) Structural decoding of netrin-4 reveals a regulatory function towards mature basement membranes. *Nat. Commun.*, **7**, 13515.
116. Dagenais, P., Desjardins, G. and Legault, P. (2021) An integrative NMR-SAXS approach for structural determination of large RNAs defines the substrate-free state of a trans-cleaving neurospora varukd satellite ribozyme. *Nucleic Acids Res.*, **49**, 11959–11973.
117. Niu, X., Sun, R., Chen, Z., Yao, Y., Zuo, X., Chen, C. and Fang, X. (2021) Pseudoknot length modulates the folding, conformational dynamics, and robustness of xrn1 resistance of flaviviral xrRNAs. *Nat. Commun.*, **12**, 6417–6417.
118. Kim, D.N., Thiel, B.C., Mrozowich, T., Hennelly, S.P., Hofacker, I.L., Patel, T.R. and Sanbonmatsu, K.Y. (2020) Zinc-finger protein CNBP alters the 3-D structure of lncRNA braveheart in solution. *Nat. Commun.*, **11**, 148.
119. Castillo-Martínez, J., Fan, L., Szewczyk, M.P., Wang, Y.X. and Gallego, J. (2022) The low-resolution structural models of hepatitis c virus RNA subdomain 5BSL3.2 and its distal complex with domain 3'X point to conserved regulatory mechanisms within the flaviviridae family. *Nucleic Acids Res.*, **50**, 2287–2301.
120. Janecki, D.M., Swiatkowska, A., Szpotkowska, J., Urbanowicz, A., Kabacińska, M., Szpotkowski, K. and Ciesiolka, J. (2021) Poly(C)-binding protein 2 regulates the p53 expression via interactions with the 5'-Terminal region of p53 mRNA. *Int. J. Mol. Sci.*, **22**, 13306.
121. Kappel, K., Zhang, K., Su, Z., Watkins, A.M., Kladwang, W., Li, S., Pintilie, G., Topkar, V.V., Rangan, R., Zheludev, I.N. *et al.* (2020) Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat. Methods*, **17**, 699–707.
122. Zhang, K., Li, S., Kappel, K., Pintilie, G., Su, Z., Mou, T.C., Schmid, M.F., Das, R. and Chiu, W. (2019) Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat. Commun.*, **10**, 5511.