

Article

A New Approach to 3D Modeling of Inhomogeneous Populations of Viral Regulatory RNA

Patrick S. Osmer¹, Gatikrushna Singh² and Kathleen Boris-Lawrie^{2,*}¹ Department of Astronomy, The Ohio State University, Columbus, OH 43210, USA; osmer.1@osu.edu² Department of Veterinary and Biomedical Sciences, University of Minnesota, Saint Paul, MN 55108, USA; gsingh@umn.edu

* Correspondence: kbl@umn.edu; Tel.: +1-612-625-2100

Received: 3 July 2020; Accepted: 27 September 2020; Published: 29 September 2020



Abstract: Tertiary structure (3D) is the physical context of RNA regulatory activity. Retroviruses are RNA viruses that replicate through the proviral DNA intermediate transcribed by hosts. Proviral transcripts form inhomogeneous populations due to variable structural ensembles of overlapping regulatory RNA motifs in the 5'-untranslated region (UTR), which drive RNAs to be spliced or translated, and/or dimerized and packaged into virions. Genetic studies and structural techniques have provided fundamental input constraints to begin predicting HIV 3D conformations in silico. Using SimRNA and sets of experimentally-determined input constraints of HIV^{NL4-3} trans-activation responsive sequence (TAR) and pairings of unique-5' (U5) with dimerization (DIS) or AUG motifs, we calculated a series of 3D models that differ in proximity of 5'-Cap and the junction of TAR and PolyA helices; configuration of primer binding site (PBS)-segment; and two host cofactors binding sites. Input constraints on U5-AUG pairings were most compatible with intramolecular folding of 5'-UTR motifs in energetic minima. Introducing theoretical constraints predicted metastable PolyA region drives orientation of 5'-Cap with TAR, U5 and PBS-segment helices. SimRNA and the workflow developed herein provides viable options to predict 3D conformations of inhomogeneous populations of large RNAs that have been intractable to conventional ensemble methods.

Keywords: 5'-cap; centroid; complex 5'-untranslated region; intramolecular folding; medoid; retrovirus; RNA structure ensemble; supercomputer cluster

1. Introduction

Intramolecular folding of RNA secondary structures drives tertiary conformation of RNA molecules [1–3]. The HIV-1 5'-untranslated region (HIV 5'-UTR) contains overlapping structural motifs that regulate early and late replication events [4–10]. Within the 5'-UTR, unspliced proviral transcripts engage host ribonucleoproteins (RNPs) that catalyze splicing or translation to virion proteins, or viral nucleocapsid RNPs drive pairs of RNA molecules into diploid genomic RNPs (gRNPs) to be packaged into virions. Rather than be trapped in a single multifunctional conformation, HIV 5'-UTRs exist as inhomogeneous populations transiting metastable structures that are impractical to capture experimentally [11].

Ample secondary structure information is available on individual segments of the HIV 5'-UTR that are necessary for dimerization and packaging of diploid gRNA into virions [6,9,12–20]. The dimer-prone secondary structure exhibits pairing of primary sequences of the unique 5' region (U5) and AUG regions (U5-AUG model) [4,8]. U5-AUG pairing orients the dimer initiation sequence (DIS) for intermolecular dimerization and the core encapsidation signal (CES) for nucleocapsid binding, and packaging of the diploid genomic ribonucleoprotein into virions [8].

In-solution studies have established the dimer-prone secondary structure (U5-AUG model) exists in thermodynamic equilibrium with U5 pairing DIS (U5-DIS model) [4]. U5-DIS pairings reorient metastable AUG and CES sequences into branched multiple hairpins that characterize the monomer 5'-UTR (non-dimer prone) [4,8]. Nucleotide pairings favoring the dimer-prone 5'-UTR conformation or monomer conformation have been examined in-solution. Substitutions in DIS or destabilizing U5-AUG pairings diminish formation of dimers in synthetic RNAs and or RNA preparations from cells and virions [21].

Select base substitutions that destabilize U5-AUG and unpair adjacent PolyA nts significantly upregulate the HIV RNA translation rate in infected lymphocytes, demonstrating monomer structure significantly affects activity [22]. Unpaired PolyA nts have been shown to influence the structural topology of the 5'-cap site and modulate engagement by host cap-binding proteins [6,23]. The phylogenetic conservation of PolyA indicates that experimentally changing these nt-nt pairings may promote maximum viral translation efficiency, but does not necessarily maintain optimal efficiency of other structural motifs within the viral 5'-UTR [22].

RNA molecules fold in a hierarchical pathway, primary sequence folding into two-dimensional (2D) multi-helical loops, bulges and stems that undergo intramolecular folding by three-dimensional (3D) interactions [1,24–27]. The 3D interactions are composed of pseudoknots, non-canonical base pairings, and single or unstacked base pairs, non-canonical (not A–U, G–C, and G–U) base pairs, pseudoknots, triplet and G-quadruplex interactions. Importantly, biological functionality of RNA in 3D requires metastable regions to be energetically compatible with helices trapped in energetic minima [1,3,24].

3D models can provide a testing ground to interrogate base substitutions, including those that destabilized U5-AUG and PolyA stem to significantly upregulate HIV translation rate [28]. Current computational methods have the ability to predict 3D RNA conformations within small parameter space (<160 nt) or within larger molecules given biologically-determined input constraints [3,29,30]. Tertiary structure prediction tools, such as SimRNA have been compared and validated in the RNA Puzzles experiments [31–33]. SimRNA begins with known structural constraints in PDB format that have been derived from chemical and enzymatic mapping, crystallography, NMR, small angle X-ray scattering (SAXS), or Förster resonance energy transfer (FRET) and optimally, are validated by mutagenesis and biological assays. Structural information validated by genetic studies has been published of progressively larger fragments of HIV 5'-UTRs, providing a solid foundation of 2D input constraints for 3D modeling [4,6,9,12–20,34,35]. 3D models have the potential to begin to guide experiments characterizing contiguous HIV 5'-UTRs in large 3D ensembles.

Herein we used SimRNA to perform simulations of the contiguous HIV 5'-UTR starting from published 2D pairings of TAR, PolyA, CES and U5-DIS (monomer) or dimer-prone U5-AUG (herein designated dimer). SimRNA evaluated the positive predictive value (PPV) of the input constraints and the sensitivity of output models. Input constraints were varied based on experimentally determined metastable U5 pairings and theoretical variables. Perspective on the capability of SimRNA to simulate HIV RNA interactions with centroid secondary structure, or no secondary structure constraint was also developed.

2. Materials and Methods

The HIV-1^{NL4-3} 5'-UTR primary sequence was downloaded from NCBI <https://www.ncbi.nlm.nih.gov/nucore/AF324493> 4 September 2015 and the 5'-UTR starting with two 5' guanosine residues (Cap + G-356 nt) was studied using the SimRNA program [32,33]. The input to SimRNA was primary sequence and experimentally-determined 2D structural constraints provided in the dot-bracket language [36]. Variables in the input constraints were designated Monomer (U5-DIS model), Dimer (dimer-prone, U5-AUG model) and centroid WT with/out PolyA unpaired [4,37]. The specific parameters for the SimRNA calculations were 30 million iterations with 10 replica exchanges, for a total of 300 million iterations. Each SimRNA run required on order 60 h of clock time on the Pitzer Cluster at

The Ohio State Supercomputer Center. SimRNA provided outputs in dot-bracket language, 3D images and PDB files. Agreement between input restraints and output base pairings were benchmarked for sensitivity and positive-predictive values. Input and Output nt-nt pairings were compared in the dot-bracket language, converted to Circle plots by CircleCompare (<https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/CircleCompare.html>, 1 December 2018) and traditional 2D sequence models were drawn in Adobe Photoshop 6. Visual Molecular Dynamics software package (VMD 1.9.3) [38] produced static views and movies of the Output models.

3. Results

3.1. SimRNA Evaluated 3D Properties of HIV^{NL4-3} 5'-UTR Beginning with 5'-Capped-Guanosine

All simulations used the same HIV^{NL4-3} primary sequence of 356 nt commencing with guanosine residues (5'-Cap + G). The workflow relied upon experimentally determined 5'-UTR conformers (Figure 1A). Perspective on the capability of SimRNA to simulate HIV RNA was developed by comparing in silico constraints of centroid 2D or no secondary structure (noSS) inputs. The dimer-prone (herein designated Dimer) and Monomer 2D models (Figure 1B,C) were previously determined by analysis of HIV RNA fragments [4,6,9,37]. The color scheme for each 5'-UTR conformation visualized TAR, light blue; PolyA stem, navy; U5 stem, cyan; PBS, green; DIS stem, orange (Figure 1B).

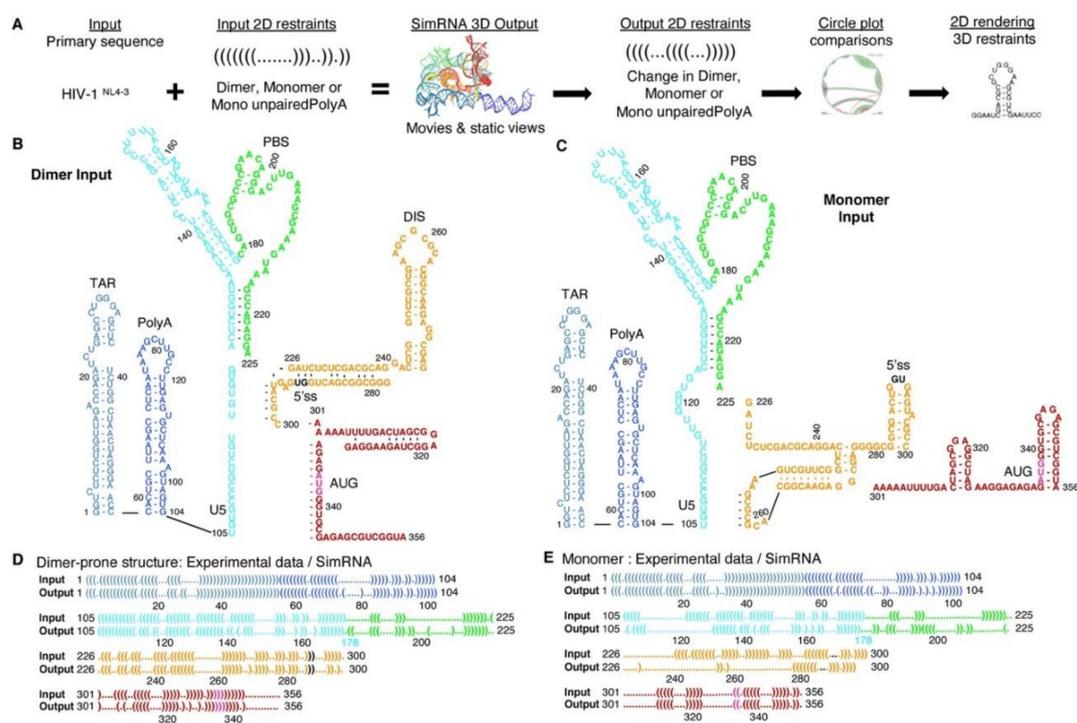


Figure 1. Approach to generate tertiary structure models of the HIV^{NL4-3} 5'-UTR. (A) Workflow used to process the tertiary structure predictions by SimRNA from experimentally determined models expressed in the Vienna dot-bracket language. (B) Input Dimer secondary structure input to SimRNA. (C) Comparison of the Input and Output structures in the Vienna dot-bracket language. (D) Monomer secondary structure input to SimRNA. (E) Comparison of the experimental Input constraints and SimRNA Output models in the Vienna dot-bracket language. Label colors designate: TAR, light blue; PolyA stem, navy; U5 stem, cyan; PBS, green; DIS and 3-way junction, orange; AUG stem, red; 5'-splice site (5'ss), black; gag start codon, magenta.

Input constraints were provided to SimRNA in Vienna dot-bracket language showing unpaired bases (dot) in relation to paired bases (brackets) (Figure 1D). SimRNA output the medoid of the three

most populated clusters of nt–nt pairings having the lowest 2% of free energy (Cluster 1, 2 or 3). Medoid of the largest cluster is the most representative tertiary conformation engendered by the input restraints (Cluster 1) and was displayed using VMD software and dot-bracket language according to the workflow (Figure 1A).

SimRNA Output from the Dimer and Monomer Input restraints generated similar numbers of objects in the top 3 clusters, as expected for identical primary sequences (Table S1). The dot-bracket language readily visualized the position of each difference between the input and output (Figure 1D). Dimer exceeded Monomer in the number of nt–nt pairings in common between the Input restraints and Output medoid (95% and 74%, respectively) (sensitivity, Table 1). Dimer Output also exceeded Monomer in the number of nt–nt pairings retained in 3D Output model that are in the Input experimentally-determined constraints (PPV, Table 1). These measurements indicated the dimer-prone U5-AUG pairings were energetically compatible with intramolecular folding of metastable regions between helices trapped in energetic minima.

Table 1. SimRNA parsed nt–nt pairings of the experimentally-determined input restraints.

5' Input Restraints ^a	Sensitivity ^b	PPV ^c
Dimer	106/111 (96%)	105/111 (95%)
Monomer	73/99 (74%)	73/99 (74%)
Mono Unpaired PolyA ^d	77/99 (78%)	77/89 (87%)
WT centroid	94/94 (100%)	94/123 (76%)

^a HIV^{NL4-3} 5'-untranslated sequence beginning with 5'-Cap-guanosine. ^b Sensitivity, pairs common to 2D Input constraints and 3D Output model. ^c PPV, positive predictive value. Pairs retained in 3D Output model that are in the Input structure. ^d Monomer Unpaired nt +59 through +103. The nt–nt pairings of Input and Output were enumerated and compared by the CircleCompare program (<https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/CircleCompare.html>). The percentage of nt pairs in the Input that were identified by SimRNA indicated sensitivity. The percentage of nt pairs identified by SimRNA that were in the Input structure measured positive predictive value (PPV).

3.2. Dimer Output 3D Model Predicted TAR-PolyA-U5 Converge near the 5'-Cap

The Dimer 3D tertiary prediction placed TAR (light blue) and PolyA (navy) helices stacking on one another (Figure 2A; Movie 1). These results agree with published results of chemical mapping [8,10,39,40] and SAXS [34,35]. The 5'-Cap (magenta space-filling atom) lay at the junction of TAR and PolyA helices and the base of the U5 stem near G104-U105. In Figure 2B, the 3D Output focused on the U5 nt (cyan atoms) pairing with AUG (red atoms), matching the input constraint. 5'-Cap was stacked on G104-U105 (white and yellow atoms) at the junction of TAR and PolyA, and AUG was in close proximity, albeit paired with U5.

In Figure 2C, the 3D Output focused on the green PBS-segment (nt 180–225) and that DIS was surface-exposed (see yellow atoms). The 5'-splice site (ss) lay between helices formed by nts 300–330 of CES (Movie 1) [37]. To emphasize the differences in nt–nt pairings between input and output, Dimer Output pairings were displayed in a Circle plot and transcribed into the traditional secondary structure format (Figure 3). Since 96% of the Input constraints maintained (Table 1), the ensemble exhibited minimal change from the Input nt–nt pairings provided for Dimer (Figure 2B). The 3D model predicted stacking of G104-U105 that had not been apparent in 2D Input constraints. The model predicted the PBS-segment (nt 134–224) containing the double-stranded primer activation signal (PAS) [41] and the tRNA-like element (TLE) at the apex of helices formed of nt 134–179 [42] and, at the base of the helices, the double-stranded RNA binding site of DHX9/RNA helicase A (RHA) [28].

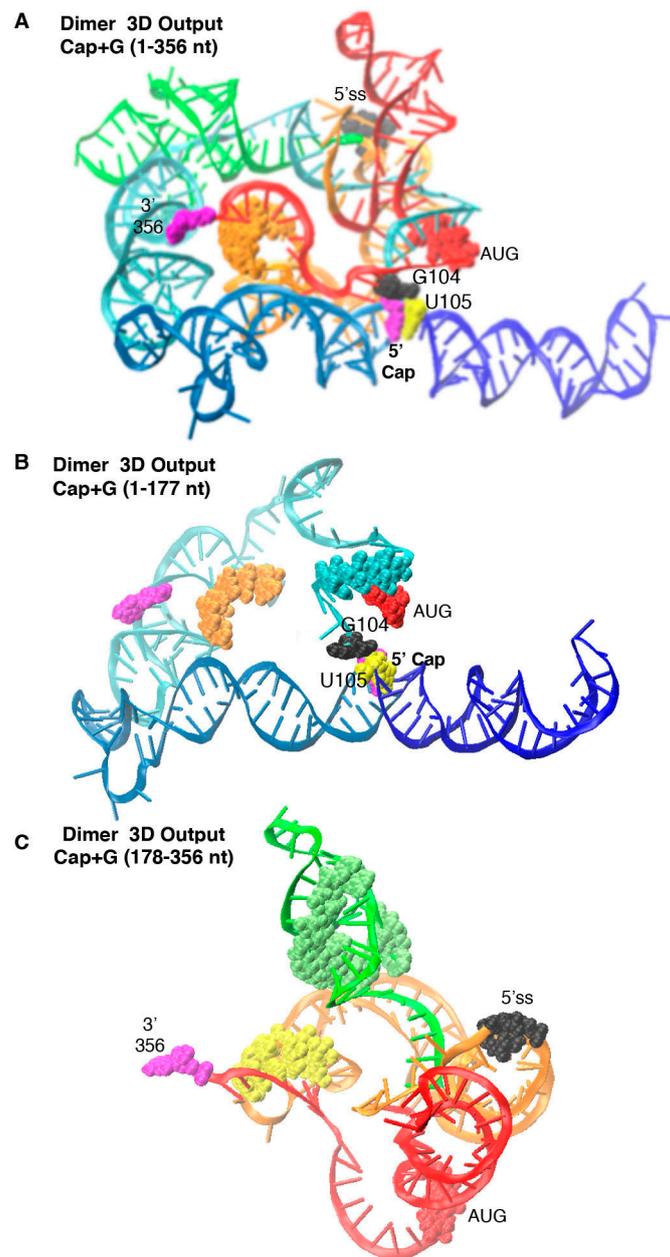


Figure 2. Dimer 3D model identified by SimRNA. (A) Most representative model of Dimer-prone 5'-UTR. (B) Focus on 5'-Cap + 1–177. (C) Focus on nt 178–356. Label colors designate: TAR, light blue; PolyA stem, navy; U5 stem, cyan; PBS, green; DIS and three-way junction, orange; 5'-Cap and 3' terminus, magenta atoms; G104, black atoms; U105, yellow atoms; AUG stem, red atoms. 5'-splice site (5'ss), black atoms; DIS atoms, orange (A,B), yellow; gag start codon, red atoms.

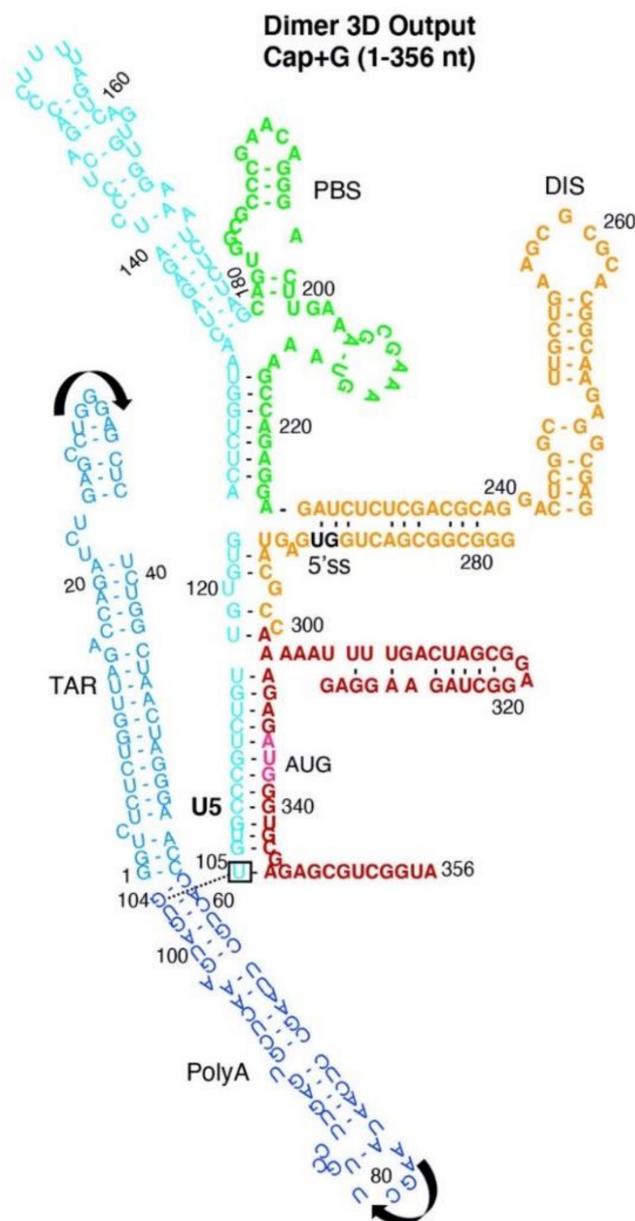


Figure 3. 2D rendering of SimRNA 3D Dimer Model. Arrows represent the orientation of TAR and PolyA helices. Box and dashed line indicates stacking of G104-U105. Label colors: TAR, light blue; PolyA, navy; U5 stem, cyan; PBS-segment, green; DIS stem, orange; AUG stem, red; 5'-splice site (5'ss), black; AUG translation start codon, magenta.

3.3. Monomer Output 3D Model Altered Accessibility of 5'-Cap and 5'-ss and Reoriented U5 Stem

The Monomer 3D model identified TAR and PolyA helices, but in a different orientation than Dimer (Figure 4A; Movie 2). Whereas Dimer modeled coaxial arrangement of the TAR-PolyA helices, Monomer modeled V-shape due to differences in stacking interactions. Figure 4B focused on the close proximity of 5'-Cap, G104, U105 and DIS (orange atoms). Helices formed of U5-DIS pairs and downstream orange nt (e.g., 256–260) showed the 5'-Cap intercalated between these helices (compare Figure 4A,B; Movie 2). Figure 4C focused on the 5'ss at the apex of a stem loop. Gag AUG was paired close to the 3'-terminus of the 5'-UTR. The models depict TAR-PolyA-U5 stems intersect around the 5'-Cap (Figure 4A).

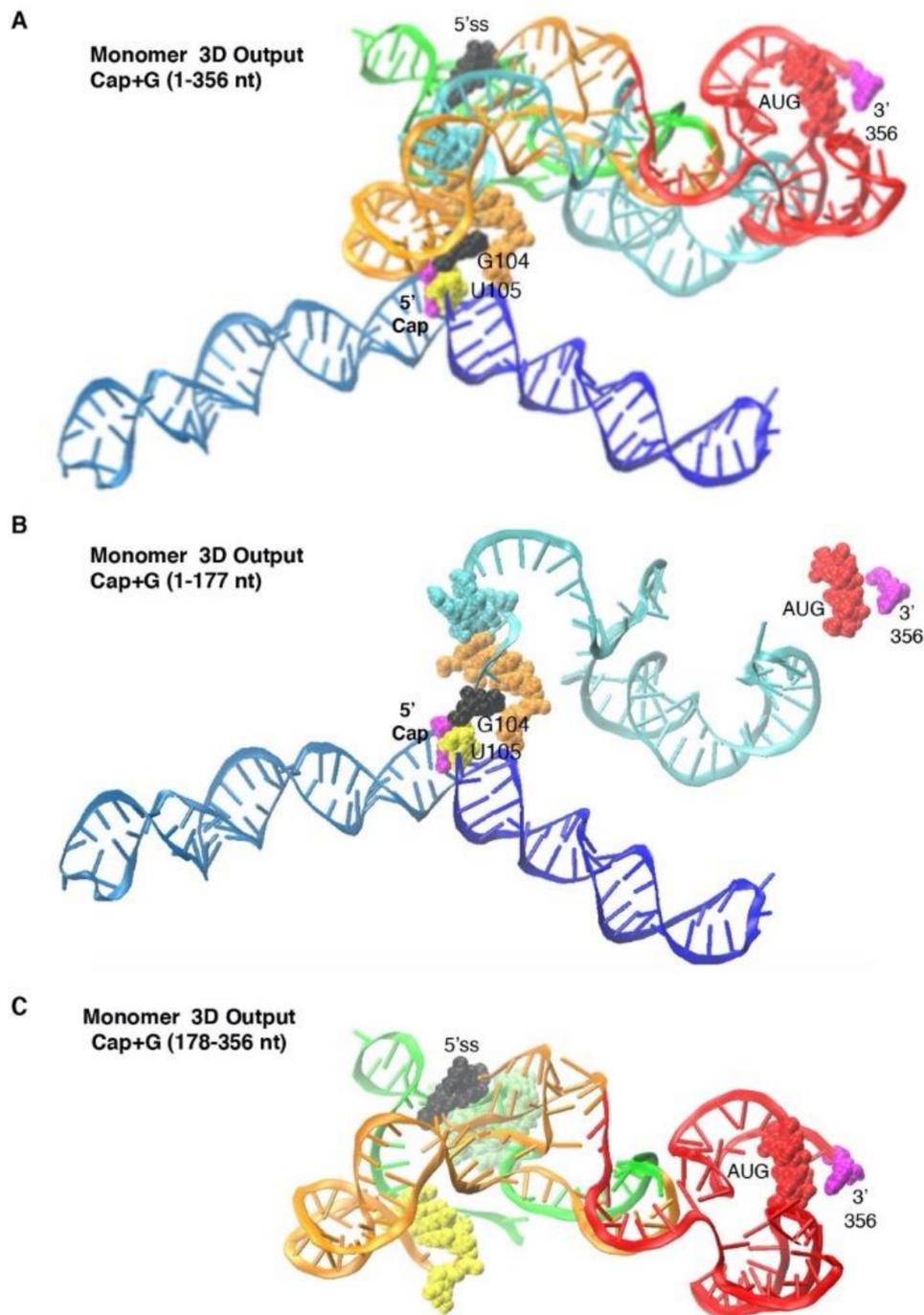


Figure 4. Monomer 3D model identified by SimRNA. (A) Most representative model of Monomer-prone 5'-UTR. (B) Focus on nt 1–177. (C) Focus on nt 178–356. Label colors designate: TAR, light blue; PolyA stem, navy; U5 stem, cyan; PBS, green; DIS stem and three-way junction, orange; 5'-Cap and 3' terminus, magenta; G104, black; U105, yellow; AUG stem, red. 5'-splice site (5'ss), black; DIS, orange (A,B), yellow (C); gag start codon, red.

Notably, only 74% of the Input nt–nt pairings maintained in Monomer Output (Table 1). The lower positive predictive value of Monomer versus Dimer (74% versus 87%, Table 1) suggested U5-DIS helices significantly changed intramolecular folding of 5-UTR relative to Dimer.

Circle plots compared the different nt pairings between Monomer Input restraints and Monomer Output (Supplementary Figure S1) and were used to generate the traditional secondary structure

drawing (Figure 5). The nt-nt pairings of TAR and PolyA maintained in the 3D Output model, but new pairings encompassed the U5 stem (cyan) through the PBS-segment (green) (compare Figure 5 with Figure 1C), consistent with the long-distance interaction model of Huthoff and Berkhout [8]. The 3D Monomer model eliminated the nt-nt pairings of PAS [41], TLE [42] and the binding site of DHX9/RHA [28]. We concluded tertiary interactions drove rearrangement of metastable Monomer Input restraints in silico. The Output model predicted three elements important to reverse transcription were incompatible with U5-DIS pairings.

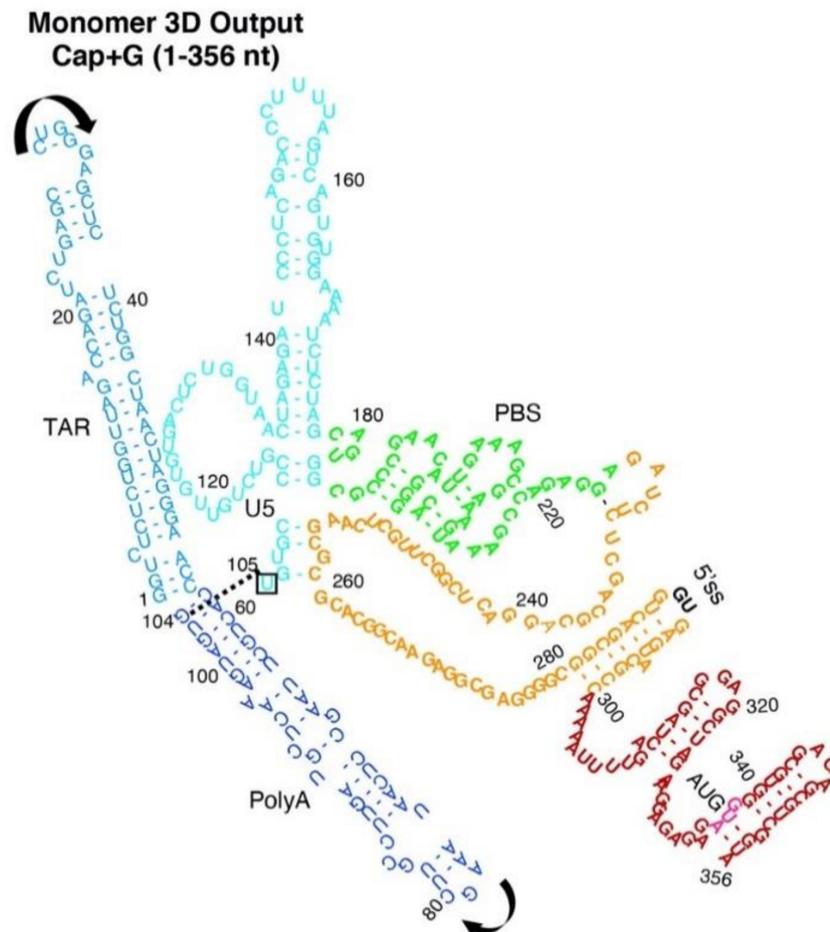


Figure 5. 2D rendering of SimRNA 3D Monomer model. Arrows represent the direction of TAR or PolyA helices. Label colors designate: TAR, light blue; PolyA, navy; U5 stem, cyan; PBS-segment, green; DIS stem, orange; AUG stem, red; 5'-splice site (5'ss), black; nt of the 3-way junction, orange. Space filling atoms: 5'-Cap, white; G104, yellow; U105, black; AUG translation start codon, red or magenta.

3.4. The Unpairing of PolyA Nts Reduced Local Energy Minima in the Thermodynamic Equilibrium to Monomer Tertiary Structure

Unpaired residues modulate local energy minima in the thermodynamic equilibrium between structural intermediates [43,44]. Biological evidence has unequivocally demonstrated the nt-nt pairings observed in TAR are essential for HIV transcriptional trans-activation by Tat [45–48], whereas the nt-nt pairings of the PolyA stem are less stringently required [49]. To test Input constraints without PolyA pairing, we provided SimRNA the identical Monomer Input restraints, except residues 59–103 were unpaired (Mono PolyA unpaired).

SimRNA calculated 300 million iterations of the input restraints comparing Monomer and Mono PolyA unpaired. Output benchmarks improved for Mono PolyA unpaired, indicating metastable PolyA region favored intramolecular folding of the nt-nt pairings in energetic minima (Table 1).

The metastable PolyA region changed proximity of 5'-Cap to G104 and U105 at the base of the U5 stem (Figure 6A,B; Movie 3) and reoriented AUG near to U5, rather than near to PBS as observed in Monomer (Figure 6A,B compared with Figure 4B). The 5'ss remained at the apex of a stem loop, yet was not intercalated between the PBS and residues of the 3-way junction, as observed in Monomer (Figure 6A,B compared with Figure 4C).

Approximately 80% of the nt-nt pairings were in common with Monomer and all of the differences were downstream from PolyA stem (Figure S2). We concluded the ensemble space of 5'-UTR conformation was reduced by PolyA nt pairings. These observations agree with the prior finding that unpaired residues play an important, passive role in HIV 5'-UTR secondary structure [7,50]. In sum, stacking of 5'-Cap with G104-U105 favored by input restraints pairing the PolyA nts.

3.5. TAR Input 2D Restraints Significantly Influence SimRNA Models

To add perspective on the power of laboratory versus computationally-derived constraints useful for SimRNA modeling, we ran two additional models, the first using the secondary structure of the centroid of the HIV ensemble as input and the second using no 2D constraint, only the primary sequence as input data (noSS). The centroid is an ab initio thermodynamic average of HIV^{NL4-3} 2D ensemble (herein designated Centroid WT) [36,51,52]. Whereas the thermodynamic average of 2D RNA structures integrates the minimum free energy (MFE) of canonical base pairings (G-C, A-U, G-U) and folding constraints of nearest neighbor bases within short RNA strands [51], advances in dynamic programming algorithms take this approach a step further by monitoring ensembles of metastable structures [52]. This dynamic approach samples nt-nt pairings, calculating probabilities of concurrent sub-structures from the ensemble of metastable structures and structures trapped in energetic minima. Convergence on probabilistic structures is required to produce the predominant stems, bulges, internal and multi-helical loops and provide the most representative model, or centroid.

Centroid WT Input constraints did not demonstrate input constraints additional to the experimentally-determined Monomer and Dimer Input 2D constraints (Figure S3A,B). Centroid WT secondary structure prediction mimicked the U5-AUG pairings observed for Dimer. Next, the computationally-derived Centroid WT input was processed 300 million iterations by SimRNA and the workflow processed results of the theoretical input constraints (Figure 7). Centroid WT Input pairings maintained and the top output cluster had 12 additional pairings (Figure S2C).

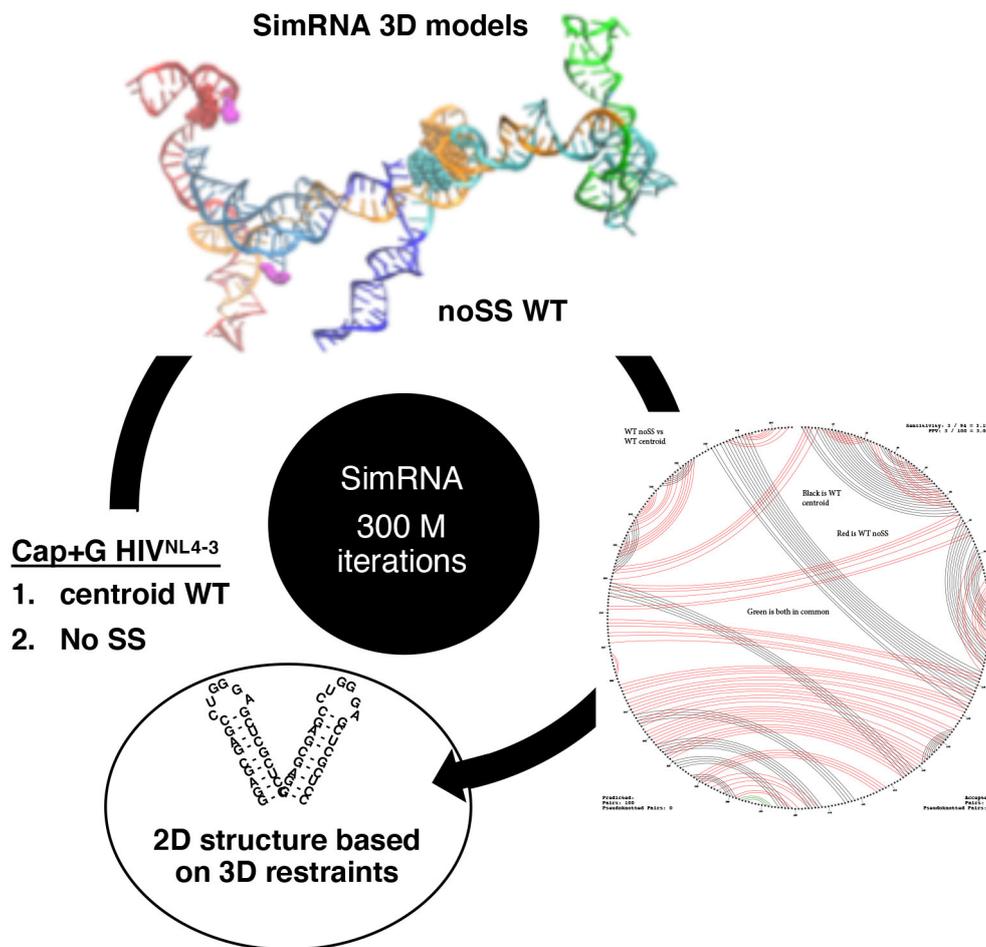


Figure 7. Approach to generate tertiary structure of HIV^{NL4-3} 5'-UTR using theoretical Input constraints. Workflow presents two sources of theoretical Input restraints provided to SimRNA to predict 3D models of the HIV 5-UTR in the course of 300 million iterations on the Pitzer Supercomputer Cluster. The most representative 3D models were backed into Vienna dot-bracket language and 2D features were analyzed using CircleCompare and used to render 2D comparison with the input 2D model. The 3D model and Circleplot present the SimRNA simulation of HIV 5'-UTR with no secondary structure restraint (no SS). Label colors designate: TAR, light blue; PolyA stem, navy; U5 stem, cyan; PBS stem, green; DIS stem, orange; AUG stem, red. Atoms: 5'-Cap and 3' nt, magenta; U5, cyan; DIS, orange; 5'-splice site (5'ss), black; gag start codon, red.

Centroid WT predicted TAR and PolyA arranged coaxial helices and base stacking interaction between Cap, G104-U105 (Figure 8A,B compared with Figure 4A,B). Unlike Dimer, the 5'ss was positioned at the apex of the stem loop formed by nt 280–300, as observed in Monomer (Figure 8A,C compared with Figure 4A,C). The U5 stem was an extended helix (nt 141–177 and 225–280) topped with PBS (green) (Figures 8A,C and 9), similar to structural intermediates of HIV 5'-UTR dimerization [43, 44,46,50].

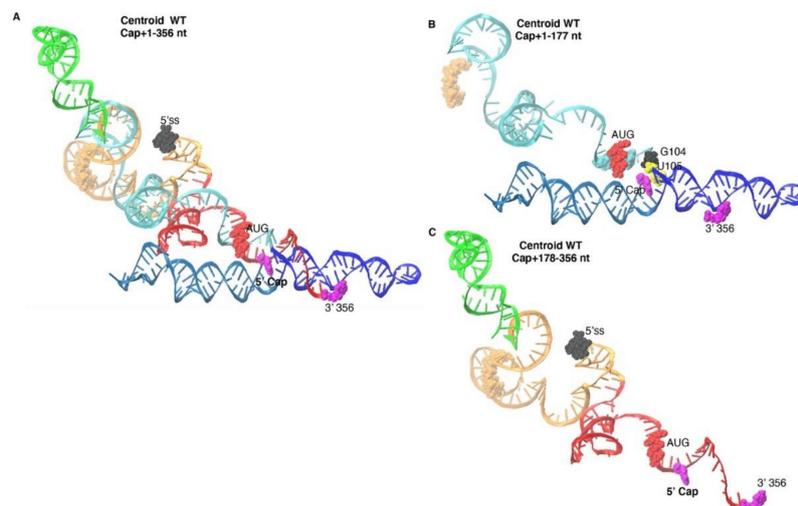


Figure 8. Output of Centroid WT predicted U5-AUG pairing and apical position of 5'ss. (A) 3D model predicted by SimRNA based on Centroid WT theoretical input constraints. (B) Focus on 5'-Cap + 1–177. (C) Focus on nt 178–356. Colors designate: TAR, light blue; PolyA, navy; U5 stem, cyan; PBS-segment, green; DIS stem and 3-way junction, orange; AUG stem, red; 5'-splice site (5'ss), black. Space filling atoms: 5'-Cap and 3' terminus, magenta; G104, yellow, U105, black, AUG translation start codon, red.

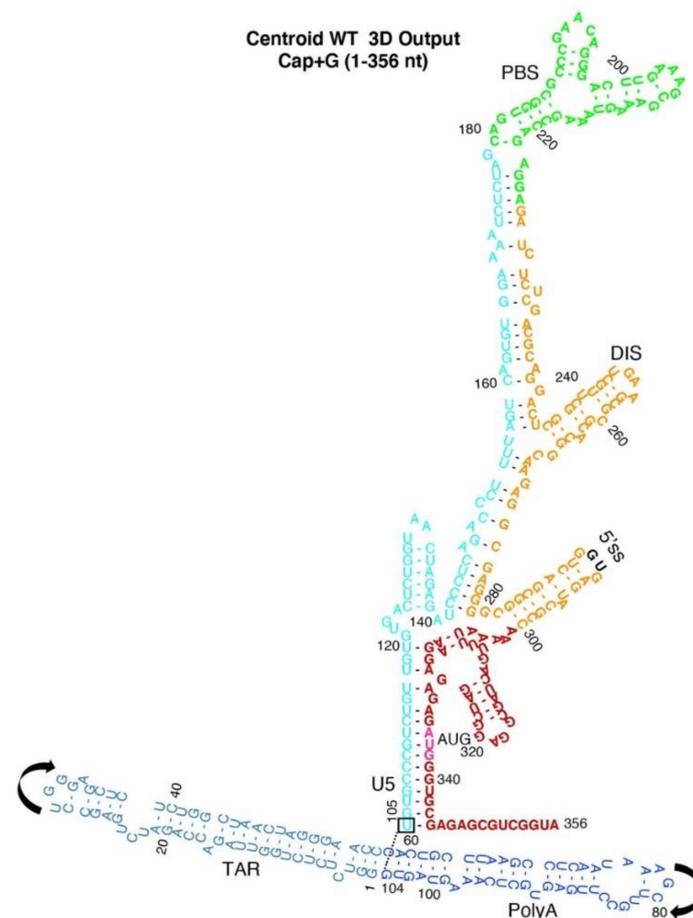


Figure 9. 2D rendering of Centroid WT Output. Arrows represent the direction of TAR or PolyA helices. The colored font designates: TAR, light blue; PolyA, navy; U5 stem, cyan; PBS-segment, green; DIS stem, orange; AUG stem, red; 5'-splice site (5'ss), black; nt of the 3-way junction, orange. Space filling atoms: 5'-Cap, white; G104, yellow, U105, black, AUG translation start codon, red or magenta.

The Output from Centroid WT recapitulated TAR and PolyA stems and pairing of U5 nts 134–140 (Figures 8 and 9). As summarized in Table 2, 60% of the identified pairings were maintained in Dimer and Monomer (Table 2), indicating overlap in the parameter spaces. The 3D simulations of Centroid WT Input restraints predicted thermodynamic stability of U5-AUG pairings, similar to Dimer, and apical position of the 5'ss that is similar to Monomer.

Table 2. Input restraints strongly influence positive predictive value of SimRNA.

HIV ^{NL4-3} 5'-UTR 2D Restraints ^a	Sensitivity ^b	PPV ^c
Input	Versus Centroid WT Output	
Monomer	54/94 (58%)	54/99 (55%)
Dimer	58/94 (62%)	58/111 (52%)
noSS WT ^d	3/100 (3%)	3/94 (3%)

^a HIV^{NL4-3} 5'-untranslated sequence beginning with 5'-Cap-guanosine. ^b Sensitivity, pairs identified in both Input and SimRNA Output. ^c PPV, positive predictive value, pairs identified by SimRNA that are in the Input structure. The nt-nt pairings of SimRNA models were enumerated and compared by the CircleCompare program (<https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/CircleCompare.html>) 12 December 2018. The percentage of nt pairs in each Input model that were identified by SimRNA in centroid WT model indicated sensitivity. The percentage of nt pairs identified in each Input that were in centroid WT model measured positive predictive value (PPV). ^d noSS, no secondary structure input constraint.

Output ensembles of centroid WT and noSS had similar numbers of total objects as Dimer and Monomer (Table S1). However, the Output from noSS Input constraints recapitulated only 3% of the pairings within centroid WT (Table 2). These results indicated the energetic cost of TAR was excessive within the parameter space of 300 million iterations. Given noSS Input constraints, SimRNA sampled metastable structures without constraint on TAR, a condition comparable to basal HIV transcription that is Tat/TAR-independent independent. The results suggested the TAR helix significantly influences the thermodynamic equilibrium of the first ~100 nt of the 5'-UTR.

4. Discussion

In the course of 300 million iterations of dynamic assemblies, SimRNA evaluated the 3D parameter space of the 5'-UTR beginning with experimentally-determined 2D constraints, theoretical constraints on PolyA nt pairing, and the secondary structure centroid prediction of its whole Boltzmann ensemble. SimRNA predictions suggest TAR helical stem and metastable PolyA region are fundamental input constraints determining the proximity of 5'-Cap to the junction of TAR and PolyA helices and the base of the U5 stem near G104-U105. The SimRNA 3D modeling agreed in principle with recent NMR results that PolyA nt pairing/unpairing propagates structural rearrangements throughout the 5'-UTR, perturbing the orientation of 5'-Cap, TAR and U5 residues, [6,53].

The metastable nt pairings in U5 and PBS-segment likely contribute to the alternate functionalities of the 5'-UTR in cellulo. Destabilized PBS-segment reduces affinity for DHX9/RHA and diminishes processivity of reverse transcriptase and virus infectivity [28,54,55]. Destabilized A59-U103 pairing significantly increase Gag translation rate in cells [22] and increase biochemical affinity of the HIVMAL 5'-Cap for host eIF4E [6]. The HIV late RNAs have been shown to specifically retain nuclear cap-binding complex (CBC) in polysomes and virions, instead of exchanging to eIF4E [56]. Recently, polysomal CBC RNPs were shown to activate eIF4E-independent translation of stress response proteins, overcoming AP-1 protein translation attenuation by mTOR inhibition [57]. Structural malleability near the HIV 5'-Cap may be beneficial to regulate cap-exchange and antagonize antiviral mTOR inhibition.

Tertiary modeling of polypeptide conformations has been integral to elucidating the functional mechanisms of enzymes, chaperones, many structural proteins and receptor-ligand interactions and for in silico modeling of small molecule therapeutics. Tertiary modeling of RNA structures has robust potential to guide the rational design of antiviral therapeutics [58–61]. Supportive evidence that alternative nt-nt pairings propagate structural changes throughout the HIV-1 leader RNA includes

TAR-binding compounds exert structural effects outside TAR [62,63] and small molecule binding within the CES reduce virus titer [64]. Our results suggest programs like SimRNA can help to begin to predict RNA structure-function relationships testable in structural studies and cell-based studies [3].

5. Conclusions

By incorporating experimental and theoretical constraints, and performing 300 million iterations of dynamic ensembles, the SimRNA software package was used to simulate tertiary conformations of the HIV^{NL4-3} 5'-UTR. The Output models predicted tertiary interactions in context of U5-AUG pairings placed G104-U105 stacking on 5'-Cap and maintained helices of the PBS-segment previously identified as PAS, TLE [41,42] and the double-stranded RNA binding site for DHX9/RHA [28]. Alternative U5 pairing with DIS reduced SimRNA agreement of input constraints with tertiary models. In conclusion, secondary structures that support early and late events in HIV replication divergently effect the 5'-UTR 3D models identified by the SimRNA algorithm. The simulation workflow developed herein provides a viable approach to model inhomogeneous large RNA populations.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1999-4915/12/10/1108/s1>, Figure S1: Circle plots of Input and Output nt-nt pairings identified for Dimer-prone and Monomer HIV^{NL4-3}. Figure S2: Circle plots comparing 3D output of Monomer with the top 3 clusters of Monomer UnpairedPolyA (cluster 1, 2, 3). Figure S3. Comparison of Input restraints of centroid WT to Dimer or Monomer. Table S1. Number of objects in the top 3 ensembles of objects having the lowest 2% of free energy, Video S1: SimRNA Output 3D Model of Dimer-prone 5'-UTR. Video S2: SimRNA Output 3D Model of Monomer 5'-UTR. Video S1: SimRNA Output 3D Model of Mono PolyA unpaired.

Author Contributions: Conceptualization, P.S.O. and K.B.-L.; methodology, P.S.O., K.B.-L. and G.S.; software, P.S.O.; validation, P.S.O., G.S. and K.B.-L.; formal analysis, P.S.O. and K.B.-L.; investigation, K.B.-L.; resources, P.S.O. and K.B.-L.; data curation, P.S.O., G.S., K.B.-L.; writing—original draft preparation, K.B.-L.; writing—review and editing, P.S.O., G.S. and K.B.-L.; visualization, P.S.O., K.B.-L. and G.S.; supervision, K.B.-L.; project administration, K.B.-L. All authors have read and agreed to the published version of the manuscript.

Funding: The College of Veterinary Medicine, University of Minnesota—Twin Cities to K.B.-L. and The Ohio State University to P.S.O.

Acknowledgments: Computing time used in this work was provided by The Ohio Supercomputer Center. 2018. Pitzer Supercomputer. Columbus, OH: Ohio Supercomputer Center.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Westhof, E.; Fritsch, V. RNA folding: Beyond Watson–Crick pairs. *Structure* **2000**, *8*, R55–R65. [[CrossRef](#)]
- Westhof, E.; Masquida, B.; Jossinet, F. Predicting and Modeling RNA Architecture. *Cold Spring Harb. Perspect. Biol.* **2010**, *3*, a003632. [[CrossRef](#)] [[PubMed](#)]
- Miao, Z.; Westhof, E. RNA Structure: Advances and Assessment of 3D Structure Prediction. *Annu. Rev. Biophys.* **2017**, *46*, 483–503. [[CrossRef](#)]
- Lu, K.; Heng, X.; Garyu, L.; Monti, S.; Garcia, E.L.; Kharytonchyk, S.; Dorjsuren, B.; Kulandaivel, G.; Jones, S.; Hiremath, A.; et al. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science* **2011**, *334*, 242–245. [[CrossRef](#)] [[PubMed](#)]
- De Guzman, R.N.; Wu, Z.R.; Stalling, C.C.; Pappalardo, L.; Borer, P.N.; Summers, M.F. Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3 -RNA Recognition Element. *Science* **1998**, *279*, 384–388. [[CrossRef](#)] [[PubMed](#)]
- Brown, J.D.; Kharytonchyk, S.; Chaudry, I.; Iyer, A.S.; Carter, H.; Becker, G.; Desai, Y.; Glang, L.; Choi, S.H.; Singh, K.; et al. Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science* **2020**, *368*, 413–417. [[PubMed](#)]
- Berkhout, B. Structure and Function of the Human Immunodeficiency Virus Leader RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **1996**, *54*, 1–34. [[CrossRef](#)]
- Huthoff, H.; Berkhout, B. Two alternating structures of the HIV-1 leader RNA. *RNA* **2001**, *7*, 143–157. [[CrossRef](#)]

9. Keane, S.C.; Heng, X.; Lu, K.; Kharytonchyk, S.; Ramakrishnan, V.; Carter, G.; Barton, S.; Hosis, A.; Florwick, A.; Santos, J.; et al. Structure of the HIV-1 RNA packaging signal. *Science* **2015**, *348*, 917–921. [[CrossRef](#)]
10. Abbink, T.E.M.; Ooms, M.; Haasnoot, P.C.J.; Berkhout, B. The HIV-1 Leader RNA Conformational Switch Regulates RNA Dimerization but Does Not Regulate mRNA Translation. *Biochemistry* **2005**, *44*, 9058–9066. [[CrossRef](#)]
11. Smyth, R.P.; Negroni, M.; Lever, A.M.; Mak, J.; Kenyon, J.C. RNA Structure—A Neglected Puppet Master for the Evolution of Virus and Host Immunity. *Front. Immunol.* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
12. Pham, V.V.; Salguero, C.; Khan, S.N.; Meagher, J.L.; Brown, W.C.; Humbert, N.; De Rocquigny, H.; Smith, J.L.; D'Souza, V.M. HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. *Nat. Commun.* **2018**, *9*, 4266. [[CrossRef](#)] [[PubMed](#)]
13. Aboul-Ela, F. Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res.* **1996**, *24*, 3974–3981. [[CrossRef](#)]
14. Du, Z.; Lind, K.E.; James, T.L. Structure of TAR RNA Complexed with a Tat-TAR Interaction Nanomolar Inhibitor that Was Identified by Computational Screening. *Chem. Biol.* **2002**, *9*, 707–712. [[CrossRef](#)]
15. Ennifar, E.; Walter, P.; Ehresmann, B.; Ehresmann, C.; Dumas, P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Genet.* **2001**, *8*, 1064–1068. [[CrossRef](#)]
16. Baba, S.; Takahashi, K.-I.; Noguchi, S.; Takaku, H.; Koyanagi, Y.; Yamamoto, N.; Kawai, G. Solution RNA Structures of the HIV-1 Dimerization Initiation Site in the Kissing-Loop and Extended-Duplex Dimers. *J. Biochem.* **2005**, *138*, 583–592. [[CrossRef](#)]
17. Renisio, J.-G.; Cosquer, S.; Cherrak, I.; El Antri, S.; Mauffret, O.; Femandjian, S. Pre-organized structure of viral DNA at the binding-processing site of HIV-1 integrase. *Nucleic Acids Res.* **2005**, *33*, 1970–1981. [[CrossRef](#)]
18. Lebars, I.; Legrand, P.; Aimé, A.; Pinaud, N.; Fribourg, S.; Di Primo, C. Exploring TAR–RNA aptamer loop–loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Res.* **2008**, *36*, 7146–7156. [[CrossRef](#)]
19. Ennifar, E.; Dumas, P. Polymorphism of Bulged-out Residues in HIV-1 RNA DIS Kissing Complex and Structure Comparison with Solution Studies. *J. Mol. Biol.* **2006**, *356*, 771–782. [[CrossRef](#)]
20. Ferner, J.; Suhartono, M.; Breitung, S.; Jonker, H.R.A.; Hennig, M.; Wöhnert, J.; Göbel, M.; Schwalbe, H. Structures of HIV TAR RNA-Ligand Complexes Reveal Higher Binding Stoichiometries. *ChemBioChem* **2009**, *10*, 1490–1494. [[CrossRef](#)]
21. Dubois, N.; Marquet, R.; Paillart, J.-C.; Bernacchi, S. Retroviral RNA Dimerization: From Structure to Functions. *Front. Microbiol.* **2018**, *9*, 527. [[CrossRef](#)] [[PubMed](#)]
22. Boeras, I.; Seufzer, B.; Brady, S.; Rendahl, A.; Heng, X.; Boris-Lawrie, K. The basal translation rate of authentic HIV-1 RNA is regulated by 5'UTR nt-pairings at junction of R and U5. *Sci. Rep.* **2017**, *7*, 6902. [[CrossRef](#)] [[PubMed](#)]
23. Khoury, G.; MacKenzie, C.; Ayadi, L.; Lewin, S.R.; Branlant, C.; Purcell, D. Tat IRES modulator of tat mRNA (TIM-TAM): A conserved RNA structure that controls Tat expression and acts as a switch for HIV productive and latent infection. *Nucleic Acids Res.* **2019**, *48*, 2643–2660. [[CrossRef](#)]
24. Nowakowski, J.; Tinoco, I. RNA Structure and Stability. *Semin. Virol.* **1997**, *8*, 153–165. [[CrossRef](#)]
25. Brion, P.; Westhof, E. Hierarchy and Dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113–137. [[CrossRef](#)] [[PubMed](#)]
26. Tinoco, I.; Bustamante, C. How RNA folds. *J. Mol. Biol.* **1999**, *293*, 271–281. [[CrossRef](#)]
27. Batey, R.T.; Rambo, R.P.; Doudna, J.A. Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed.* **1999**, *38*, 2326–2343. [[CrossRef](#)]
28. Boeras, I.; Song, Z.; Moran, A.; Franklin, J.; Brown, W.C.; Johnson, M.; Boris-Lawrie, K.; Heng, X. DHX9/RHA Binding to the PBS-Segment of the Genomic RNA during HIV-1 Assembly Bolsters Virion Infectivity. *J. Mol. Biol.* **2016**, *428*, 2418–2429. [[CrossRef](#)]
29. Zuker, M.; Sankoff, D. RNA secondary structures and their prediction. *Bull. Math. Biol.* **1984**, *46*, 591–621. [[CrossRef](#)]
30. McCaskill, J.S. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **1990**, *29*, 1105–1119. [[CrossRef](#)]

31. Lu, Z.J.; Gloor, J.W.; Mathews, D.H. Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA* **2009**, *15*, 1805–1813. [[CrossRef](#)] [[PubMed](#)]
32. Boniecki, M.; Lach, G.; Dawson, W.K.; Tomala, K.; Lukasz, P.; Soltysiński, T.; Rother, K.M.; Bujnicki, J. SimRNA: A coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **2015**, *44*, e63. [[CrossRef](#)]
33. Magnus, M.; Boniecki, M.J.; Dawson, W.K.; Bujnicki, J. SimRNAweb: A web server for RNA 3D structure modeling with optional restraints. *Nucleic Acids Res.* **2016**, *44*, W315–W319. [[CrossRef](#)]
34. Danhart, E.M.; Bakhtina, M.; Cantara, W.A.; Kuzmishin, A.B.; Ma, X.; Sanford, B.L.; Vargas-Rodriguez, O.; Košutić, M.; Goto, Y.; Suga, H.; et al. Conformational and chemical selection by a trans-acting editing domain. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E6774–E6783. [[CrossRef](#)]
35. Brigham, B.S.; Kitzrow, J.P.; Reyes, J.-P.; Musier-Forsyth, K.; Munro, J.B. Intrinsic conformational dynamics of the HIV-1 genomic RNA 5'UTR. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 10372–10381. [[CrossRef](#)] [[PubMed](#)]
36. Gruber, A.R.; Lorenz, R.; Bernhart, S.H.; Neuböck, R.; Hofacker, I.L. The Vienna RNA Websuite. *Nucleic Acids Res.* **2008**, *36*, W70–W74. [[CrossRef](#)] [[PubMed](#)]
37. Heng, X.; Kharytonchyk, S.; Garcia, E.L.; Lu, K.; Divakaruni, S.S.; Lacotti, C.; Edme, K.; Telesnitsky, A.; Summers, M.F. Identification of a minimal region of the HIV-1 5'-leader required for RNA dimerization, NC binding, and packaging. *J. Mol. Biol.* **2012**, *417*, 224–239. [[CrossRef](#)] [[PubMed](#)]
38. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
39. Abbink, T.E.M.; Berkhout, B. A Novel Long Distance Base-pairing Interaction in Human Immunodeficiency Virus Type 1 RNA Occludes the Gag Start Codon. *J. Biol. Chem.* **2002**, *278*, 11601–11611. [[CrossRef](#)]
40. Mueller, N.; Van Bel, N.; Berkhout, B.; Das, A.T. HIV-1 splicing at the major splice donor site is restricted by RNA structure. *Virology* **2014**, *468*, 609–620. [[CrossRef](#)]
41. Beerens, N.; Berkhout, B. The tRNA Primer Activation Signal in the Human Immunodeficiency Virus Type 1 Genome Is Important for Initiation and Processive Elongation of Reverse Transcription. *J. Virol.* **2002**, *76*, 2329–2339. [[CrossRef](#)]
42. Jones, C.P.; Saadatmand, J.; Kleiman, L.; Musier-Forsyth, K. Molecular mimicry of human tRNA^{Lys} anti-codon domain by HIV-1 RNA genome facilitates tRNA primer annealing. *RNA* **2012**, *19*, 219–229. [[CrossRef](#)]
43. Laughrea, M.; Jetté, L. HIV-1 Genome Dimerization: Kissing-Loop Hairpin Dictates Whether Nucleotides Downstream of the 5' Splice Junction Contribute to Loose and Tight Dimerization of Human Immunodeficiency Virus RNA. *Biochemistry* **1997**, *36*, 9501–9508. [[CrossRef](#)]
44. Laughrea, M.; Shen, N.; Jetté, L.; Wainberg, M.A. Variant Effects of Non-Native Kissing-Loop Hairpin Palindromes on HIV Replication and HIV RNA Dimerization: Role of Stem–Loop B in HIV Replication and HIV RNA Dimerization. *Biochemistry* **1999**, *38*, 226–234. [[CrossRef](#)] [[PubMed](#)]
45. Berkhout, B.; Silverman, R.H.; Jeang, K.-T. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell* **1989**, *59*, 273–282. [[CrossRef](#)]
46. Berkhout, B.; Gagnon, A.; Silver, J.; Jeang, K.T. Efficient trans-activation by the HIV-2 tat protein requires a duplicated tar RNA structure. *Nucleic Acids Res.* **1990**, *18*, 1839–1846. [[CrossRef](#)] [[PubMed](#)]
47. Rounseville, M.P.; Kumar, A. Binding of a host cell nuclear protein to the stem region of human immunodeficiency virus type 1 trans-activation-responsive RNA. *J. Virol.* **1992**, *66*, 1688–1694. [[CrossRef](#)]
48. Ippolito, J.A.; Steitz, T.A. A 1.3-Å resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9819–9824. [[CrossRef](#)]
49. Das, A.T.; Klaver, B.; Berkhout, B. A Hairpin Structure in the R Region of the Human Immunodeficiency Virus Type 1 RNA Genome Is Instrumental in Polyadenylation Site Selection. *J. Virol.* **1999**, *73*, 81–91. [[CrossRef](#)]
50. Van Bel, N.; Ghabri, A.; Das, A.T.; Berkhout, B. The HIV-1 leader RNA is exquisitely sensitive to structural changes. *Virology* **2015**, *483*, 236–252. [[CrossRef](#)]
51. Ding, Y.; Chan, C.Y.; Lawrence, C.E. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* **2005**, *11*, 1157–1166. [[CrossRef](#)] [[PubMed](#)]
52. Mathews, D.H. Revolutions in RNA Secondary Structure Prediction. *J. Mol. Biol.* **2006**, *359*, 526–532. [[CrossRef](#)] [[PubMed](#)]

53. Kharytonchyk, S.; Monti, S.; Smaldino, P.J.; Van, V.; Bolden, N.C.; Brown, J.D.; Russo, E.; Swanson, C.; Shuey, A.; Telesnitsky, A.; et al. Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 13378–13383. [[CrossRef](#)] [[PubMed](#)]
54. Brady, S.; Singh, G.; Bolinger, C.; Song, Z.; Boeras, I.; Weng, K.; Trent, B.; Brown, W.C.; Singh, K.; Boris-Lawrie, K.; et al. Virion-associated, host-derived DHX9/RNA helicase A enhances the processivity of HIV-1 reverse transcriptase on genomic RNA. *J. Biol. Chem.* **2019**, *294*, 11473–11485. [[CrossRef](#)]
55. Bolinger, C.; Sharma, A.; Singh, D.; Yu, L.; Boris-Lawrie, K. RNA helicase A modulates translation of HIV-1 and infectivity of progeny virions. *Nucleic Acids Res.* **2010**, *38*, 1686–1696. [[CrossRef](#)]
56. Sharma, A.; Yilmaz, A.; Marsh, K.; Cochrane, A.; Boris-Lawrie, K. Thriving under Stress: Selective Translation of HIV-1 Structural Protein mRNA during Vpr-Mediated Impairment of eIF4E Translation Activity. *PLoS Pathog.* **2012**, *8*, e1002612. [[CrossRef](#)]
57. Singh, G.; Fritz, S.E.; Seufzer, B.; Boris-Lawrie, K. The mRNA encoding the JUND tumor suppressor detains nuclear RNA-binding proteins to assemble polysomes that are unaffected by mTOR. *J. Biol. Chem.* **2020**, *295*, 7763–7773. [[CrossRef](#)]
58. Childs-Disney, J.L.; Tran, T.; Vummidi, B.R.; Velagapudi, S.P.; Haniff, H.S.; Matsumoto, Y.; Crynen, G.; Southern, M.R.; Biswas, A.; Wang, Z.-F.; et al. A Massively Parallel Selection of Small Molecule-RNA Motif Binding Partners Informs Design of an Antiviral from Sequence. *Chem* **2018**, *4*, 2384–2404. [[CrossRef](#)]
59. Patwardhan, N.N.; Ganser, L.R.; Kapral, G.J.; Eubanks, C.S.; Lee, J.; Sathyamoorthy, B.; Al-Hashimi, H.M.; Hargrove, A.E. Amiloride as a new RNA-binding scaffold with activity against HIV-1 TAR. *Medchemcomm* **2017**, *8*, 1022–1036. [[CrossRef](#)]
60. Wong, R.W.; Balachandran, A.; Cheung, P.K.; Cheng, R.; Pan, Q.; Stoilov, P.; Harrigan, P.R.; Blencowe, B.J.; Branch, D.R.; Cochrane, A. An activator of G protein-coupled receptor and MEK1/2-ERK1/2 signaling inhibits HIV-1 replication by altering viral RNA processing. *PLoS Pathog.* **2020**, *16*, e1008307. [[CrossRef](#)]
61. Singh, G.; Rife, B.D.; Seufzer, B.; Salemi, M.; Rendahl, A.; Boris-Lawrie, K. Identification of conserved, primary sequence motifs that direct retrovirus RNA fate. *Nucleic Acids Res.* **2018**, *46*, 7366–7378. [[CrossRef](#)] [[PubMed](#)]
62. Sztuba-Solinska, J.; Shenoy, S.R.; Gareiss, P.; Krumpe, L.R.H.; Le Grice, S.F.J.; O’Keefe, B.R.; Schneekloth, J.J.S. Identification of Biologically Active, HIV TAR RNA-Binding Small Molecules Using Small Molecule Microarrays. *J. Am. Chem. Soc.* **2014**, *136*, 8402–8410. [[CrossRef](#)] [[PubMed](#)]
63. Abulwerdi, F.A.; Shortridge, M.D.; Sztuba-Solinska, J.; Wilson, R.; Le Grice, S.F.J.; Varani, G.; Schneekloth, J.J.S. Development of Small Molecules with a Noncanonical Binding Mode to HIV-1 Trans Activation Response (TAR) RNA. *J. Med. Chem.* **2016**, *59*, 11148–11160. [[CrossRef](#)] [[PubMed](#)]
64. Chadwick, D.R.; Level, A.M.L. Antisense RNA sequences targeting the 5’ leader packaging signal region of human immunodeficiency virus type-1 inhibits viral replication at post-transcriptional stages of the life cycle. *Gene Ther.* **2000**, *7*, 1362–1368. [[CrossRef](#)]

