A highly ordered, nonprotective MALAT1 ENE structure is adopted prior to triplex formation

MICHAEL J. YONKUNAS and NATHAN J. BAIRD
Department of Chemistry & Biochemistry, University of the Sciences, Philadelphia, Pennsylvania 19104, USA

ABSTRACT
The 3′ end of the ∼7 kb IncRNA MALAT1 contains an evolutionarily and structurally conserved element for nuclear expression (ENE) which confers protection from cellular degradation pathways. Formation of an ENE triple helix is required to support transcript accumulation, leading to persistent oncogenic activity of MALAT1 in multiple cancer types. Though the specific mechanism of triplex-mediated protection remains unknown, the MALAT1 ENE triplex has been identified as a promising target for therapeutic intervention. Interestingly, a maturation step of the nascent IncRNA 3′ end is required prior to triplex formation. We hypothesize that disruption of the maturation or folding process may be a viable mechanism of inhibition. To assess putative cotranscriptional ENE conformations prior to triplex formation, we perform microsecond MD simulations of a partially folded ENE conformation and the ENE triplex. We identify a highly ordered ENE structure prior to triplex formation. Extensive formation of U•U base pairs within the large U-rich internal loops produces a global rod-like architecture. We present a three-dimensional structure of the isolated ENE motif, the global features of which are consistent with small angle X-ray scattering (SAXS) experiments. Our structural model represents a nonprotective conformation of the MALAT1 ENE, providing a molecular description useful for future mechanistic and inhibition studies. We anticipate that targeting stretches of U•U pairs within the ENE motif will prove advantageous for the design of therapeutics targeting this oncogenic IncRNA.

Keywords: ENE; molecular dynamics; RNA dynamics; small angle X-ray scattering; IncRNA

INTRODUCTION
RNAs arising from noncoding regions comprise roughly two-thirds or more of the human genome (Woo 2018). A growing number of long noncoding RNA (IncRNA) actively alter gene expression and promote cellular dysfunction in cancer cells (Arun et al. 2018). Regulation of IncRNA transcripts is not fully understood but is thought to follow a deadenylation-dependent degradation pathway (Conrad et al. 2006; Thompson and Parker 2007; Geisler et al. 2012; Tycowski et al. 2012). Uniquely, several viral and mammalian IncRNAs, including those without a canonical poly(A) tail, maintain high transcript levels by evading degradation (Conrad et al. 2006; Brown et al. 2012; Tycowski et al. 2012, 2016; Wilusz et al. 2012). An evolutionarily conserved RNA element for nuclear expression (ENE) near the 3′ ends of these transcripts confers stability. This ENE motif functions as a cis-acting RNA stability element and has been identified across diverse genomes including viruses, plants, fungi, and mammals (Tycowski et al. 2016). Reports indicate stabilization of both IncRNA and mRNA transcripts by the ENE motif, which leads to evasion of cellular degradation pathways, resulting in high levels of transcript accumulation in the nucleus and cytoplasm (Wang et al. 1999; Conrad and Steitz 2005; Muhrad and Parker 2005; Conrad et al. 2006; Gameau et al. 2008; Brown et al. 2012; Wilusz et al. 2012). The protective function of the ENE is dependent on formation of a triplex between a large U-rich internal loop and the transcript 3′ A-rich tail (Mitton-Fry et al. 2010; Brown et al. 2012; Wilusz et al. 2012). Sequestration of the 3′ end within the triplex precludes degradation in vitro (Ageeli et al. 2018). Two ENE triplex crystal structures have revealed the RNA 3′ end embedded within a triple helical architecture (Mitton-Fry et al. 2010; Brown et al. 2012). The ENE triplex is further stabilized by stacking atop canonical duplexes forming a continuous duplex–triplex–duplex structure.

The IncRNA MALAT1 is overexpressed in many human cancer types (Ji et al. 2003; Lin et al. 2007; Lai et al. 2012) and high transcript levels are dependent on the protective ENE triplex (Brown et al. 2012; Wilusz et al. 2012).
Maturation of IncRNA MALAT1 involves cleavage of a tRNA-like domain on the 3′ end of the nascent MALAT1 transcript by RNase P followed by formation of the ENE triple helix with the mature A-rich tail. The speed of processing and triplex formation is not precisely known, but the dynamic interconversion between an isolated ENE and a well-formed triplex regulates in vitro degradation (Ageeli et al. 2018). Because the nuclear abundance of the oncocgenic MALAT1 is regulated by the ENE triplex, this small 94-nt protective element may be a viable drug target (Brown et al. 2012). Furthermore, the crystal structure of the MALAT1 ENE triplex supports direct structure–function relationship studies. Overall, the small size of the MALAT1 ENE triple helix (M1\textsuperscript{TH}) is readily tractable by both experimental and computational methods, facilitating detailed investigations of triplex-mediated RNA protection and the discovery of potential inhibitory mechanisms to ameliorate cancer progression supported by MALAT1.

Singlet structures do not define RNA functions in cells; the interplay between alternate conformations determines RNA function and ultimately cellular regulation (Russell et al. 2002; Mahen et al. 2005; Wickiser et al. 2005; Dethoff et al. 2008; Serganov and Nudler 2013; Saldi et al. 2018). With the discovery of an increasing number of IncRNAs connected to cancer and neurobiological dysfunction (Johnsson and Morris 2014; Huarte 2015; Arun et al. 2018), targeting cotranscriptional structures using high-throughput screening is increasing in popularity (Ganser et al. 2018). In fact, owing to the exceptional thermostability of M1\textsuperscript{TH} following triplex formation (Brown et al. 2012; Ageeli et al. 2018), targeting cotranscriptional structures of the MALAT1 ENE may represent a feasible approach to drug discovery, where at minimum, a putative highly ordered structure of the isolated ENE motif may become the focus of therapeutic interventions. Several questions remain unanswered regarding putative cotranscriptional structures of the ENE motif prior to triplex formation. Specifically, the absence of the A-rich tail may render the large U-rich internal loop devoid of structure or may bring the U-rich regions in close proximity to form nonnative interactions. The functional implications of either structural arrangement have not been adequately investigated. Previous reports suggest residual structure within the U-rich internal loop. A temperature-dependent peak in the NMR spectrum of an isolated ENE motif, lacking the triplex-forming A-rich sequence, was implicated in formation of U•U pairs within the U-rich loop (Mitton-Fry et al. 2010). Similarly, UV melt analysis of the isolated ENE motif indicates residual structure within the U-rich regions (Ageeli et al. 2018).

Here we directly investigate the structure and dynamics of the MALAT1 ENE triplex core (M1\textsuperscript{TH}\textsubscript{core}) and a partially folded ENE conformation (M1\textsuperscript{TH}\textsubscript{partial}) using microsecond molecular dynamics (MD) simulations. We present a quantitative description of the ENE motif structure prior to triplex formation. Limited structural flexibility within the ENE results from highly ordered U•U pairs between the U-rich regions. Small angle X-ray scattering (SAXS) experiments confirm the global properties of our simulated ENE motif structure. Importantly, our results provide a molecular description of a nonprotective conformation of the MALAT1 ENE prior to triplex formation (Ageeli et al. 2018), the structure of which represents a putative target for therapeutic intervention.

RESULTS

The 3′ terminus of the nascent MALAT1 transcript forms a tRNA-like domain. A requisite maturation step is required to form the protective triple helix. Cleavage of the tRNA-like domain by RNase P produces the mature 3′ A-rich tail and subsequent formation of the triplex (Brown et al. 2012; Wilusz et al. 2012; Ageeli et al. 2018), comprising stretches of five and four U•A-U triples interrupted by a C•G-A-C triplet and a G-C doublet (Fig. 1A,B; Brown et al. 2012; Wilusz et al. 2012). To better understand the molecular underpinnings guiding triplex formation, we performed all-atom MD simulations. Although folding of RNAs can occur on the single microsecond timescale (Hyeon and Thirumalai 2012; Hori et al. 2018), robust folding pathways require millisecond simulations. Because millisecond all-atom simulations of monitoring folding of RNAs of this size are currently not accessible, we performed comparative analyses of two different structural models of the MALAT1 triplex: (i) a structural model representing the native, functional structure of the triplex core, defined by a crystal structure (PDBID:4PLX) (Brown et al. 2012), which we refer to herein as M1\textsuperscript{TH}\textsubscript{core} (Fig. 1A,B) and (ii) a partially folded structure wherein the A-rich tail is not hybridized with the ENE, referred to as M1\textsuperscript{TH}\textsubscript{partial} (Fig. 1C). We generated the initial M1\textsuperscript{TH}\textsubscript{partial} structural model by rotating the 3′ A-rich tail (3′-tail) away from the ENE core (Fig. 1C). We reasoned that this arrangement of the tail represents a conformation that may generally describe the ENE motif and 3′ tail structures after maturation by RNase P and prior to formation of the triplex. Our initial M1\textsuperscript{TH}\textsubscript{partial} structural model treats the ENE and 3′ tail as independent structural regions within a single transcript. We use this structural model to evaluate dynamic structural changes that may occur after RNase P maturation, rather than immediately before triplex formation. Importantly, we anticipate that movements of the unstructured 3′ tail may be influenced by nucleotides within the linker. Therefore, in an effort to represent M1\textsuperscript{TH}\textsubscript{partial} in a more biologically relevant conformation space, we replaced the truncated linker sequence present in the crystal structure with the wild-type human MALAT1 sequence, which we modeled as an ideal A-form helix (see Materials and Methods). Thus, the M1\textsuperscript{TH}\textsubscript{partial} structure contains 15 linker
nucleotides, six additional nucleotides relative to the M1 THEM core. Overall, the starting structural models of M1 THEM core and M1 THEM partial have identical atomic coordinates for nucleotides 1–53 (crystallographic numbering), which comprise the ENE core (Fig. 1C). Structural distinctions between the initial models are confined between nucleotide 54 and the 3’ end of the RNA.

Intrinsic global dynamics of M1 THEM core and M1 THEM partial

MD simulations of both M1 THEM core and M1 THEM partial were monitored using RMSD of backbone atoms aligned and compared to the final equilibrated structure over simulation time at an increment of 2 psec per frame (see Materials and Methods). From a 1.2 µsec simulation of each RNA system, the RMSD of M1 THEM core (Fig. 2, black line) and M1 THEM partial (Fig. 2, red line) was calculated over an 800 and 760 nsec equilibrium trajectory, respectively (see Materials and Methods). The M1 THEM core average RMSD of 5 Å (Fig. 2A) is consistent with simulations of nonribosomal RNAs of similar size (Réblová et al. 2006; Priyakumar and MacKerell 2010; Huang et al. 2013; Aytenfisu et al. 2015; Suresh et al. 2016). The M1 THEM partial RMSD fluctuates between 15 and 30 Å during the first 100 nsec (Fig. 2A). After this initial period of fluctuating structure, M1 THEM partial continues with RMSD ~ 18 ± 4 Å over the course of the simulation. Large fluctuations would be expected for a partially folded RNA, though the global RMSD analysis does not indicate whether these structural dynamics are equally distributed throughout the M1 THEM partial structure. To address this, we also evaluated the RMSD for nucleotides 1–53, which constitute the ENE motif in both M1 THEM core and M1 THEM partial (Fig. 2A, gray and pink lines, respectively). This comparison highlights the stability of the ENE motif despite the distinct
RNA structures. While the M1\textsuperscript{TH\ partial} ENE motif (light pink) makes a conformational transition at the 100–150 nsec time frame (Fig. 2A), large increases in RMSD of the total M1\textsuperscript{TH\ partial} structure (red) can be primarily attributed to dynamics of the linker (54–70) and the 3′ tail (71–81), which are both peripheral to the ENE core (Supplemental Fig. S1). Overall, the RMSD calculations clearly indicate a globally dynamic behavior for M1\textsuperscript{TH\ partial} and a globally static behavior for M1\textsuperscript{TH\ core}.

To assess the degree of local fluctuations in the structure, we calculated the root mean square fluctuation (RMSF) by comparing the RNA backbone for each trajectory frame to the average structure from the last 800 nsec of simulation. In contrast to RMSD, the RMSF provides average dynamic information at the single nucleotide scale. The RMSF for M1\textsuperscript{TH\ core} (black line) and M1\textsuperscript{TH\ partial} (red line) are shown on the same nucleotide axes (Fig. 2B), where a dashed line indicates the truncated region of the M1\textsuperscript{TH\ core} linker (Fig. 1B,C). The fluctuations in the M1\textsuperscript{TH\ partial} structure are overall larger than M1\textsuperscript{TH\ core}. Using the lowest fluctuation for each simulation as a respective normalization (~4 Å for M1\textsuperscript{TH\ core} and ~8 Å for M1\textsuperscript{TH\ partial}, Supplemental Fig. S1), we evaluated specific regions of heightened flexibility between the two simulated RNAs. This comparison reveals reduced fluctuations in the U-rich strands (Fig. 1B,C) of M1\textsuperscript{TH\ partial} relative to M1\textsuperscript{TH\ core} (Supplemental Fig. S1). The contrasting relative fluctuation within these regions is surprising because this region is specifically involved in a three-stranded triplex in M1\textsuperscript{TH\ core} but a heretofore unknown structure in M1\textsuperscript{TH\ partial}. In fact, there are only three regions where M1\textsuperscript{TH\ partial} exhibits a higher relative fluctuation compared to M1\textsuperscript{TH\ core}: the 5′ portion of P2 (near nucleotide 15), the 3′ region of P1 (near nucleotide 50), and the 3′ A-rich tail (Fig. 2B; Supplemental Fig. S1). The increased fluctuation within P1 and P2 are localized to initial regions of the duplex-triplex junctions due to local structural rearrangements upon removal of the 3′ tail. Increased fluctuations within the 3′ tail are expected given the single-stranded nature of this region in M1\textsuperscript{TH\ partial}. Mapping the raw fluctuations (Fig. 2B) to the surface of each structure gives a view of the flexibility differences between M1\textsuperscript{TH\ core} and M1\textsuperscript{TH\ partial} (Fig. 2C). Although our simulations demonstrate large increases in both RMSD and RMSF for M1\textsuperscript{TH\ partial}, structural dynamics are limited to the single-stranded 3′ tail rather than indicative of multiple transiently populated conformations.

The M1\textsuperscript{TH\ partial} ENE motif forms distinct local interactions

Over the time-scale of our simulation M1\textsuperscript{TH\ partial} does not fold into a triplex conformation, nor does M1\textsuperscript{TH\ core} dissociate into an unfolded conformation. To assess convergence of our simulations, we use the RMS average correlation function (RAC) (Galindo-Murillo et al. 2015), where a RAC value of 0 Å represents a completely converged time-series data set (see Supplemental Methods). Convergence values of <0.6 Å and 0.1 Å for M1\textsuperscript{TH\ core} and M1\textsuperscript{TH\ partial}, respectively, were calculated (Supplemental Fig. S2). The 90% convergence value of the M1\textsuperscript{TH\ core} simulation occurs more quickly than the M1\textsuperscript{TH\ partial} simulation, demonstrating the structural integrity of our M1\textsuperscript{TH\ partial} system. We quantitatively assess each trajectory using a 2D RMSD-Rg coordinate space, where Rg represents the radius of gyration, comparing the global motion and overall size for each RNA (Supplemental Fig. S3). Overall, the M1\textsuperscript{TH\ core} and M1\textsuperscript{TH\ partial} simulations are stable around a single minimum, where structural variations are attributed to global stretching or compression dynamics (Supplemental Fig. S4; Supplemental Information).

Not surprisingly, the M1\textsuperscript{TH\ core} triplex structure is highly stable (Supplemental Fig. S3A), consistent with systematic UV melt and FRET experiments demonstrating stable triplex formation across a broad range of multiplexed monovalent and MgCl\textsubscript{2} solution conditions (Ageeli et al. 2018). To assess the level of local structural variations for the simulated M1\textsuperscript{TH\ core}, we quantify the fraction of specific base pair interactions (f\textsubscript{bp}) over simulation time (see Materials and Methods), represented on a per nucleotide basis using a heat map (Fig. 3A). Local variations near the G40–C71 base pair are quite dynamic. Specifically, U10 (U1) interacts with A68 and A69 (Tail) equivalently over simulation time. Additionally, aside from interacting with their cognate base-pairing partners, U8 and U9 in the U1 region interact with nucleotides A67 and A66 (Tail), respectively, for >20% of simulation time. These interactions constitute alternative base-pairings or “slippage” and suggest conformational dynamics in the center of the triplex. Regardless, the primary interactions within the triplex region are base-triple interactions between the Tail (nt. 65–75), U2 (nt. 36–46) and U1 (nt. 6–15). Direct interactions between U1 and U2 are minimal (<6% of simulation time). Unfolding of the triplex or dissociation of the 3′ tail does not occur during M1\textsuperscript{TH\ core} simulation. Therefore, this interaction map is consistent with a well-formed triplex, and is supported by experimental analysis wherein the triplex is resistant to 3′–5′ exoribonuclease degradation (Ageeli et al. 2018).

Previous studies have suggested that the internal loop within the U1 and U2 regions of the isolated ENE motif is highly flexible (Conrad and Steitz 2005; Conrad et al. 2006; Mitton-Fry et al. 2010; Brown et al. 2012; Tycowski et al. 2012, 2016; Donlic et al. 2018). Structural plasticity within this region would result in a dynamic, partially collapsed RNA exhibiting minimal, or highly transient, interactions involving U1 and U2 regions. To assess structural interactions, or lack thereof, within the U-rich regions of the ENE, we calculate f\textsubscript{bp} for the M1\textsuperscript{TH\ partial} simulation (Table 1; Fig. 3B). This quantitative description of base pair interactions within the M1\textsuperscript{TH\ partial} ENE motif reveals a significantly altered fingerprint when compared to M1\textsuperscript{TH\ core} (Fig. 3A). The loss of triplex interactions involving the
duplex strand (U2), Hoogsteen strand (U1), and the 3′ tail (Tail) are expected, though it is informative to note that their occupancy remains near zero for the duration of the simulation. One new region of interaction is clearly formed during simulation of M1TH partial (Fig. 3B); our analysis reveals a high degree of base pair interactions within the U-rich regions. The types of base pair interactions observed on average in the newly formed U1–U2 region from the M1TH partial simulation are listed in Table 1 and are consistent with cis-Watson–Watson interactions (cWW, Leontis–Westhof notation) observed in antiparallel RNA strands (Leontis and Westhof 2001). However, the first U•U pair between U8 and U45 is dominated by a trans-Watson–Watson interaction 69% of the time and a trans-Sugar–Watson interaction 10% of the time. The most dynamic nucleotide in the region, U9, interacts with U42, U43, and U44 in a cis-Watson–Watson configuration ranging from 13% to 55% of simulation time. The majority of remaining U•U pairs are present in cis-Watson–Watson for >99% of simulation time. This new U1–U2 duplex is formed of primarily U•U base pairs (Fig. 4C), some of which exhibit a high degree of alternative base-pair interactions. We depict the alternative pair interactions observed during MD simulations on the secondary structure of M1TH partial (Fig. 4C), where a solid blue arrow indicates the highest populated interaction (typically $f_{bp} > 0.8$) and a dotted arrow indicates transient interactions (10% < $f_{bp} < 20$%). Formation of the U•U pairs leads to a highly ordered backbone structure in which the phosphate–phosphate distances are decreased relative to a standard A-form duplex (Fig. 4D).

Our analysis indicates for the first time that the ENE maintains a significantly ordered structure, rather than adopting highly dynamic, flexible conformations as previously suggested. Our simulation results imply specific structures are maintained within the U-rich regions U1 and U2. To determine whether such structure can be experimentally corroborated, we performed SAXS experiments to compare the global structural properties of the RNA in solution with the properties of our simulated structure (see Materials and Methods). The calculated Rg of our simulated ENE is in excellent agreement with experimental SAXS results (Table 2). The experimental Rg for the isolated ENE motif is 27.6 ± 0.6 Å; the calculated Rg from our simulated ENE structure is 27 Å. Additionally, comparison of the probability distribution, or P(r), profiles from experiment and simulation (Supplemental Fig. S5) indicates that both structures are extended, mostly rod-like, with an identical maximum distance (Dmax), within error (Table 2). We attribute small differences in the probability distributions between experimental and simulated ENE motifs to differences in solvation layers between calculated and experimental systems as well as possible over-winding of the M1TH partial structure induced by the MD force-field. Overall, both the simulated and experimental ENE structures maintain a markedly extended conformation, rather than a collapsed structure. Given that MD simulations of M1TH partial result in a highly ordered ENE structure represented by a single conformational minimum, we reasoned that internal structure within the U-rich regions would be necessary to achieve such stability.

A nonprotective conformation of the MALAT1 ENE

Based on the MD-derived high probability base pairs within M1TH partial, we present a new M1TH partial secondary structure (Fig. 5A) depicting the structural interactions within the ENE motif, in which well-ordered U•U pairs are defined with $f_{bp} > 0.8$ (Fig. 4C). The structure reveals a contiguous double helical stack. This distinct conformation, compared to M1TH core (Fig. 1B), is a result of rearrangement of the U1

**FIGURE 3.** Fraction of base pairs ($f_{bp}$) observed during MD simulations. The $f_{bp}$ interactions for each nucleotide are presented as a grayscale heat map for (A) M1TH core and (B) M1TH partial. Each heat map represents a fingerprint for highly stable base pairs (black) and transient base pairs (lighter gray) throughout the MD simulation. Regions of interest are labeled with text and colored according to the structural diagrams shown in Figure 1.
and U2 strands where predominantly new U•U base pairs are formed amid two unpaired nucleotides (U43 and U44 on U2).

In light of the very limited interactions between the ENE and adjacent linker and 3′ tail, the ENE motif represents an independent structural domain, comprising only nucleotides 1–53 of our simulated structure. Consequently, we generated an average structure of the ENE motif from the last 500 nsec of production simulation. However, this average ENE structure lacks the wild-type P2 sequence, which was truncated to facilitate crystallization (Fig. 1A, B). Therefore, to achieve a wild-type model of the ENE motif structure, we modeled the wild-type P2 helix as an A-form helix (Fig. 5B). After manually fitting the P2 helix to the apex of the simulated ENE motif structure, the system was re-solvated and equilibrated for 2 nsec to achieve a final structure of the wild-type ENE (Fig. 5C). Theoretical scattering calculations of the wild-type ENE structure (ENE + WT P2) are in excellent agreement with experimental SAXS measurements of the wild-type ENE motif RNA of the same sequence (Table 2; Supplemental Fig. S5). Overall, the SAXS data indicate an extended structure consistent with the elongated duplex-like stack within the ENE + WT P2 structure. It is clear from the new

| Interaction | f_{bp} | Leontis-Westhof type |
|-------------|--------|----------------------|
| 15          | UU     | 99 cWW               |
| 14          | UU     | 99 cWW               |
| 13          | UU     | 99 cWW               |
| 12          | UU     | 99 cWW               |
| 11          | CG     | 88 cWW               |
| 10          | UC     | 70 cWW               |
| 10          | UU     | 17 cWW               |
| 9           | UU     | 55 cWW               |
| 9           | UU     | 13 cSW, cWW          |
| 9           | UU     | 19 cWW               |
| 8           | UU     | 86 tSW, tWW, cWW     |
| 8           | UG     | 17 cWW               |
| 7           | UG     | 99 cWW               |
| 6           | UG     | 99 cWW               |

f_{bp} is the fraction of base pairs observed over simulation time for the U1, U2, Linker, and Tail regions of M1TH partial as shown in Figure 3B. Types are determined using Leontis-Westhof classification (Leontis and Westhof 2001).

**FIGURE 4.** Base-pair interactions in the M1^{TH} ENE motif. (A) Example of U•U pair observed in M1^{TH} simulation at site U8–U45 in a trans-Watson–Crick interaction. The hydrogen bonds in this base pair are present for more than 86% of simulation time. (B) A snapshot of a dynamic region within the ENE including nucleotides U7–U10 on U1 and C41–G47 on U2. Blue dotted lines indicate potential hydrogen bonding. This view effectively depicts the promiscuity of U9, which is within reach of U42 (f_{bp} ∼ 55%), U43 (f_{bp} ∼ 13%), and U44 (f_{bp} ∼ 19%) (Table 1). (C) Zoomed in partial secondary structure of the ENE motif (Fig. 1C) with f_{bp} (Table 1) mapped to each nucleotide pair where a solid blue arrow indicates the highest populated interaction (typically f_{bp} > 80%) and a dotted arrow indicates transient interactions (10% < f_{bp} < 20%). (D) For each base pair detected, the average phosphate–phosphate distance (P–P) was calculated over the MD simulation time. The right axis corresponds to M1^{core} base pairs (black line), and the left axis corresponds to M1^{TH} partial base pairs (red line). Missing values at bulged nucleotides are closed with dotted lines. The horizontal blue dashed lines indicate a structural transition between U1–U2 region with the canonical A-form duplexes (P1 and P2) within each structure. The average phosphate–phosphate distance for helical A-form RNA is indicated by a purple line (van Knippenberg and Hilbers 1986).
structure of the wild-type ENE (Fig. 5C) that the newly formed U•U base pairs tolerate a non-A-form helical conformation (Fig. 4D). Stability of this region is maintained by a well-ordered helical U•U stack. This unique topology, while rarely observed experimentally (Baeyens et al. 1995), is highly stable, in contrast to structures containing singular U•U base pairs whose stability is weakened relative to U•G or C•A mismatches (Kierzek et al. 1999; Sheng et al. 2013).

DISCUSSION

Many biological processes utilize alternative conformers to regulate RNA function (Russell et al. 2002; Mahen et al. 2005; Wickiser et al. 2005; Dethoff et al. 2008; Serganov and Nudler 2013; Saldi et al. 2018). Cotranscriptional RNA folding may populate alternative structures including nonproductive conformations (Hua et al. 2018). For the IncRNA MALAT1, 3’ end maturation by RNase P is a requisite step for proper formation of a protective 3’ ENE triplex, which supports high cellular abundance and persistent IncRNA oncogenic processes (Brown et al. 2012; Wilusz et al. 2012). Therefore, a cotranscriptionally folded ENE motif, if highly structured prior to triplex formation, would represent an important non-functional structure for therapeutic intervention. Recently, we investigated the mechanism of M1TH triplex formation using a battery of biochemical and biophysical approaches. We determined that prior to liberation of the 3’ end upon RNase P cleavage, a stable P2 and a weak P1 helix are partially formed, suggesting a metastable ENE structure exists prior to triplex formation. (Ageeli et al. 2018). Protection from degradation is conferred through formation of the triplex (Brown et al. 2016; Ageeli et al. 2018). Dynamic structural interconversions between the isolated ENE structure and the well-formed triplex regulate triplex-mediated protection in vitro. Here, we perform microsecond MD simulations to investigate structural variations within the ENE triplex and the isolated ENE motif. We assess the conformational dynamics of the MALAT1 ENE prior to triplex formation, which we refer to as M1TH\textsuperscript{partial}. Our quantitative results support a highly ordered ENE structure comprising nonnative U•U pairs, leading to an overall extended structure. The global properties of this ENE structure are supported by SAXS R\textsuperscript{g} and P(r) data, confirming the extended duplex conformation of the MALAT1 ENE. While the ENE triplex of M1TH\textsuperscript{core} and isolated ENE of M1TH\textsuperscript{partial} are identical at the nucleotide level, no tendency to convert between conformations was observed on this timescale. A quantitative assessment of convergence shows a converging trajectory, suggesting both conformations are indeed distinct stable conformational states. Therefore, the stable MALAT1 ENE may be entirely formed cotranscriptionally; prior to 3’ end maturation of the nascent MALAT1 transcript, the two duplexes and U-rich regions form a stable, extended rod-like structure, which is insufficient to protect MALAT1 from 3’ to 5’ exoribonucleolytic degradation. We present the first evidence of a highly ordered ENE motif stabilized by nonnative U•U pairs. This stable ENE structure unveils a novel putative therapeutic target. Recent studies have utilized numerous small molecule designs to target U•U base pairs (Arambula et al. 2009; Xia et al. 2014). Similar efforts targeting the ENE structure may yield effective inhibition of the IncRNA MALAT1.

**FIGURE 5.** A structure of the MALAT1 ENE motif. (A) The secondary structure of the M1TH\textsuperscript{partial} determined by simulation highlights structure within the U1 and U2 regions; blue lines show through-space tertiary interactions. (B) Secondary structure of the isolated wild-type MALAT1 ENE structure including the full-length P2 helix. (C) Three-dimensional model of wild-type MALAT1 ENE shows a non-A-form helical stack in an extended rod-like conformation consistent with SAXS measurements (Table 2; Supplemental Fig. S5).
TABLE 2. Experimental and theoretical SAXS calculations

| RNA                                | \( R_g \)  | \( D_{\text{max}} \)  |
|------------------------------------|------------|------------------------|
| Experimental ENE (WT)              | 27.6 (0.6) | 93 (6)                 |
| Simulated ENE (truncated P2)       | 23 (0.7)  | 83 (3)                 |
| Simulated ENE + WT P2              | 27 (1)    | 92 (3)                 |

The radius of gyration (\( R_g \)) and maximum intramolecular distance (\( D_{\text{max}} \)) were determined experimentally using SAXS or calculated from structural models using CRYOSOL (Svergun et al. 1995) and GNOM (Svergun 1992).

MATERIALS AND METHODS

Simulation system preparation

We simulated two different RNA systems, which we refer to as M1\textsuperscript{TH\_core} and M1\textsuperscript{TH\_partial}. The M1\textsuperscript{TH\_core} structure includes the triplex region, a truncated P2 helix, and truncated linker sequence. The atomic coordinates for M1\textsuperscript{TH\_core} were obtained from the only available crystal structure of the MALAT1 triplex (Brown et al. 2012). For this study, we used chain A of PDB:4PLX (Brown et al. 2012) and placed under harmonic restraints during minimization and initial equilibration. Approximately, 800 nsec of equilibrium MD from each simulation system was used in subsequent analyses as determined by RMSD calculations. Translational and rotational motions were removed from the trajectory by fitting all frames to the first frame of the analysis data set using CPPTRAJ in AMBERTOOLS 16 (Case et al. 2016). The M1\textsuperscript{TH\_core} system was simulated for 1.2 µsec. The M1\textsuperscript{TH\_partial} system was simulated for 1 µsec.

The root mean square deviation (RMSD) was calculated as

\[
\text{RMSD} = \left[ \frac{1}{N} \sum_{i=1}^{N} \left| \text{r}_i(t) - \text{r}_i^{\text{ref}}(t) \right|^2 \right]^{1/2},
\]

where \( N = \sum_{t}^{n} \), \( \text{r}_i(t) \) are the coordinates of atom \( i \) at time \( t \) after fitting to the reference coordinates \( \text{r}_i^{\text{ref}}(t) \), for which the first frame of production simulation was used. The root mean square fluctuation (RMSF) was calculated per nucleotide as

\[
\text{RMSF} = \left[ \frac{1}{T} \sum_{t=1}^{T} \left| \text{r}_i(t) - \overline{\text{r}}_i \right|^2 \right]^{1/2},
\]

where \( T \) is the time and \( \overline{\text{r}}_i \) is the average position of particle \( i \) over \( T \).

Fraction of base pairs

Base pairs were identified using the DSSR package (Lu et al. 2015) of the 3DNA program (Lu and Olson 2003). All MD trajectories were read into DSSR using the -nmr flag, all 29 base pair references were searched at every frame using the -more option and output as a JSON formatted file. Analyses were performed using custom R (R Core Team 2018) scripts. The fraction of base pair, \( f_{bp} \), is calculated as

\[
f_{bp} = \frac{N_{bp}}{N},
\]

where \( N_{bp} \) is the number of frames in a simulation a base pair is observed, and \( N \) is the total number of frames in the simulation. Heat map plots were constructed using R (R Core Team 2018).
Small angle X-ray scattering experiments

In vitro transcription of the wild-type isolated ENE motif was performed using PCR templates containing two 2′OMe at the 5′ end of the template strand as previously described (Ageelli et al. 2018). The sequence of the isolated ENE motif was GGAAAGTTTTTTCTTTCTGAGAAGAACACGTTATTTCTTCGAGTTTTGCTTTTGCCCTTT. Samples were purified by size-exclusion chromatography in 20 mM HEPES, pH 7.4, 150 mM NaCl and KCI, and 1 mM MgCl₂. Data were collected under continuous flow on beamline ID-12 at the Advanced Photon Source (APS) at Argonne National Laboratories, controlled and collected using beamline-specific programs and scripts. Data was collected for buffer samples and RNA samples separately. Igor Pro (WaveMetrics) was used to examine data quality. A Guinier analysis was used to calculate R₀ and estimate error from scattering intensities. Initial calculations of the pair-distance probability distribution, P(r), were conducted using a scan of the maximum molecule distance Dmax using Python scripts (Lipfert et al. 2007) and GNOM (Svergun 1992) within the ATLAS analysis package (embil-hamburg.de). Dmax was determined based on a smooth plot of P(r) where Dmax corresponds to an abscissa intercept at r=0. The error in Dmax is reported as the interval over which Dmax results in a P(r) function with a smooth intercept at r=0.

Theoretical scattering calculations

Coordinates were extracted from single frames of MD simulations and imported into CRYSOl (Svergun et al. 1995) where default parameters were selected. The calculated R₀ and maximum molecular diameter excluding hydrogen volume were recorded. Theoretical intensities generated by CRYSOl (Svergun et al. 1995) were imported into GNOM (Svergun 1992) where the predicted maximum molecular diameter is input as Dmax. A scan of P (r), the pair distribution function, was performed at Dmax ±5 Å based on a smooth plot of P(r) where Dmax corresponds to an abscissa intercept at r=0. The error in Dmax calculated from theoretical scattering profiles generated from CRYSOl are not explicitly reported by GNOM due to the lack of experimental error in the scattering intensities. A 3% error was calculated and added to the theoretical intensities. The error in Dmax is reported as the interval over which Dmax results in a P(r) function with a smooth intercept at r=0.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We would like to thank Dr. Kayleigh McGovern-Gooch and Abeer Ageelli for helpful discussions. We gratefully acknowledge use of the SAXS core facility of the Center for Cancer Research, National Cancer Institute (NCI). The shared scattering beamline 12-ID-B resource is allocated under the PUP-24152 agreement between the NCI and Argonne National Laboratory (ANL). We thank Dr. Lixin Fan (NCI) and Dr. Xiaobing Zuo (ANL) for their expert support. Use of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility, was operated for the DOE Office of Science by ANL under contract no. DE-AC02-06CH11357. Support was also provided by University of the Sciences (to N.J.B.) and National Institutes of Health (K22-HL121113A to N.J.B.).

Received December 8, 2018; accepted May 13, 2019.

REFERENCES

Ageelli AA, McGovem-Gooch KR, Kaminska MM, Baird NJ. 2018. Finely tuned conformational dynamics regulate the protective function of the IncRNA MALAT1 triple helix. Nucleic Acids Res 47: 1468–1481. doi:10.1093/nar/gky1171

Arambula JF, Ramisetty SR, Baranger AM, Zimmerman SC. 2009. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. Proc Natl Acad Sci 106: 1668–1673. doi:10.1073/pnas.0901824106

Arun G, Diermeier SD, Spector DL. 2018. Therapeutic targeting of long non-coding RNAs in cancer. Trends Mol Med 24: 257–277. doi:10.1016/j.molmed.2018.01.001

Aytenfisu AH, Liberman JA, Wedekind JE, Mathews DH. 2015. Molecular mechanism for preQ₁–II riboswitch function revealed by molecular dynamics. RNA 21: 1898–1907. doi:10.1261/rna.051367.115

Baeyens KJ, De Bondt HL, Holbrook SR. 1995. Structure of an RNA double helix including uracil–uracil base pairs in an internal loop. Nat Struct Mol Biol 2: 56–62. doi:10.1038/nsb0195-56

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. J Chem Phys 81: 3684–3690. doi:10.1063/1.448118

Brown JA, Valenstein ML, Yario TA, Tycowski KT, Steitz JA. 2012. Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENI noncoding RNAs. Proc Natl Acad Sci 109: 19202–19207. doi:10.1073/pnas.1217338109

Brown JA, Kinzig CG, DeGregorio SJ, Steitz JA. 2016. Hoogsteen-pyrimidines promote the stability and function of the MALAT1 RNA triple helix. RNA 22: 743–749. doi:10.1261/rna.055707.115

Case DA, Ben-Shalom Y, Brozell SR, Cheatham TE III, Case DA, Brozell SR, Cerutti DS, Cheatham TE III, Conrad NK, Mili S, Marshall EL, Shu M-D, Steitz JA. 2006. Finely tuned conformational dynamics regulate the protective function of the IncRNA MALAT1 triple helix. Mol Cell 24: 943–953. doi:10.1016/j.molcel.2006.10.029

Dethoff EA, Hansen AL, Musselman C, Watt ED, Andricioaei I, Al-Hashimi HM. 2008. Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. Biophys J 95: 3906–3915. doi:10.1529/biophysj.108.140285

Donlic A, Morgan BS, Xu JL, Liu A, Roble C, Hargrove AE. 2018. Discovery of small molecule ligands for MALAT1 by tuning an RNA-binding scaffold. Angew Chem Int Ed Engl 57: 13242–13247. doi:10.1002/anie.201808823

Galindo-Murillo R, Roe DR, Cheatham TE. 2015. Convergence and reactivity in molecular dynamics simulations of the DNA duplex d(GCAGCAAGCAGCAGAAGG). Biochim Biophys Acta 1850: 1041–1058. doi:10.1016/j.bbagen.2014.09.007

Ganser LR, Lee J, Rangadurai A, Merriman DK, Kelly ML, Kansal AD, Sathyamoorthy B, Al-Hashimi HM. 2018. High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble. Nat Struct Mol Biol 25: 425–434. doi:10.1038/s41594-018-0062-4

Gameau NL, Sokoloski KJ, Opyrchal M, Neff CP, Wilusz CJ, Wilusz J. 2008. The 3′ untranslated region of sindbis virus represses
deadenylation of viral transcripts in mosquito and mammalian cells. J Virol 82: 880–892. doi:10.1128/JVI.01205-07

Geisler S, Lojek L, Khalil AM, Baker KE, Coller J. 2012. Decapping of long noncoding RNAs regulates inducible genes. Mol Cell 45: 279–291. doi:10.1016/j.molcel.2011.11.025

Hor N, Denesyuk NA, Thirumalai D. 2018. Frictional effects on RNA folding; speed limit and kramers turnover. J Phys Chem B. doi:10.1021/acs.jpcb.8b07129

Hua B, Panja S, Wang Y, Woodson SA, Ha T. 2018. Mimicking co-transcriptional RNA folding using a superhelicase. J Am Chem Soc 140: 10067–10070. doi:10.1021/acs.jacs.8b03784

Huang W, Kim J, Jha S, Aboul-ela F. 2013. The impact of a ligand binding on strand migration in the SAM-I riboswitch. PLoS Comput Biol 9: e1003069. doi:10.1371/journal.pcbi.1003069

Huarte M. 2015. The emerging role of IncRNAs in cancer. Nat Med 21: 1253–1261. doi:10.1038/nm.3981

Humphreys W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. J Mol Graph 14: 33–38. doi:10.1016/0263-7855(96)00018-5

Hyeon C, Thirumalai D. 2012. Chain length determines the folding rates of RNA. Biophys J 102: L1–L13. doi:10.1016/j.bpj.2012.01.003

Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, et al. 2003. MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene 22: 8031–8041. doi:10.1038/sj.onc.1206928

Johnsson P, Morris KV. 2014. Expanding the functional role of long noncoding RNAs. Cell Res 24: 1284–1285. doi:10.1038/cr.2014.104

Kierzek R, Burkard ME, Turner DH. 1999. Thermodynamics of single mismatches in RNA duplexes. Biochemistry 38: 14214–14223. doi:10.1021/bi991186l

Lai M, Yang Z, Zhou L, Zhu Q, Xie H, Zhang F, Wu L, Chen L, Zheng S. 2012. Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. Med Oncol 29: 1810–1816. doi:10.1007/s12332-011-0004-z

Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. RNA 7: 499–512. doi:10.1017/S1355838201002515

Lin R, Maeda S, Liu C, Karin M, Edgington TS. 2007. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. Oncogene 26: 851–858. doi:10.1038/sj.onc.1209846

Lipfert J, Das R, Chu VB, Kudaravalli M, Boyd N, Herschlag D. 2005. The yeast EDC1 mRNA undergoes deadenylation-independent decapping stimulated by Not2p, Not4p, and Not5p. EMBO J 24: 1033–1045. doi:10.1038/sj.emboj.7600560

Muhlrad D, Parker R. 2005. The yeast EDC1 mRNA undergoes deadenylation-independent decapping stimulated by Not2p, Not4p, and Not5p. EMBO J 24: 1033–1045. doi:10.1038/sj.emboj.7600560

Priyakumar UD, MacKerell AD. 2010. Role of the adenine ligand on the stabilization of the secondary and tertiary interactions in the adenine riboswitch. J Mol Biol 396: 1422–1438. doi:10.1016/j.jmb.2009.12.024

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rêblova K, Lankâs F, Râžga F, Krasovska MV, Koča J, Šponer J. 2006. Structure, dynamics, and elasticity of free 16s rRNA helix 44 studied by molecular dynamics simulations. Biopolymers 82: 504–520. doi:10.1002/bip.20503

Russell R, Zhuang X, Babcock HP, Millett IS, Doniach S, Chu S, Herschlag D. 2002. Exploring the folding landscape of a structured RNA. Proc Natl Acad Sci U S A 99: 155–160. doi:10.1073/pnas.221593598

Rycckaert J-P, Ciccotti G, Berendsen HJ. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23: 327–341. doi:10.1002/1097-1758(199709098-5

Salid T, Fong N, Bentley DL. 2018. Transcription elongation rate affects nascent histone pre-mRNA folding and 3′ end processing. Genes Dev 32: 297–308. doi:10.1101/gad.310896.117

Serganov A, Nucler E. 2013. A decade of riboswitches. Cell 152: 17–24. doi:10.1016/j.cell.2012.12.024

Sheng J, Gan J, Soares AS, Salon J, Huang Z. 2013. Structural insights of non-canonical U-U pair and Hoogsteen interaction probed with Se nucleic acid. Nucleic Acids Res 41: 10476–10487. doi:10.1093/nar/gkt799

Suresh G, Srinivasan H, Nanda S, Priyakumar UD. 2016. Ligand-induced stabilization of a duplex-like architecture is crucial for the switching mechanism of the SAM-III riboswitch. Biochemistry 55: 3349–3360. doi:10.1021/acs.biochem.5b00973

Svergun D. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 25: 495–503. doi:10.11002/1097-8898992021663

Svergun D, Barberato C, Kach MHJ. 1995. CRYSO – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr 28: 768–773. doi:10.11002/1097-8898992007047

Thompson DM, Parker R. 2007. Cytoplasmic decay of intergenic transcripts in Saccharomyces cerevisiae. Mol Cell Biol 27: 92–101. doi:10.1128/MCB.01023-06

Tycowski KT, Shu M-D, Borah S, Shi M, Steitz JA. 2012. Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. Cell Rep 2: 26–32. doi:10.1016/j.celrep.2012.05.020

Tycowski KT, Shu M-D, Steitz JA. 2016. Myriad triple-helix-forming structures in the transposable element RNAs of plants and fungi. Cell Rep 15: 1266–1276. doi:10.1016/j.celrep.2016.04.010

van Knippenberg PH, Hilbers CW. 1986. Structure and dynamics of RNA. Springer, Boston. doi:10.1007/978-1-4684-5173-3

Wang Z, Day N, Trifillis P, Kiledjian M. 1999. An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA in vitro. Mol Cell Biol 19: 4552–4560. doi:10.1128/MCB.19.7.4552

Wickiser JK, Winkler WC, Breakey RR, Crothers DM. 2005. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. Mol Cell 18: 49–60. doi:10.1016/j.molcel.2005.02.032

Wilsue JE, Freier SM, Spector DL. 2008. 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell 135: 919–932. doi:10.1016/j.cell.2008.10.012

Wilsue JE, JnBaptiste CK, Lu LY, Kuhn C-D, Joshua-Tor L, Sharp PA. 2012. A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. Genes Dev 26: 2392–2407. doi:10.1101/gad.204438.112

Woo CJ. 2018. The therapeutic targeting of long noncoding RNA. In RNA therapeutics, Vol. 27 of topics in medicinal chemistry (ed. Garner AL), pp. 207–235. Springer International Publishing, Cham, Switzerland.

Xia X, Piao X, Fredrick K, Bong D. 2014. Bifacial PNA complexation inhibits enzymatic access to DNA and RNA. Chembiochem 15: 31–36. doi:10.1002/cbic.201300536