Structural Insights of Non-canonical U*U Pair and Hoogsteen Interaction Probed with Se Atom

Jia Sheng  
*Georgia State University*, jsheng1@student.gsu.edu

Jianhua Gan  
*Georgia State University*, ganjhh@fudan.edu.cn

Alexei S. Soares  
*Brookhaven National Laboratory*, soares@bnl.gov

Jozef Salon  
*Georgia State University*, jsalon@gsu.edu

Zhen Huang  
*Georgia State University*, huang@gsu.edu

Follow this and additional works at: [https://scholarworks.gsu.edu/chemistry_facpub](https://scholarworks.gsu.edu/chemistry_facpub)

Recommended Citation  
Structural insights of non-canonical U*U pair and Hoogsteen interaction probed with Se atom Nucl. Acids Res. (2013) 41 (22): 10476-10487 first published online September 5, 2013 doi:10.1093/nar/gkt799

This Article is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Faculty Publications by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
Structural insights of non-canonical U\textbullet{}U pair and Hoogsteen interaction probed with Se atom

Jia Sheng\textsuperscript{1}, Jianhua Gan\textsuperscript{1}, Alexei S. Soares\textsuperscript{2}, Jozef Salon\textsuperscript{1} and Zhen Huang\textsuperscript{1,*}

\textsuperscript{1}Department of Chemistry, Georgia State University, Atlanta, GA, 30303, USA and \textsuperscript{2}Department of Biology, Brookhaven National Laboratory, Upton, NY, 11973, USA

Received May 16, 2013; Revised August 1, 2013; Accepted August 14, 2013

ABSTRACT

Unlike DNA, in addition to the 2'-OH group, uracil nucleobase and its modifications play essential roles in structure and function diversities of non-coding RNAs. Non-canonical U\textbullet{}U base pair is ubiquitous in non-coding RNAs, which are highly diversified. However, it is not completely clear how uracil plays the diversifying roles. To investigate and compare the uracil in U-A and U\textbullet{}U base pairs, we have decided to probe them with a selenium atom by synthesizing the novel 4-Se-uridine (\textsuperscript{Se}U) phosphoramidite and Se-nucleobase-modified RNAs (\textsuperscript{Se}U-RNAs), where the exo-4-oxygen of uracil is replaced by selenium. Our crystal structure studies of U-A and U\textbullet{}U pairs reveal that the native and Se-derivatized structures are virtually identical, and both U-A and U\textbullet{}U pairs can accommodate large Se atoms. Our thermostability and crystal structure studies indicate that the weakened H-bonding in U-A pair may be compensated by the base stacking, and that the stacking of the trans-Hoogsteen U\textbullet{}U pairs may stabilize RNA duplex and its junction. Our result confirms that the hydrogen bond (O4'–H–C5) of the Hoogsteen pair is weak. Using the Se atom probe, our Se-functionality studies reveal more insights into the U\textbullet{}U interaction and U-participation in structure and function diversification of nucleic acids.

INTRODUCTION

Unlike natural DNA, which merely stores genetic information in cells (1), natural RNA is highly diversified in structure and function. Because of the RNA diversity, RNA plays essential functions in cells and expands complexity of living systems by serving as genetic information carrier, catalyst and regulator (2–10). Recently, tremendous functional RNAs have been discovered as non-coding RNAs (ncRNA), such as ribozymes, riboswitches, small interfering RNA (siRNA), microRNA (miRNA), small nuclear RNA (snRNA) and RNAs regulating biological pathways. ncRNAs can control gene expressions selectively through transcription and translation regulations (11,12), participate in chromatin silencing and remodeling (13), regulate the retroviruses activity (14), catalyze biochemical reactions (15,16), recognize metabolites (17), as well as facilitate gene function study and drug discovery (18,19). ncRNAs play highly specific roles by folding into various 3D structures and binding specifically with other molecules or ligands (such as proteins and metabolites), which may trigger cascades of biological events.

However, considering the similar chemical structures of nucleic acid building blocks (such as almost the same nucleobases in RNA and DNA), it is striking that RNA with the extra 2'-OH is able to establish much more diversified structures and functions than DNA (20,21). In addition to the 2'-OH group, it appears that the RNA modifications and non-canonical base pairings are the two major strategies to overcome the structural homogeneity limit caused by the four similar nucleobases and to achieve huge diversities in both structure and function (22–24). Especially, uracil nucleobase can form multiple non-canonical base pairings and play essential roles in diversifying RNA structure and function. Non-canonical U\textbullet{}U base pair is ubiquitous in ncRNA, and Watson–Crick U-A pair can often be replaced with U-G wobble pair without significant duplex destabilization, which increases structure and function diversity of ncRNAs. U\textbullet{}U pairs are often observed in RNA duplex junction and loops (25–27), whereas U-A pair is normally not formed at these places. Replacing U-A pair in duplex with U\textbullet{}U pair significantly destabilizes the duplex structure. It is not completely clear how uracil plays the diversifying roles in these base pairs to achieve the structure and function diversity. To investigate and compare the uracil roles played in these non-canonical and canonical pairs, we have decided to probe the U\textbullet{}U and U-A pairs with a Se atom, where the exo-4-oxygen of uracil is replaced by selenium.

*To whom correspondence should be addressed. Tel: +1 404 413 5535; Fax: +1 404 413 5535; Email: huang@gsu.edu

© The Author(s) 2013. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Though 4-Se-uridine was synthesized over three decades ago (28,29), it has not been incorporated into RNAs because of the synthetic challenges. Recently, our successes on the synthesis and biophysical studies of the Se-nucleobase modifications (30–35) have encouraged us to overcome the SeU-RNA synthesis challenge, meet the urgent needs in ncRNA investigation and probe U-A and U*U pairs by a Se atom. Herein, we report the first synthesis of the 4-Se-uridine phosphoramidite (SeU) and the corresponding SeU-RNAs by replacing 4-oxygen with selenium. We have found that this Se-modification does not cause significant perturbation and that the native and modified structures are virtually identical. We also found that via the stacking and hydrogen bonding, the uracil nucleobase interacts differently in RNA duplex and duplex junction. Moreover, the accommodation of the larger selenium atom by both U-A and U*U pairs implies the RNA flexibility. Our studies suggest that by presenting their different faces and edges, uracil and uridine are capable of diversifying structure and function of ncRNAs. Furthermore, this Se-modified uridine offers the Se-RNAs with additional UV absorption (λmax = 370 nm; ε: 1.30 × 10^4 M⁻¹·cm⁻¹). Excitingly, after a single-oxygen atom replacement with selenium, we have observed for the first time the color RNAs (light yellow) as well as color RNA crystals (dark yellow). The color property of the SeU-RNAs is unique and has great potentials in RNA visualization, detection, spectroscopic study and crystallography of RNAs and protein-RNA complexes and interactions, demonstrating the usefulness of selenium-derivated nucleic acids (SeNA) (36,37) in structural biology. In addition, both the anomalous phasing and molecular replacement approaches result in the identical crystal structures. Our new method provides a unique atomic tool for probing structure and function of ncRNAs and their protein complexes.

**MATERIALS AND METHODS**

**Synthesis of the 4-Se-uridine phosphoramidite**

3-(1-((2R,3S,4S,5R)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-(tert-butylidimethyl-silyloxy)-4-hydroxy-tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidin-4-ylselanyl)propanenitrile

To a dry THF solution (10 ml) of the starting material compound (1, 1.34 g, 2 mmol), 4,4'-dimethylenopyridine (24.5 mg, 0.2 mmol) and triethylamine (0.56 ml, 4 mmol) under argon, the dry tetrahydrofuran (THF) solution (10 ml) of 2,4,6-trisopropylbenzenesulfonyl chloride (906 mg, 0.38 ml) was added slowly into this selenide solution. After the selenium incorporation was completed in 45 min (monitored on TLC, 5% MeOH in CH₂Cl₂, product Rf = 0.60), water (100 ml) was added to the reaction flask. The solution was adjusted to pH 7–8 using CH₃COOH (10%) and was then extracted with ethyl acetate (3 × 100 ml). The organic phases were combined, washed with NaCl (sat., 100 ml), dried over MgSO₄ (s) for 30 min and evaporated to minimum volume under reduced pressure. The crude product was then dissolved in methylene chloride (5 ml) and purified on a silica gel column equilibrated with hexanes/methylene chloride (1:1). The column was eluted with a gradient of methylene chloride (CH₂Cl₂, 0.5%, 1% and 2% MeOH in CH₂Cl₂, 300 ml each). After the collected fraction evaporation and dry under high vacuum, pure compound 2 was obtained as a slightly yellow foam product (1.27 g, 81% yield).

1H-NMR (400 MHz, CDCl₃) δ: 2.05 (s, 3H, CH₃), 0.38 (s, 3H, CH₃), 0.95 (s, 6H, 2×CH₃), 2.31–2.37 (m, 1H, H-2'), 3.00 (dd, J = 6.5 and 6.7 Hz, 2H, CH₂-Se), 3.37-3.41 (m, 2H, CH₂-CN), 3.50-3.52 (m, 2H, 1H-5′), 3.81 (s, 6H, 2×OCH₃), 4.17–4.22 (m, 1H, H-3′), 4.31 (s, 1H, 3′-OH), 4.40–4.50 (m, 1H, H-4′), 5.78 (s, 1H, H-1′), 5.90 (d, 1H, J = 6.8 Hz, H-5), 6.8–6.90 (m, 4H, aromatic), 7.20–7.46 (m, 9H, aromatic), 8.31 (d, 1H, J = 6.8 Hz, H-6). 13C-NMR (100 MHz, CDCl₃) δ: –4.30, –4.40 (CH₃), 18.1 (CH₂-CN), 19.0 (CH₂-Se-CN), 20.5 [(CH₃)₂C(Ct-Bu)], 25.9 (CH₃), 55.3 (OCH₃), 68.7 (C-3′), 76.4 (C-2′), 83.1 (C-4′), 91.0 (C-1′), 106.0 (C-5), 118.8 (CN), 113.3, 127.1, 128.0, 128.2, 130.1, 135.0, 135.3, 144.2, 158.7 (Ar-C), 140.4 (C-6), 153.3 (C-2), 175.0 (C-4). HRMS (ESI-TOF): molecular formula, C₃₉H₅₉N₃O₇SeSi; [M+H]^+ : 778.2413 (calc: 778.2426).

(2R,3S,4S,5R)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-(tert-butylidimethyl-silyloxy)-4-(2-cyanoethylselenanyl)-2-oxopyrimidin-1(2H)-yl)-tetracydrofuran-3-yl-2-cyanoethyl disopropylphosphoramidite

To the flask (25 ml) containing 2 (453 mg, 0.68 mmol) under argon, dry methylene chloride (2.5 ml), N,N-diisopropylethylamine (0.17 ml, 1.03 mmol, 1.5 eq.), and 2-cyanoethy N,N-diisopropyl-chlorophosphoramidite (195 mg, 0.83 mmol, 1.2 eq.) were added sequentially (3). The reaction mixture was stirred at −10°C in an ice-salt bath under argon for 10 min, followed by removal of the bath. The reaction was completed in 2 h at room temperature, generating a mixture of two diastereomers (indicated by TLC, 5% MeOH in CH₂Cl₂, product Rf = 0.63 and 0.68). The reaction was then quenched with NaHCO₃ (5 ml, sat.) and stirred for 5 min, followed by the extraction with CH₂Cl₂ (3 × 8 ml). The combined organic layer was washed with NaCl (10 ml, sat.) and dried over MgSO₄ (s) for 30 min, followed by filtration. The solvent was then evaporated under reduced pressure, and the crude product was re-dissolved in CH₂Cl₂ (2 ml). This solution was added drop-wise to cold petroleum ether (or hexane) (200 ml) under vigorous stirring, generating a white precipitate. The petroleum ether layer was decanted. The crude product was re-dissolved again in CH₂Cl₂ (2 ml) and then loaded on Al₂O₃ column (neutral) that was equilibrated with CH₂Cl₂/Hexanes (1:1). The column
was eluted with a gradient of methylene chloride and ethyl acetate [CH₂Cl₂ to CH₃Cl/EtOAc (7:3)]. After solvent evaporation and dry over high vacuum, the compound 3 (612 mg) was obtained as a white foamy product (92% yield). ¹H-NMR (400 MHz, CDCl₃, two sets of signals from a mixture of two diastereomers) δ: 0.2–0.4 (m, 12H, 4× CH₃), 0.85–1.20 [m, 36H, 8× CH₃-ipl and 4× Si(CH₃)], 2.30–2.38 and 2.70–2.82 (2× m, 4H, 2× H-2'), 2.34 and 2.64 (2× t, J = 6.4 Hz, 4H, 2× O-CH₂-CH₂-CN), 3.00–3.04 (m, 4H, 2× Se-CH₂-CH₂-CN), 3.32–3.44 (m, 6H, 2× H-5', 2× Se-CH₃), 3.52–3.64 (m, 8H, 4× CH-ipl, 2× O-CH₂-CH₂-CN), 3.73–3.84 (m, 2H, 2× H-5'), 3.82 and 3.83 (2× s, 12H, 4× OCH₃), 4.12–4.35 (m, 2H, 2× H-3'), 4.43–4.48 (m, 2H, 2× H-4'), 5.70–5.90 (m, 4H, 2× H-5' and 2× H-1'), 6.83–6.88 (m, 8H, aromatic), 7.27–7.43 (m, 18H, aromatic), 8.30 and 8.39 (2× s, 2H, 2× H-6). HRMS (ESI-TOF): molecular formula, C₄₈H₆₄N₅O₈PSeSi·[M + H]⁺: 978.3479 (calc. 978.3505).

**Synthesis of the SeU-RNAs**

All the RNA oligonucleotides were chemically synthesized in 1.0 µmol scale on solid phase. The ultra-mild RNA phosphorimidates protected with 2′-TBDMS were used (Glen Research). The concentration of the Se-U-phosphorimidate was 0.08 M in acetonitrile, compared with the regular ones (0.1 M). Coupling was carried out using 5’-(benzylmercapto)-1H-tetrazole solution (0.25 M) in acetonitrile with 12 min coupling time for both native and Se-modified phosphorimidates. Three percent trichloroacetic acid in methylene chloride was used for the 5′-detritylation. Synthesis was performed on control-pore glass (CPG-500) immobilized with the appropriate nucleoside through a succinate linker. All oligonucleotides were eluded in a 20-min linear gradient of 100% buffer A to 60% of buffer B. The HPLC analysis for both DMTr-on and DMTr-off oligonucleotides were carried out with up to 60% of buffer B in a linear gradient in the same period of time. The collected fractions were lyophilized, and the purified RNAs were re-dissolved in water for the detritylation and precipitation steps.

**Thermodenaturation of the SeU-RNAs**

Solutions of the duplex RNAs (1 or 2 µM) were prepared by dissolving the purified RNAs in sodium phosphate [10 mM (pH 6.5)] buffer containing 100 mM NaCl. The solutions were heated to 75°C for 3 min, then cooled down slowly to room temperature and stored at 4°C overnight before Tm measurement. Before thermal denaturation, the Se-RNA samples were bubbled with argon for 5 min. Each denaturizing curves were acquired at 260 nm by heating and cooling from 5 to 70°C for four times in a rate of 0.5°C/min, using Cary-300 UV-Visible spectrometer equipped with temperature controller system.

**Se-RNA crystallization and diffraction data collection**

The purified RNA oligonucleotides (1 mM) were heated to 70°C for 2 min and cooled down slowly to room temperature. Both native buffer and Nucleic Acid Mini Screen Kit (Hampton Research) were applied to screen the crystalization conditions at different temperatures using the hanging drop method by vapor diffusion (1 µl of RNA and 1 µl of buffer). Thirty percent glycerol, PEG 400 or the perfluoropolyether was used as a cryoprotectant during the crystal mounting, and data collection was taken under the liquid nitrogen stream at 99°C. The Se-RNA crystal data were collected at beam line X12B and X12C in NSLS, Brookhaven National Laboratory. A number of crystals were screened to find the ones with strong anomalous scattering at the K-edge absorption of selenium. The distance of the detector to the crystals was set to 150 mm. The radiation wavelength at 0.9795 Å was chosen for diffraction data collection and selenium single-wavelength anomalous dispersion (SAD) phasing. The crystals were exposed for 10 s per image with 1° oscillation, and a total of 180 images were taken for each data set. All data were processed using HKL2000 and DENZO/SCALEPACK (38).

**Structure determination and refinement**

The structures of Se-RNAs were solved by both SAD with HKL2MAP and molecular replacement with Phaser (39), followed by the refinement with Refmac. Both SAD phasing and molecular replacement led to the same crystal structure. The refinement protocol includes simulated annealing, positional refinement, restrained B-factor refinement and bulk solvent correction. The stereochemical topology and geometrical restrain parameters of DNA/RNA (40) have been applied. The topologies and parameters for the uridine modified with selenium (US) were constructed and applied. After several cycles of refinement, a number of highly ordered waters were added. Finally, the occupancies of selenium were adjusted. Cross-validation (41) with a 5–10% test set was monitored during the refinement. The σA-weighted maps (42) of the (2m|Fo| - D|Fc|) and the difference (m|Fo| - D|Fc|)
density maps were computed and used throughout the model building.

RESULTS AND DISCUSSION

Synthesis of the 4-Se-uridine (SeU) phosphoramidite

We have developed a facile strategy to synthesize the Se-phosphoramidite. As showed in Scheme 1, our synthesis started from the partially protected 2'-TBDMS-5'-trityl-uridine (1). To simplify the synthesis, we used a bulky reagent (2,4,6-trisopropylbenzenesulfonyl chloride, TIBS-Cl) to selectively activate position 4, thus avoiding the protection and deprotection steps of the 3'-hydroxyl group. Without purifying the activated intermediate, the selenium functionality was introduced by substituting TIBS group at position 4 with 2-cyanoethylselenide in the yield of 81%. Sodium 2-cyanoethylselenide was generated by the reduction of di-(2-cyanoethyl) diselenide with NaBH4 in ethanol solution (30). This protected Se-functionality is compatible with the solid-phase synthesis and can be removed by weak base treatment (K2CO3 in methanol). Finally, the 4-Se-uridine derivative (2) was converted to the corresponding phosphoramidite (3) in 92% yield. The analysis data are shown in the supporting information (Supplementary Figures S1–S7).

Synthesis of the SeU-RNAs

The ultramild phosphoramidites, where the base-labile protecting groups can be deprotected with a weak base (K2CO3 in methanol) (30-32,33,35,43), were used because the 4-Se-functionality is sensitive to strong base cleavage (such as ammonia, causing deselenization). We found that this Se-modified phosphoramidite is compatible with the longer coupling time (12 min), I2 oxidation and trichloroacetic acid treatment without deselenization.

To determine the extinction coefficient of 4-Se-uridine residue (SeU) by comparing with the native nucleotide, we synthesized and purified the 5'-SeU-3'. Their HPLC profiles are presented in Figure 3. The HPLC assistance, which removes and minimizes the interference of impurities, allows accurate measurement of the extinction coefficients (43). Our experimental results indicate that 5'-SeU-3' residue absorbs at both 260 and 370 nm (Figure 3A). The absorption ratio at these two wavelengths is 5.71, calculated on the basis of the HPLC peak areas. As the extinction coefficient is proportional to the absorption, Equation (1) is deduced. In addition, from the HPLC profile (Figure 3B) of 5'-SeU-3', the ratio between the absorption at 260 nm (contributed by both native U and SeU) and 370 nm (only by SeU) is determined as 0.920. Thus, Equation (2) is deduced. As the extinction coefficient of native U at 260 nm (ε260u = 9.66 × 10^3 M^-1 cm^-1) is known (45), we calculated the extinction coefficient of 4-Se-uridine residue (ε260seu) by subtracting the extinction coefficient of uridine (ε260u) from the extinction coefficient of 5'-SeU-3' (ε260seu = ε260seu - ε260u).
coefficient of $\text{SeU}$ at 370 nm ($e_{370}^{\text{SeU}}$) and 260 nm ($e_{260}^{\text{SeU}}$) are 13.0 and 2.28 $/C_210^3$ $/C_01cm$ $/C_01$, respectively.

$$e_{370}^{\text{SeU}} = e_{260}^{\text{SeU}} \approx 5.71$$

$$\left( \frac{e_{260}^{\text{U}} + e_{260}^{\text{SeU}}}{e_{370}^{\text{SeU}}} \right) = 0.920$$

Thermodenaturation study

The rationales of using a Se atom to probe the U-A and U$\cdot$U base pairs are that selenium, a large-size atom, can probably strengthen the stacking interaction and is a poorer hydrogen-bond acceptor (30,32,33) that can likely weaken the hydrogen-bond (H-bond) interaction. The polarizable and large Se atom with delocalizable electrons can increase the stacking interaction by narrowing the gap between the stacked nucleobases, which is observed in our crystal structure presented in this work. Furthermore, the increase of the stacking interaction by this Se atomic probe is consistent with the computational study of the Se-modified thymidine in DNA duplex (46). Thus, the Se-atom probe that alters the stacking and H-bonding interactions may provide novel insights into the base pairs. To investigate the RNA duplex recognition and stability, we carried out the UV-melting study with RNAs containing the 4-Se-uracil in duplexes or in duplex junctions (or overhang regions). Typical curves of
Se-RNA melting-temperatures (Tm) are showed in Figure 4, and all the Tm data are summarized in Table 2, compared with the corresponding native RNA duplexes. When the Se-atom probe is introduced to the uracil in RNA duplexes, no significant Tm differences between the native and Se-modified duplexes were observed (entry 1–8 in Table 2), and the free energy \((\Delta G)\) differences with the corresponding natives were almost zero. This suggests that the Se-atom probe in RNA duplex regions may not cause significant perturbation in duplex stability. As selenium is a poor H-bond acceptor, it is anticipated that the Se-mediated H-bond in the U-A pair is weak. The zero (or very small) free energy difference between the native and Se-modified RNA duplexes also indicates that the stability increase via the stronger stacking compensates the stability decrease caused by the weaker H-bonding. This observation reveals that the modified U-A base-pair can maintain a fine balance between the stacking and H-bonding interactions.

It is reported that a U●U pair is less stable comparing with a U-G or C-A mispair in a RNA duplex (33,47). In RNA duplex junctions and loops, however, the two consecutive U●U pairs are more stable than the two consecutive A-A pairs (48). Thus, the Se-atom probe is used to investigate the non-canonical U●U pair, and we chose and modified the RNAs forming RNA duplex and UU junction (Table 2). The UV-thermal denaturation study was carried out, and the melting-temperatures (Tm) of the Se-RNAs and their corresponding natives are summarized in Table 2 (entry 9–14). Excitingly, when the atomic probe is introduced to the RNA duplex junctions, the melting temperatures increased by 1.5–2.4 \(^\circ\)C per Se-modification of these RNA duplexes. Consistently, the free energy \((\Delta G)\) calculation indicates that each Se atom contributed additional stabilization (0.4–0.8 kcal/mol) to the stability of the RNA duplexes. This increased RNA duplex stability is attributed to the increased stacking interaction of SeU on the duplex ends; the support from the high-resolution structure data is presented later. Via the Se-atom probe, the UV-melting study of the duplex RNAs containing the UU junction indicates that the
uracil stacking contributes significantly to RNA duplex stability.

**Crystallization, diffraction data collection and crystal structure determination**

To investigate the Se-nucleobase modification and its structural property, we have crystallized two Se-RNA sequences [hexamer (5'-rUUUAUUAUUAUAA-3')$_2$ with overhangs and octamer (5'-rGUGG-UACAC-3')$_2$ with a perfect duplex]. Crystals of both Se-RNA sequences were formed in 2–5 days at room temperature (25°C) with the Hampton nucleic acid mini-screen kit (total 24 buffers with broad conditions). Excitingly, all crystals of both Se-RNAs had strong yellow or dark yellow color because of the selenium modification (Figures 5 and 6). The Se-RNA hexamer formed crystals in 22 of 24 buffers using the kit, whereas the corresponding native RNA formed crystals only in 4 of 24 buffers (in 3 weeks) using the kit. Most of these Se-RNA crystals (one example shown in Figure 5) diffracted very well, up to 1.3 Å resolution (the orthorhombic space group, C222$_1$). Similarly, the Se-RNA octamer formed crystals in 22 of 24 buffers using the same kit, and these crystals (examples shown in Figure 6) could diffract up to 2.5 Å resolution (the rhombohedral space group, R32). In contrast, the corresponding native (5'-rGUGG-UACAC-3')$_2$ did not crystallize under any conditions over several weeks, which is consistent with the literature (49). The native octamer (5'-rGUGGUACAC-3')$_2$ is difficult to crystallize, and its structure has not been reported in literature. Finally, several high-quality crystals from these two Se-RNAs were mounted and cryo-protected for the diffraction data collection. The structures were determined using the best data sets and diffractions collected from the crystals grown in buffer No.10 [10% MPD, 40 mM Na Cacodylate (pH 6.0), 12 mM Spermine tetra-HCl, 12 mM NaCl and 80 mM KCl] for the Se-hexamer and No.12 [10% MPD, 40 mM Na Cacodylate (pH 6.0), 12 mM Spermine tetra-HCl, 80 mM KCl and 20 mM BaCl$_2$] for the Se-octamer. The statistic data of the structural analysis are summarized in Table 3, and the determined Se-RNA structures are presented in Figures 5 and 6.

**Structures of 4-Se-derivatized RNAs**

The structure of the Se-RNA hexamer (Figure 5) revealed formation of the right-handed Watson–Crick duplex (Supplementary Table S1) and Hoogsteen base pairs. The structures determined via SAD and molecular replacement approaches are identical. The Se-modified structure (PDB ID: 3HGA; 1.30 Å resolution) and the corresponding native structure (PDB ID: 1OSU; 1.40 Å resolution) (50) are virtually identical as well. They can superimpose on each other perfectly well (Figure 5C) with the RMSD as 0.09 Å, indicating the fine structure superimpose on each other perfectly well (Figure 5C) with the RMSD as 0.09 Å, indicating the fine structure reproducibility approaches are identical. The Se-modified structure (PDB ID: 3HGA; 1.30 Å) and the corresponding native structure (PDB ID: 1OSU; 1.40 Å) (50) are virtually identical as well. They can superimpose on each other perfectly well (Figure 5C) with the RMSD as 0.09 Å, indicating the fine structure isomorphism. Moreover, the electron delocalization of the large Se atom on the uracil may facilitate the nucleobase stacking interaction, also supported by the computational study of the Se-modified nucleobase (46). Furthermore, Se atom is 0.43 Å larger than O, and the distances between U2 4-exo-Se and the 3'-cytosine atoms (N3, exo-N4, C4 and C5) are similar to the corresponding native distances between U2 4-exo-O and the 3'-cytosine atoms (Figure 5D and E); the distances between the 4-Se or 4-O atom and the 3'-C atoms are also displayed. Thus, the comparison of the Se-modified and native structures (Figure 5D-I) suggests that the Se-nucleobase may better stack on the 3'-cytosine than the native nucleobase. The stronger stacking interaction can rigidify the local conformation and strengthen the RNA duplexes, which are consistent with the stronger duplex stability in the presence of the UU overhang (or duplex junction; Table 2). These results are also consistent with the faster crystal growth after the selenium modification. Similar to the corresponding native structure (50), two SeU-U pairs (Hoogsteen pair) have been observed in the Se-RNA (Figure 5F and G). In the Se-modified and native structures, both SeU-U and UsU pairs participate in formation of a pseudo-fiber and long duplex through the overhang

### Table 2. UV-melting temperatures of SeU-RNAs

| Entry | Modified region | RNA sequences | Tm (°C) | ΔTm (°C) |
|-------|-----------------|---------------|---------|----------|
| 1     | Duplex          | (5'-rUUUAUUAUUAUAA-3')$_2$ | 38.0 ± 0.3 |          |
| 2     | Duplex          | (5'-rUUUA-3'SeU-AUUUAUAA-3')$_2$ | 37.3 ± 0.5 | −0.7     |
| 3     | Duplex          | (5'-rGGG-UUUCGGGUACC-3')$_2$ | 45.0 ± 0.4 |          |
| 4     | Duplex          | (5'-rGGG-SeU-AUUGCGGUACC-3')$_2$ | 44.2 ± 0.3 | −0.8     |
| 5     | Duplex          | (5'-rCGCGGAU-3'SeU-CGG-3')$_2$ | 39.4 ± 0.4 |          |
| 6     | Duplex          | (5'-rCGCGGAU-3'SeU-CGG-3')$_2$ | 39.0 ± 0.3 | −0.4     |
| 7     | Duplex          | 5'-AUCACCUCCCUA-3' | 43.2 ± 0.3 |          |
| 8     | Duplex          | 3'-UAGUUGAGGAAAU-5' | 42.8 ± 0.4 | −0.4     |
| 9     | Duplex junction | 5'-U-3'SeU-UCCCUUCAU-3' | 42.2 ± 0.2 |          |
| 10    | Duplex junction | 5'-U-3'SeU-AGCUAGCU-3' | 47.1 ± 0.3 | +4.9     |
| 11    | Duplex junction | 5'-UUCGCGUACGG-3' | 59.4 ± 0.3 |          |
| 12    | Duplex junction | 5'-U-3'SeU-CGCGUACGG-3' | 63.2 ± 0.3 | +3.8     |
| 13    | Duplex junction | 5'-UUCAGUGACC-3' | 48.2 ± 0.3 |          |
| 14    | Duplex junction | 5'-U-3'SeU-CAUGUGACC-3' | 49.9 ± 0.4 | +1.7     |
Hoogsteen-base pairs. The 5′-UU sequence allows the RNAs (both the Se-modified and native ones) infinitely stacking and elongating along the 2₁ screw axis in the crystals with nicks on the 5′-end of each 5′-U(SeU). This 5′-U-SeU sequence forms the two symmetrical SeU/C₁₅U base pairs, which is virtually identical to the native U/C₁₅U pair (Figure 5G). Namely, this junction sequence forms the two symmetrical SeU•U base pairs, which glue the RNA duplexes together in a head-to-tail linear fashion.

The results of our crystal structure study are consistent with the UV-melting study. The 5′-UU of one RNA molecule (e.g. the red one in Figure 6A) forms two U•U

![Figure 5](image-url)
Figure 6. Schematic diagram and local structures of the native and modified U•U pairs in overhang regions (or duplex junctions). (A) Schematic diagram of the RNA duplex with five strands, the nicks, SeU•U pairs and normal Watson–Crick C-G pairs. (B) Superimposition comparison of SeU14•U1 (in red) with native U14•U1 pair (in cyan); the numbers represent the H-bond lengths (Å). (C) The stacking of two SeU•U pairs with the distance (3.11 Å) between the two neighbor Se atoms in the modified U14 and U2. (D) The stacking of two native U•U pairs with the distance (3.29 Å) between the two neighbor O atoms in native U14 and U2. The 2Fo-Fc maps of Se-4 and O-4 are showed.

Table 3. Diffraction data collection and refinement statistics of the Se-RNA structures

| Structure (PDB ID) | U-SeU-CGCC (3HGA) | GUG-SeU-ACAC (4IQS) |
|--------------------|-------------------|---------------------|
| Data collection    |                   |                     |
| Space group        | C222              | R3_2                |
| Cell dimensions: a, b, c (Å) | 30.255, 34.079, 28.931 | 47.006, 47.006, 354.105 |
| Resolution range, Å (last shell) | 50.00-1.30 (1.32-1.30) | 50.0-2.60 (2.69-2.60) |
| Unique reflections | 3773 (162)        | 8915 (846) |
| Completeness%      | 95.9 (90.0)       | 99.2 (95.8) |
| Rmerge%            | 4.5 (26.1)        | 5.3 (35.8) |
| I/σ(I)             | 40.5 (1.2)        | 35.9 (1.0) |
| Redundancy         | 11.7 (4.2)        | 10.0 (6.1) |
| Refinement         |                   |                     |
| Resolution range, Å | 22.62-1.30 | 31.73-2.60 |
| Rwork%             | 18.9              | 19.4 |
| Rfree%             | 22.5              | 25.8 |
| Number of reflections | 3586     | 4776 |
| Number of atoms    |                   |                     |
| Nucleic acid (single) | 157             | 1002 |
| Heavy atoms and ion| 1 Se              | 6 Se |
| Water              | 42                | 0 |
| R.m.s. deviations  |                   |                     |
| Bond length, Å     | 0.005             | 0.008 |
| Bond angle, °      | 0.931             | 1.846 |

Rmerge = Σ|I – ⟨I⟩|/ΣI
pairs with the second RNA molecule (the blue one), whereas its consecutive CGCG sequence forms regular Watson–Crick base pairs with the third RNA molecule (the green one). As showed in Figure 5F, the $\text{SeU} \cdot \text{U}$ pair displays a conventional hydrogen bond between O4 of the native uracil (U1) and N3 of the Se-uracil (U14) and an unusual C-H-Se hydrogen bond between C5 of native U and Se4 of Se-U, through the Hoogsteen edge of native U and the Watson–Crick edge of Se-U. These interactions result in a trans-Hoogsteen U\text{U} \cdot \text{U} pair (Figure 5F). Compared with the native structure, the substitution of the uridine 4-oxygen with a selenium atom does not change the structure significantly (Figure 5C), suggesting that the Hoogsteen U\text{U} \cdot \text{U} pair has space available at 4-position of the Watson–Crick edge. A slight shift (0.09 Å) on the Se-modified nucleobase is observed (Figure 6B). The Hoogsteen C-H-Se (or O) hydrogen bond (bond length: 3.36 Å in the Se case), between C5 of native U and Se4 of Se-U (the corresponding native H-bond: 3.27 Å; Figure 6B), is still retained. Because selenium atom (1.16 Å in atomic radius) is 0.43 Å larger than oxygen (0.73 Å in atomic radius), it is surprising to find the nucleobase shift only by 0.09 Å to accommodate the big selenium atom, confirming that the native hydrogen bond (O4-H-C5) of the Hoogsteen pair is weak. Thus, the large Se atom probe indicates that the Hoogsteen H-bond is less important in the U\text{U} pairing. This also suggests that the trans-Hoogsteen pair can tolerate a larger substitution and that the Hoogsteen pair is not rigid, which gives the duplex junction sufficient flexibility. Moreover, it is counterintuitive that the distance (3.11 Å) between these two big neighboring 4-Se atoms (Figure 6C) is even smaller (by 0.18 Å) than the native distance (3.29 Å) between two small O atoms (Figure 6D), implying the enhanced stacking interactions between these two U\text{U} pairs. Using electron-rich selenium as the atomic probe, our structural result suggests the strong electron delocalization and stacking interaction between these two U\text{U} pairs. The structure study provides new insights into the Hoogsteen U\text{U} \cdot \text{U} pair and the uracil-mediated interactions in ncRNAs.

The Se-octamer structure (Figure 7), where the two Se atoms point to the major groove, reveals formation of the $\text{SeU} \cdot \text{A}$ pair and the typical right-handed A-form duplex by the Se-RNA (Supplementary Table S2). Moreover, we have superimposed the structures of $\text{SeU} \cdot \text{A}$ (or $\text{SeU}4 \cdot \text{A}13$ pair) and U2-A15 pair (Figure 7D), as the corresponding native structure is not available (from literature or us) for direct comparison. This comparison of the base pair structures has demonstrated that the Se-modified and native U-A pairs are similar. The major difference is the slight shift of the $\text{SeU}$ nucleobase to accommodate the large selenium atom, revealing the flexibility of RNA duplex structure. The distance between $\text{SeU}4$ exo-Se4 and A13 exo-N6 is 3.54 Å, which was increased from the original 2.99 Å. Considering that the atomic size of Se is 0.43 Å larger than that of O and that a typical H-bond length is 2.8–3.2 Å, this distance (3.54 Å) suggests a weak hydrogen bond after the Se-modification. On the other hand, the polarizable and large Se atom with delocalizable electrons may facilitate the base stacking interaction, supported by the narrower base-pair gap and the computational study of the Se-nucleobase-modified DNA (46). Using the Se atom probe, we found that the increased

![Figure 7](http://nar.oxfordjournals.org/)

**Figure 7.** The yellow crystals and structures of the 4-Se-U RNA octamer, \((5’-\text{GUG-}^{3\text{Se}}\text{U-ACAC-3})_2\). The Se atoms are labeled as purple balls. (A) Crystal image. (B) The Se-RNA duplex structure (PDB ID: 4IQS; 2.75 Å resolution). (C) Electron density map (2Fo-Fc) and model of the $\text{SeU} \cdot \text{A}$ pair at the level of 1.0 $\sigma$. (D) Superimposition of $\text{SeU} \cdot \text{A}$ pair (in pink) with native U2-A15 pair (in cyan); the H-bond lengths are indicated individually.
stacking interaction can compensate the loss of the H-bond interaction, which is consistent with the virtually identical duplex stability after the Se-modification (Table 2). Moreover, most of the 2'-hydroxyl groups are involved in the H-bonding interactions with its 3'-sugar ring oxygen (O4') or 3'-phosphate oxygen, which restrains the conformations of the sugar-phosphate backbone, thereby facilitating the intramolecular interaction and reducing molecular dynamics. The Se-RNA crystallization is consistent with the Se-enhanced base stacking and conformation rigidification. In the crystal lattice, the duplexes are stacked on the top of each other in a head-to-tail fashion and three Se-RNA duplexes present in an asymmetric unit, where the three duplexes are virtually identical (r.m.s. < 0.1 Å). Chain A and B are shown in Figure 7.

Furthermore, X-ray crystallography is one of the most powerful methodologies for structure and function studies of RNAs and their complexes with ligands, including protein-RNA complexes and RNA-small molecule complexes, at the atomic resolution. However, owing to the difficulties in crystallization and phasing (phase determination or phase problem), progress in RNA crystallography is limited, especially in the ncRNA structure study. Inspired by the protein Se-derivatization, multi-wavelength anomalous dispersion phasing and SAD phasing (51–55), our laboratory has pioneered SeRNA (36,37), which has great potential as a general strategy for RNA X-ray crystallography (37). This research work on the synthesis and structure studies of the 4-Se-uridine RNAs has further demonstrated that the selenium modification is a useful approach for structural biology, as the Se-functionalization can facilitate phase determination, crystallization, RNA color and atomic probing.

CONCLUSION

To probe uracil-mediated interactions and base-pairs with a single selenium atom, we have synthesized the 4-Se-uridine phosphoramidite and Se-RNAs. Our thermostability and structure studies indicate that the modified and native structures are virtually identical, that the H-bonding decrease in U-A pair can be compensated by the base-stacking increase, and that the uracil stacking in duplex junction may increase duplex thermostability. We also found that the stacking interaction of the two trans-Hoogsteen U•U pairs is the main contributor to the duplex junction stability, whereas the Hoogsteen H-bond is weak. Moreover, the accommodation of larger Se atoms in uracil by both U-A and U•U pairs implies the RNA flexibility. Using the Se atom probe, our studies confirm that uracil is capable of interacting in multiple modes, thereby diversifying U•U and U-A pairs in structure and function. Our thermodynamic and structural studies have also demonstrated that this Se-modification can facilitate the nucleobase stacking interaction and potential crystal growth without significant perturbation. Furthermore, this Se-modification generates color RNA for the first time by single atom replacement, and it shifts the uridine UV spectrum over 100 nm (λmax: 370 nm; ε: 1.30 x 10⁴ M⁻¹ cm⁻¹). This color property is useful for RNA-protein co-crystallization, RNA visualization, detection and spectroscopic study. This work provides a new strategy for crystallization, phasing, structure and function studies of ncRNAs and protein-RNA complexes.

ACCESSION NUMBERS

3HGA, 4IQS

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank the mail-in program and PXrr in Brookhaven National Laboratory (BNL). The National Synchrotron Light Source (NSLS) at BNL is funded by NIH’s National Center for Research Resources and DOE’s Office of Biological and Environmental Research.

FUNDING

NIH [R01GM095881] and Georgia Cancer Coalition (GCC) Distinguished Cancer Clinicians and Scientists. Funding for open access charge: NIH [R01GM095881] and Georgia Cancer Coalition (GCC) Distinguished Cancer Clinicians and Scientists.

Conflict of interest statement. None declared.

REFERENCES

1. Watson, J. D. and Crick, F. H. (1953) Genetical implications of the structure of deoxyribonucleic acid. Nature, 171, 964–967.
2. Imai, M., Richardson, M. A., Ikegami, N., Shatkin, A. J. and Furuichi, Y. (1983) Molecular cloning of double-stranded RNA virus genomes. Proc. Natl Acad. Sci. USA, 80, 373–377.
3. Bass, B. L. and Cech, T. R. (1984) Specific interaction between the self-splicing RNA of Tetrahymena and its guanosine substrate: implications for biological catalysis by RNA. Nature, 308, 820–826.
4. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. and Doudna, J. A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. Science, 273, 1678–1685.
5. Lohse, P. A. and Szostak, J. W. (1996) Ribozyme-catalysed amino-acid transfer reactions. Nature, 381, 442–444.
6. Cheeth, M. T., Wachter, A., Sudarsan, N. and Breaker, R. R. (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature, 447, 497–500.
7. Bradshaw, N., Neher, S. B., Booth, D. S. and Walter, P. (2009) Signal sequences activate the catalytic switch of SRP RNA. Science, 323, 127–130.
8. Lincoln, T. A. and Joyce, G. F. (2009) Self-sustained replication of an RNA enzyme. Science, 323, 1229–1232.
9. Tsai, M.-C., Manor, O., Wan, Y., Mosammeparast, N., Wang, J. K., Lun, F., Shi, Y., Segal, E. and Chang, H.-Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science, 329, 689–693.
10. Weil, L., James, L., Ulyrc, N., Chamond, N., Herbretneau, C. H., Ohlmann, T. and Sargueil, B. (2010) A new type of IRES within gag coding region recruits three initiation complexes on HIV-2 genomic RNA. Nucleic Acids Res., 38, 1367–1381.
11. Lee, J. T. (2012) Epigenetic regulation by long noncoding RNAs. Science, 338, 1435–1439.
12. Rinn,J.L. and Chang,H.Y. (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.*, 81, 145–166.
13. Rinn,J.L., Kertesz,M., Wang,J.K., Squazzo,S.L., Xu,X., Brugmann,S.A., Goodnough,L.H., Helms,J.A., Farnham,P.J., Segal,E. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.*, 129, 1311–1323.
14. Coley,W., Kulmi-Halik,. Van Duyne,R. and Kashanchi,F. (2009) Novel HIV-1 therapeutics through targeting altered host cell pathways. *Expert Opin. Biol. Ther.*, 9, 1369–1382.
15. Klein,D.J. and Ferre-D’Amare,A.R. (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*, 313, 1752–1756.
16. Ke,A., Zhou,K., Ding,F., Cate,J.H. and Doudna,J.A. (2004) Structural basis of 3 nucleic acid hydroxyl groups by warming-up. *Nucleic Acids Research*, 32, 719–724.
17. Rottiers,V. and Naar,A.M. (2012) MicroRNAs in metabolism and metabolic disorders. *Nat. Rev. Mol. Cell Biol.*, 13, 239–250.
18. Qureshi,I.A. and Mehler,M.F. (2012) Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat. Rev. Neurosci.*, 13, 40–51.
19. Esteller,M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, 12, 861–874.
20. Gilbert,W. (1986) The RNA World. *Nature*, 319, 618.
21. Gesteland,R.F., Cech,T.R. and Atkins,J.F. (2006) The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA catalysis. *Nature*, 429, 201–205.
22. Czerwoniec,A., Dunin-Horkawicz,S., Purta,E., Kaminska,K.H., Kasprzak,J.M., Bujnicki,J.M., Grosjean,H. and Rother,K. (2009) MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res.*, 37, D118–D121.
23. Carr,bew,R.W. and Sontheimer,E.J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 136, 642–655.
24. Phizicky,E.M. and Alfonzo,J.D. (2010) Do all modifications benefit all tRNAs? *FEBS Lett.*, 584, 265–271.
25. Michel,F., Jacquier,A. and Dujon,B. (1982) Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie*, 64, 867–881.
26. Davies,R.W., Waring,R.B., Ray,J.A., Brown,T.A. and Scacciazzochio,C. (1982) Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature*, 300, 719–724.
27. Michel,F., Netter,P., Xu,M.Q. and Shub,D.A. (1990) Mechanism of 3′ splice site selection by the catalytic core of the sunY intron of bacteriophage T4: the role of a novel base-pairing interaction. *Genes Dev.*, 4, 777–788.
28. Wise,D.S. and Townsend,L.B. (1972) Synthesis of the tetragonal hen egg-white lysozyme. *Acta. Crystallogr. D Biol. Crystallogr.*, 28, 1398–1405.
29. Read,R.J. (1986) Improved Fourier coefficients for maps using single 4-selenothyrimidine nucleosides 2-seleno- and 4-selenouridine J. *Org. Chem.*, 51, 265–271.
30. Rossi,J., Jiang,J., Sheng,J., Gerlits,O.O. and Huang,Z. (2008) Derivatization of DNAs with selenium at 6-position of guanine for function and crystal structure studies. *Nucleic Acids Res.*, 36, 7009–7018.
31. Salon,J., Zhang,Y. and Huang,Z. (2011) Mild detritylation of nucleic acid hydroxyl groups by warming-up. *Nucleosides Nucleotides Nucleic Acids*, 30, 271–279.
32. Cava1uzzi,M.J. and Borer,P.N. (2004) Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.*, 32, e13.
33. Christoffersen,A., Zhao,L., Sun,H., Huang,Z. and Huang,N. (2011) Theoretical studies of the base pair fidelity of selenium-modified DNAJ. *Phys. Chem. B*, 115, 10041–10048.
34. Kierzek,R., Burdk,M.E. and Turner,D.H. (1999) Thermodynamics of single mismatches in RNA duplexes. *Biochemistry*, 38, 14214–14223.
35. SantaLucia,J. Jr, Kierzek,R. and Turner,D.H. (1991) Stabilities of consecutive A.C, C.C, G.G, U.C, and U.U mismatches in RNA internal loops: evidence for stable hydrogen-bonded U.U and C.C.+ pairs. *Biochemistry*, 30, 8242–8251.
36. Wahl,M.C., Ramakrishnan,B., Ban,C., Chen,X. and Sundaralingam,M. (1996) RNA - synthesis, purification and crystallization. *Acta. Crystallogr. D Biol. Crystallogr.*, 52, 668–675.
37. Wahl,M.C., Rao,S.T. and Sundaralingam,M. (1990) The structure of r(UUCGCG) has a 5′-U-overhang exhibiting Hoogsteen-like trans U.U base pairs. *Nat. Struct. Biol.*, 3, 24–31.
38. Hendrickson,W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, 254, 51–58.
39. Hendrickson,W.A. (2000) Synchrotron crystallography. *Trends. Biochem. Sci.*, 25, 637–643.
40. Yang,W., Hendrickson,W.A., Crouch,R.J. and Satow,Y. (1990) Structure of ribonucleobase H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science*, 249, 1398–1405.