



# HHS Public Access

Author manuscript

*Crystals (Basel)*. Author manuscript; available in PMC 2021 March 26.

Published in final edited form as:

*Crystals (Basel)*. 2021 March ; 11(3): . doi:10.3390/cryst11030273.

## Affinity and Structural Analysis of the U1A RNA Recognition Motif with Engineered Methionines to Improve Experimental Phasing

Yoshita Srivastava<sup>1,‡</sup>, Rachel Bonn-Breach<sup>1,‡</sup>, Sai Shashank Chavali<sup>1,‡</sup>, Geoffrey M. Lippa<sup>1,3</sup>, Jermaine L. Jenkins<sup>1,2</sup>, Joseph E. Wedekind<sup>1,\*</sup>

<sup>1</sup>Department of Biochemistry & Biophysics and Center for RNA Biology, University of Rochester School of Medicine & Dentistry, Rochester, NY 14642, USA

<sup>2</sup>Structural Biology and Biophysics Facility, University of Rochester School of Medicine & Dentistry, Rochester, NY 14642, USA

<sup>3</sup>Present address: Division of Biology, Alfred University, Alfred, NY 14802, USA

### Abstract

RNA plays a central role in all organisms and can fold into complex structures to orchestrate function. Visualization of such structures often requires crystallization, which can be a bottleneck in the structure-determination process. To promote crystallization, an RNA-recognition motif (RRM) of the U1A spliceosomal protein has been co-opted as a crystallization module. Specifically, the U1-snRNA hairpin II (hpII) single-stranded loop recognized by U1A can be transplanted into an RNA target to promote crystal contacts and to attain phase information via molecular replacement or anomalous diffraction methods using selenomethionine. Herein, we produced the F37M/F77M mutant of U1A to augment the phasing capability of this powerful crystallization module. Selenomethionine-substituted U1A(F37M/F77M) retains high affinity for hpII ( $K_D$  of  $59.7 \pm 11.4$  nM). The 2.20 Å resolution crystal structure reveals that the mutated sidechains make new S- $\pi$  interactions in the hydrophobic core and are useful for single-wavelength anomalous diffraction. Crystals were also attained of U1A(F37M/F77M) in complex

Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

\*Correspondence: joseph.wedekind@rochester.edu; Tel.: +1 585 273-4516.

**Author Contributions:** J.E.W. and G.M.L. conceived and designed the dmU1A mutations. J.E.W. conceived of the TBP6.9 as a crystallization module. Y.S., R.B.B., and S.S.C. purified the proteins and RNA and conducted ITC. G.M.L. purified dmU1A(F37M/F77M) protein and associated RNA, and grew all crystals. J.L.J. collected the X-ray diffraction data, determined the structure and refined the structure. J.E.W. conducted the data analysis, generated the figures and wrote the manuscript with input from all authors. All authors have read and agreed to the published version of the manuscript.

<sup>‡</sup>contributed equally to this work.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2073-4352/11/3/273/s1](http://www.mdpi.com/2073-4352/11/3/273/s1), Supplementary Methods, Figure S1: Size-exclusion chromatography profiles for dmU1A and dmU1A(F37M/F77M), Figure S2: ITC analysis of dmU1A and dmU1A(F37M/F77M) to assess heats of dilution, Table S1: Intensity statistics for the Se-Met dmU1A(F37M/F77M)-preQ1-II riboswitch complex.

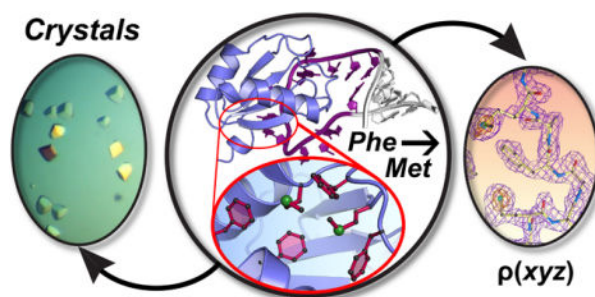
**Informed Consent Statement:** Not applicable

**Data Availability Statement:** Atomic coordinates and structure factor amplitudes for the U1A(F37M/F77M) mutant are available from the Protein Data Bank as entry 7LHX. The pET28-TEV plasmids encoding the U1A(F37M/F77M) and TBP6.9(F34M/F37M/F77M) proteins have been deposited into Addgene ([www.addgene.org](http://www.addgene.org)) and are available to the community.

**Conflicts of Interest:** The authors declare no conflicts of interest.

with a bacterial preQ<sub>1</sub>-II riboswitch. The F34M/F37M/F77M mutant was introduced similarly into a lab-evolved U1A variant (TBP6.9) that recognizes the internal bulged loop of HIV-1 TAR RNA. We envision that this short RNA sequence can be placed into non-essential duplex regions to promote crystallization and phasing of target RNAs. We show that selenomethionine-substituted TBP6.9(F34M/F37M/F77M) binds a TAR variant wherein the apical loop was replaced with a GNRA tetraloop ( $K_D$  of  $69.8 \pm 2.9$  nM), laying the groundwork for use of TBP6.9(F34M/F37M/F77M) as a crystallization module. These new tools are available to the research community.

## Graphical Abstract



## Keywords

RNA-protein interactions; RNA crystallization; isothermal titration calorimetry; X-ray crystallography; selenomethionine; anomalous diffraction; U1A RRM; S- $\pi$  interaction

## 1. Introduction

RNA is integral to cell function and has overturned old rules that erroneously credited proteins as master regulators of key biological functions [1]. Unlike the human proteome, which arises from only a small fraction of the DNA blueprint (0.05%), most of the genome is transcribed into RNA [2,3]. Accordingly, many RNA transcripts are crucial molecules that function to guide DNA synthesis, as guardians against invading nucleic acids, and as controllers of gene regulation at the transcriptional and translational levels [1]. The importance of RNA is heightened by the fact that only a small fraction of the proteome (~3.5–10%) is likely druggable [4,5]. Consequently, RNA has been viewed increasingly as a therapeutic target [6–8].

On the road to understanding RNA molecular function, near atomic-level structures of several important RNA classes have been elucidated using experimental approaches [9–22]. This work has demonstrated that RNA can adopt elegant three-dimensional folds with distinct topologies and recurrent architectural motifs [23–33]. Many of these RNAs include cavities and deep grooves poised to receive natural or artificial ligands [34]. At present, the preponderance of known RNA structures has been derived from x-ray crystallography. Albeit, successful RNA crystallization can be a rocky pathway that often represents the rate-limiting step of a structure determination. An ingenious strategy that has facilitated the success of this process has been the use of RNA binding proteins to promote crystallization

[35]. In this respect, U1A RNA-recognition motif 1 (RRM1) has shown considerable efficacy (Table 1). The main strategy has been documented.

Elsewhere in a series of thought-provoking reviews that outline practical methodology [53–55]. The essence of the approach is that the hairpin II (hpII) loop of the U1-snRNA is substituted into a non-conserved loop position of a target RNA (Figure 1a). U1A RRM1 variant Y31H/Q36R (i.e., double-mutant protein or dmU1A) is then added to promote crystallization [56,57]. In the context of the RNA target, the dmU1A variant can provide new protein-protein and protein-RNA contacts that favor formation of a well-ordered crystal lattice. A second function of the dmU1A variant is to serve as a phasing module [53–55]. The known dmU1A-hpII complex can serve as a partial search model for molecular replacement (MR). Alternatively, the five methionine residues of dmU1A can be substituted with selenomethionine for single-wavelength anomalous diffraction (SAD) or multiwavelength anomalous diffraction (MAD) phasing [53–55]. Both SAD and MAD take advantage of resonant x-ray interactions with selenium K-shell electrons, as reviewed elsewhere [58].

Under optimal circumstances, one selenium is sufficient to phase a scattering mass of approximately 75–100 amino acids [60,61], which is equivalent to ~22–33 nucleotides of RNA in terms of scattering strength [62]. To improve the signal available for de novo dmU1A-mediated selenomethionine phasing, we undertook a mutational and RNA binding analysis of dmU1A with the goal of adding additional ordered methionines for experimental, selenium-based phasing. Here we describe the bacterial expression, purification, and affinity analysis of the dmU1A(F37M/F77M) mutant wherein phenylalanine residues were changed to methionine. An RNA binding analysis indicated that the mutated, selenomethionine-labeled protein retains tight binding to hpII, and readily crystallizes in an isolated form, in complex with hpII RNA and in complex with a class II preQ<sub>1</sub> riboswitch from *Lactobacillus rhamnosus* [63]. The crystal structure of the dmU1A(F37M/F77M) variant is described and compared to other dmU1A molecules crystallized in the absence and presence of RNA. A SAD phasing analysis demonstrates that the new Se atoms contribute useful anomalous signals to phasing. In addition, we present evidence that lab-evolved TAR-binding protein variant 6.9 (TBP6.9) derived from dmU1A retains HIV-1 TAR RNA binding when three core phenylalanines are mutated to methionine. This TBP6.9(F34M/F37M/F77M) variant recognizes the 16-nucleotide internal bulged loop of TAR RNA but does not require the apical loop or stem 1a for target RNA binding (Figure 1b). Indeed, our observations herein confirm that TBP6.9(F34M/F37M/F77M) strongly binds the TAR internal bulge in the absence of the apical loop. These observations lay the groundwork for use of TBP6.9(F34M/F37M/F77M) as a crystallization and phasing module when the target RNA is not amenable to integration of a hpII RNA loop. Although a comprehensive analysis of crystallization and phasing awaits future studies, we have made our protein variants available to researchers seeking immediate access to new tools for crystallization and experimental phasing of RNA.

## 2. Experimental Procedures

### 2.1. Expression of Se-Met-labeled dmU1A(F37M/F77M) and TBP6.9(F34M/F37M/F77M)

Synthetic genes for dmU1A, dmU1A(F37M/F77M), and TBP6.9(F34M/F37M/F77M) were produced by DNA synthesis and subcloned into pET28a (GenScript Inc.). Each construct has a tobacco etch virus (TEV) protease cleavage site in place of the thrombin proteolytic site. The Y31H/Q36R mutation that defines the dmU1A variant was integrated into each sequence to promote crystallization [56,57]. Expression of methionine-containing dmU1A and TBP6.9 was described as reported [52]. Selenomethionine labeling of dmU1A(F37M/F77M) and TBP6.9(F34M/F37M/F77M) was performed essentially as described [64]. Each plasmid was transformed into *E. coli* BL21(DE3) cells (Novagen) for expression. Bacteria were streaked onto Luria–Bertani (LB) agar plates containing 50 µg mL<sup>-1</sup> kanamycin and grown at 37 °C. Single colonies were used to inoculate cultures of LB broth for overnight growth. A volume of 60 mL was used for each liter of expression media. The cells were pelleted by centrifugation for 10 min at 2880 *xg* and the LB media was decanted. Cells were suspended in 10 mL of minimal media (MM) containing 1x M9 salts (Na<sub>2</sub>HPO<sub>4</sub>, KH<sub>2</sub>PO<sub>4</sub>, NaCl, and NH<sub>4</sub>Cl), 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 0.5% (*w/v*) glucose, 2 mg L<sup>-1</sup> biotin, 2 mg L<sup>-1</sup> thiamine, and 1x MEM vitamin mix (Thermo Fisher Scientific). The cell suspension was transferred into a larger 1 L volume of MM within a baffled 3 L Fernbach flask sealed with cheesecloth. Cells were grown to an optical density at 600 nm of 0.4, followed by addition of an amino acid mixture (100 mg L<sup>-1</sup> of Lys/Phe/Thr and 50 mg L<sup>-1</sup> of Iso/Leu/Val) to repress methionine synthesis, and selenomethionine (Sigma–Aldrich) (50 mg L<sup>-1</sup>). After reaching an OD at 600 nm of 0.6 the cells were cooled to 20 °C, induced with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG), and expressed for 20 h. The cells were pelleted by centrifugation at 2880*xg* and frozen in N<sub>2</sub> (*J*).

### 2.2. Purification of dmU1A, TBP6.9, dmU1A(F37M/F77M), and TBP6.9(F34M/F37M/F77M)

Cells were thawed in a cell-lysis buffer (CLB) containing: 0.05 M Na-HEPES pH 8.0, 0.3 M NaCl, 0.02 M imidazole pH 8.0, 0.0005 M EDTA, 0.005 M β-mercaptoethanol (β-ME) and 0.01% (*v/v*) Brij35; the cell slurry was made 2 mg mL<sup>-1</sup> in hen egg white lysozyme (VWR). After 20 min, cells were sonicated with a Sonic Dismembrator 60 (Fisher Scientific) and treated with 100 µg mL<sup>-1</sup> DNase and RNase (Roche). The clarified supernatant was bound in batch to Ni-NTA resin (Pierce) equilibrated with CLB. After 2 h of nutation at 4 °C, resin was poured into a 1.5 cm × 10 cm gravity-flow column (CrystalCruz), washed with 40 column volumes of CLB and 5 column volumes of wash buffer (WB) comprising: 0.05 M Na-HEPES pH 8.0, 0.15 M NaCl, 0.04 M imidazole pH 7.5, 0.005 M EDTA, 0.005 M β-ME and 0.01% (*v/v*) Brij35. Fractions were collected in 3 mL volumes using an elution buffer (EB) comprising: 0.05 M Na-HEPES pH 8.0, 0.15 M NaCl, 0.2 M imidazole pH 7.5, 0.005 M EDTA, 0.005 M β-ME and 0.01% (*v/v*) Brij35. Fractions were pooled based on absorption at 280 nm and diluted with WB without imidazole to achieve a final imidazole concentration <0.03 M. The 6xHis-TEV protease [65] was added (1:100 TEV:protein) and the mixture was incubated at 4 °C to remove the 6His tag. After 16 h the reaction was incubated in batch with pre-equilibrated Ni-NTA to retain the uncut protein and TEV; the supernatant was decanted and saved. The protein was loaded at 0.5 mL min<sup>-1</sup> onto a 5 mL HiTrap SP FF column (Cytiva) using an ÄKTA Pure (Cytiva). Samples of dmU1A and

dmU1A(F37M/F77M) were washed and eluted using a linear gradient comprising: 0.15 to 0.85 M NaCl, 0.05 M Na-HEPES pH 8.0, 0.0005 M EDTA, and 0.005 M  $\beta$ -ME. The proteins eluted around 50–70 min corresponding to 50–70% of the maximum salt concentration. TBP6.9 and TBP6.9(F34M/F37M/F77M) were purified similarly except that a steeper salt gradient of 0.15 to 1.5 M NaCl was required for the HiTrap SP FF column to reduce non-specific RNA interactions caused by the arginine-rich content of the  $\beta$ 2– $\beta$ 3 loop; other minor changes were made to the buffers as described [52]. Each concentrated protein sample (~150  $\mu$ m) was subjected to size exclusion on a HiPrep (16/60) Sephacryl S-300 HR column (Cytiva), except TBP6.9(F34M/F37M/F77M), which was sufficiently pure after ion exchange for isothermal titration calorimetry (ITC) analysis. Molecular weight standards were obtained as a kit (Cytiva). Samples of dmU1A (MW 11.4 kDa) and dmU1A(F37M/F77M) (MW 11.5 kDa) each eluted with retention times consistent with monomeric subunits (Supplementary Figure S1). TBP6.9 (MW 11.5 kDa) exhibited longer chromatographic retention than predicted by mass, eluting at or >1 column volume (data not shown). Protein purity was estimated at > 95% by Coomassie-stained SDS-PAGE. The yield from LB media was 2 to 3 mg L<sup>-1</sup> of cells; the yield from MM was ~1 mg L<sup>-1</sup> of cells.

### 2.3. ITC Analysis of dmU1A, dmU1A(F37M/F77M), TBP6.9, and TBP6.9 (F34M/F37M/F77M)

The hpII 24-mer and GNRA-TAR 30-mer strands were made by chemical synthesis (Horizon Discovery). The HIV TAR 27-mer and *Lactobacillus rhamnosus* preQ<sub>1</sub>-II riboswitch 84-mer strands were produced by *in vitro* transcription as described [66]. The sequences were: 5'-GGAGAUCUGAGCCUGGGAGCUCUCUC for TAR and 5'-GGAAGGCCAUUGCACUCCGGUCUCCACGACGAUACUUACUUCCUUUGAUCGUCGUUACUGGCUUCGGCCACAAAGGAGA for the riboswitch. Each RNA strand was purified by denaturing gel electrophoresis, desalted and lyophilized as described [30,66]. For ITC, each RNA strand was dissolved in 0.01 M Na-HEPES pH 7.5 and heated at 65 °C. After 3 min, ITC buffer pre-heated to 65 °C (0.05 M Na-HEPES pH 7.5, 0.05 M NaCl, 0.05 M KCl, and 0.002 M MgCl<sub>2</sub>) was pipetted into the RNA, followed by a 5 min incubation at 65 °C; the sample was cooled overnight to room temperature. Each sample was dialyzed at 4 °C overnight against 4 L of ITC buffer containing 0.005 M  $\beta$ -ME. To attain a buffer match, each pre-concentrated protein sample was co-dialyzed in the same reservoir as the RNA. Protein samples were diluted in used dialysis buffer to concentrations ~10-fold higher than the RNA (7–10  $\mu$ M). ITC measurements were conducted at 20 °C using a PEAQ-ITC (Malvern Panalytical) with protein in the syringe and RNA in the cell [51,52,67]. The time between injections was 150 sec with a total of 19 injections. Thermograms were analyzed with PEAQ-ITC Analysis software using a 1:1 binding model (Supplementary Methods). This model was supported by the stoichiometry (n) of binding (i.e., the ligand–receptor ratio), which produced values close to unity. Other models were tested, including stoichiometries of 2:1, to assess sequential or two sets of binding sites (e.g., resulting from cooperativity). However, these fits were rejected due to very poor fits of the thermogram data to binding isotherms based on  $\chi^2$  values and visual inspection. For dmU1A, the 1:1 model yielded  $\chi^2$  values of 0.3, whereas the two-site and sequential-binding models produced  $\chi^2$  values of 4.0 and 159. Similarly, selenomethionine-labeled dmU1A(F37M/F77M) produced  $\chi^2$  values of 0.02, whereas the two-site and sequential-binding models produced  $\chi^2$  values of 5.0 and 55. As a control for protein oligomerization, dmU1A and

selenomethionine-labeled dmU1A(F37M/F77M) were injected into buffer alone. No significant heats of dilution were detected (Supplementary Figure 2), confirming that there is no appreciable dimer-monomer equilibrium [68]. The monomeric states of these samples at concentrations used for ITC were confirmed further by size-exclusion chromatography (Section 2.3). Formation of RNA-protein complexes in 1:1 ratios is consistent with known co-crystal structures of the parent complexes (Figure 1). For quality control, we maintained  $c$  values of ITC experiments in the preferred range of  $10 < c < 500$ , which is necessary to attain reasonable fits of the equilibrium association constant,  $K_a$ , from the binding isotherm [69]; here,  $c = nK_a[M_T]$ , where  $[M_T]$  is the receptor concentration in the cell [69]. We note titration of TBP6.9 into TAR slightly exceeded 1000 (Figure 3c). A  $c$ -value of 1000 is acceptable here because the fits were restrained by well-known concentrations of ligand and receptor, as well as the known 1:1 ligand-to-receptor stoichiometry [70]. ITC experiments were conducted in duplicate or triplicate.

#### 2.4. Crystallization, x-ray Diffraction Analysis, and Structure Determination of dmU1A(F37M/F77M)

Pure selenomethionine-labelled dmU1A(F37M/F77M) protein was concentrated in a buffer comprising 0.050 M HEPES pH 8.0, 0.15 M NaCl, 0.005 M  $\beta$ -ME, and 0.0005 M EDTA to 5–8 mg mL<sup>-1</sup> based on OD at 280 nm measured using a Nanodrop spectrophotometer (Thermo Scientific) using an extinction coefficient of 0.3319 (mg/mL)<sup>-1</sup> cm<sup>-1</sup>. The protein was subjected to crystallization trials using hanging-drop vapor-diffusion at 20 °C [71] using an in-house incomplete factor screen [66] set up in two 24-well VDX plates (Hampton Research). A volume of 2  $\mu$ L of precipitating agent was added to an equal volume of protein on siliconized glass cover slides (Hampton Research). The best single crystals were attained from 2.7–3.1 M sodium-acetate in a pH range of 6.5 to 7.0. Crystals were cryoprotected in 3.0 M Na-acetate pH 7.0 supplemented with 5% (*v/v*) glycerol. Crystals were flash cooled in thin nylon loops mounted on Crystal Cap Copper magnetic pins (Hampton Research) using a stream of N<sub>2</sub> gas at 100 K produced by a 700 Series Cryostream (Oxford Cryosystems, Ltd). A complete data set was recorded using an X8 Prospector Ultra microfocus I $\mu$ S sealed-tube X-ray source equipped with an APEX II detector (Bruker AXS). A strategy was chosen to record the anomalous diffraction data using a combination of  $\phi$  and  $\omega$  scans. X-ray diffraction data were reduced in the Proteum software package keeping output anomalous pairs separate. The structure was determined by molecular replacement using Phaser [72] implemented in Phenix [73]. The dmU1A coordinates from PDB entry 1urn [56] were employed as a search model. Visualization of electron-density maps and manual model building were conducted in Coot [74]. Additional SAD phasing experiments were conducted using HySS, Phaser EP, Resolve and autobuilding as implemented in Phenix [72,73,75–77]. All cartoons and schematic diagrams derived from coordinates were generated using PyMOL (Schrödinger, LLC). Least-squares superpositions were performed in CCP4 using Lsqkab [78,79].

#### 2.5. Co-crystallization of Se-Met Labelled dmU1A(F37M/F77M) with Hpii and Riboswitch RNA

The dmU1A(F37M/F77M) protein in complex with hpII RNA was crystallized as follows. A volume of 2 L of hairpin folding buffer (HFB) was prepared comprising: 0.010 M Na-

HEPES pH 7.5, 0.05 M NaCl and 0.002 M MgCl<sub>2</sub>. Pure dmU1A(F37M/F77M) was dialyzed against ~2 L of HFB containing 0.005 M β-ME overnight at 4 °C. A 5 mL volume of HFB was heated to 65 °C in a thermostatted water bath. The synthetic 21-mer hpII RNA (Horizon Discovery), described previously [56], was purified as described in Section 4.3. The lyophilized RNA was dissolved in 0.010 M HEPES pH 7.5 to 0.40 mM and heated at 65 °C for 3 min. The RNA was then diluted slowly with 65 °C HFB to 0.040 mM. The dilute 21-mer solution was incubated at 65 °C for 5 min before turning off power to the water bath, allowing the RNA to cool slowly to room temperature overnight. Freshly dialyzed dmU1A(F37M/F77M) at a concentration of 0.04 mM U1A was added dropwise into the folded hpII RNA with gentle vortexing to give a 1:1 molar ratio. The complex was incubated for 30 min at room temperature. The protein-RNA complex was then concentrated using a Pierce 10 kDa MWCO 15 mL concentrator (Thermo Fisher Scientific) pre-washed with HFB. The sample was concentrated to 10 to 12 mg mL<sup>-1</sup> (0.55 to 0.65 mM) at 20 °C to final volume of 50 μL. The complex was crystallized by the hanging-drop vapor-diffusion method [71] using 96 well plates. Screens were established using a Mosquito liquid-handling robot (SPT Labtech) in which 0.5 μL of precipitating agent from a Nuc-Pro HT screen (Jena Bioscience) was added to an equal volume of protein-RNA complex.

The dmU1A(F37M/F77M) protein in complex with an 84-mer preQ<sub>1</sub>-II riboswitch containing the hpII binding site in P1 was crystallized as follows. A volume of 10 mL of riboswitch folding buffer with preQ<sub>1</sub> (RFBQ) was made from 0.010 M Na-cacodylate pH 7.0, 0.005 M MgCl<sub>2</sub> and 0.0015 M preQ<sub>1</sub>. A volume of 5 mL of RFBQ was heated to 65 °C in a thermostatted water bath. A small volume of lyophilized riboswitch RNA (50 μL) stored at -20 °C was dissolved in 0.01 M Na-cacodylate pH 7.0 to 0.5 mM and heated for 2 min at 65 °C. The riboswitch was diluted slowly with 65 °C RFBQ to a final volume of ~5 mL. The solution was incubated at 65 °C for 3 min before turning off the water bath to allow overnight cooling to room temperature. The dmU1A(F37M/F77M) sample freshly dialyzed against HFB containing 0.005 M β-ME was diluted to 0.04 mM. The protein was added dropwise into the folded riboswitch with gentle vortexing to give a 1:1.2 molar ratio. The complex was incubated for 30 min at room temperature. The complex was concentrated as described above to ~10–12 mg mL<sup>-1</sup>. The protein was subjected to crystallization trials using hanging-drop vapor-diffusion at 20 °C [71] using an in-house screen [66], as described in Section 4.4. The best single crystals were observed in high salt and were optimized in 1.6 M Li<sub>2</sub>SO<sub>4</sub>, 0.050 M Tris-HCl pH 8.5, and 0.060–0.080 M MgSO<sub>4</sub>.

All crystals were tested initially for mechanical fragility by probing with a cat whisker mounted on a 27-gauge needle with beeswax or by pressure from a stainless-steel acupuncture needle. Consistent with their high solvent content, our crystals were soft and easily crushed by gentle contact, as noted previously for biomacromolecules [80].

## 2.6. X-ray Diffraction Analysis of dmU1A(F37M/F77M) in Complex with a preQ<sub>1</sub>-II Riboswitch

Co-crystals of dmU1A(F37M/F77M) in complex with the *Lactobacillus rhamnosus* preQ<sub>1</sub>-II riboswitch 84-mer were cryoprotected by passing crystals through a 1:1 mixture of Parabar 10,312 and silicone oils (Hampton Research) before plunging into N<sub>2</sub>(J), as described [81].

Diffraction data were recorded at the Cornell High Energy Synchrotron Source (Ithaca, NY) on beamline A1. Due to the small crystal size, the beam was outfitted with capillary optics to generate an  $\sim 20$   $\mu\text{m}$  micro focus spot. X-ray diffraction data were collected on Q-210 CCD detector (ADSC Inc) at a distance of 300 mm. Data were recorded as  $0.5^\circ$   $\varphi$ -rotations with an exposure time of 30 sec per image. Data were reduced using HKL2000 (HKL Research, Inc) [82]. Representative intensity statistics are provided in Supplementary Table S1.

### 3. Results and Discussion

#### 3.1. Rationale of Phenylalanine-to-methionine Mutants in the dmU1A Hydrophobic Core

First, we considered where to add additional methionines and which residues to mutate. Although frequently encountered at the N-terminus of proteins due to its role as the genetic signal to start translation, methionine is non-polar making it well suited to occupy the hydrophobic cores of globular proteins. Indeed, analysis of T4 lysozyme has suggested that leucine is a first-choice site for methionine substitution in the hydrophobic core along with other residues in the series  $L > F > I > V$  [83]. Accordingly, we sought to identify non-methionine residues within dmU1A that would maintain core packing but with sufficient volume to accommodate selenomethionine, which exhibits a slightly larger van der Waals radius than sulfur (i.e., 1.85 Å versus 2.00 Å) and a longer covalent bond (1.80 Å versus 1.95 Å) [64,83]. We noted that dmU1A contains three well-ordered methionines in its hydrophobic core located above and below each face of the four-stranded  $\beta$ -sheet. Specifically, M72 and M82 are positioned on the face that opposes the hpII RNA binding site, designated site I for convenience here (Figure 2a). M72 and M82 pack against several  $\beta$ -branched amino acids including I12, I14, L19, L26, Leu 30, and I84, as well as aromatic amino acids F34, F37, F75, F77, and Y78. The opposite  $\beta$ -sheet face contains M51, which packs against F56 (Site II) part of the conserved RNP1 motif [56] that recognizes hpII RNA by stacking with Ade69 (Figure 2b). Due to the central role of site II in RNA recognition, we chose to avoid mutations in this region and opted instead for methionine insertion at site I.

Of the site I amino acids, phenylalanine seemed like the best choice for methionine substitution. The aromatic sidechain comprises  $C\beta$ ,  $C\gamma_1$ ,  $C\delta_1$ , and  $C\epsilon_1$  atoms that are analogous to the  $C\beta$ ,  $C\gamma$ ,  $S\delta$ , and  $C\epsilon$  atoms of methionine (Figure 2c). Although not aromatic, methionine has a hydrophobic alkyl chain and a thioether moiety that can engage in  $S-\pi$  interactions that can be more stabilizing than hydrophobic interactions by +1 to +1.5 kcal mol<sup>-1</sup> [84]. Prior  $S-\pi$  analysis showed sulfur to be located above or in-plane with aromatic rings at distances  $\sim 6.0$  Å [84,85]. For dmU1A, site I is characterized by two in-plane  $S-\pi$  interactions from M72 to F37 and M82 to F34 (Figure 2a). Additionally, two above-plane interactions occur from M72 to F75 and M82 to F77. Site II exhibits a single in-plane  $S-\pi$  interaction from M51 to F56 (Figure 2b). To preserve the hydrophobic core, we choose to make the F37M/F77M double mutant. We similarly choose to produce the F34M/F37M/F77M mutants for TBP6.9. While our mutations were predicted to eliminate a handful of existing  $S-\pi$  interactions, we hypothesized that the chosen Phe-to-Met changes would yield compensatory  $S-\pi$  contacts in return (described below).



### 3.2. dmU1A(F37M/F77M) and TBP6.9(F34M/F37M/F77M) Retain Affinity for hpII and TAR RNAs

We produced the dmU1A(F37M/F77M) and TBP6.9(F34M/F37M/F77M) mutants using synthetic genes of the desired sequence for protein expression in *E. coli*. The proteins were engineered with TEV protease cleavage sites to remove the N-terminal poly-histidine tag after immobilized-metal-affinity chromatography (Figure 3a). Proteins were determined to be of high quality based on Coomassie stained polyacrylamide gels and elution as a single species using size-exclusion chromatography. As an additional quality-control step, we subjected each protein to RNA binding analysis using isothermal titration calorimetry (ITC).

As control experiments, we first measured the equilibrium binding affinity of dmU1A for hpII and TBP6.9 binding to HIV TAR. Both proteins were expressed in bacteria and purified essentially as described [51,52]. U1A has been reported to be a tight binder to hpII RNA based on surface plasmon resonance [86]. Indeed, we measured an average  $K_D$  of  $152 \pm 7$  nM and an average  $\Delta G$  of  $-9.15 \pm 0.03$  kcal mol<sup>-1</sup> (Figure 3b and Table 2). The average  $\Delta H$  of  $-32.0 \pm 0.6$  kcal mol<sup>-1</sup> indicates an enthalpy-driven binding process with an unfavorable entropic contribution (average  $-\Delta S$  of  $+22.8 \pm 0.6$  kcal mol<sup>-1</sup>). As for TBP6.9 binding to HIV-1 TAR RNA, we measured an equilibrium binding constant of  $8.5 \pm 0.4$  nM and a  $\Delta G$  of  $-10.9 \pm 0.1$  kcal mol<sup>-1</sup> (Figure 3c and Table 2). Like dmU1A, TAR binding by TBP6.9 occurs via an enthalpy-driven process with a  $\Delta H$  of  $-16.0 \pm 0.1$  kcal mol<sup>-1</sup> and an opposing entropic contribution of  $+5.2 \pm 0.01$  kcal mol<sup>-1</sup>. This thermodynamic profile is similar to prior measurements of the TBP6.9 interaction with TAR reported by our lab [52].

We next evaluated how the dmU1A(F37M/F77M) and TBP6.9(F34M/F37M/F77M) samples affected binding to their respective target RNAs. We rationalized that the most relevant form of each protein would contain selenomethionine. As such, each analysis was performed on protein samples substituted with selenomethionine. The dmU1A(F37M/F77M) variant yielded a  $K_D$  of  $59.7 \pm 11.4$  nM, which was slightly tighter than dmU1A as indicated by a  $K_{rel}$  of 0.39 (Figure 3d and Table 2). We hypothesize that selenomethionine enhances the stability of the hydrophobic core of dmU1A(F37M/F77M) to improve RNA binding. Selenomethionine was shown previously to enhance the stability of T4 lysozyme, while leaving the fold unchanged [83]. Indeed, the enthalpic and entropic contributions to hpII binding agreed with trends observed for dmU1A (Table 2).

In contrast, TBP6.9(F34M/F37M/F77M) produced a  $K_D$  of  $13.5 \pm 2.3$  nM, which was 1.6-fold poorer in TAR binding compared to parental TBP6.9 (Figure 3e and Table 2). The triple mutant produced thermodynamic trends comparable to TBP6.9 with a favorable enthalpy ( $\Delta H$  of  $-14.5 \pm 0.1$  kcal mol<sup>-1</sup>) and unfavorable entropy ( $-\Delta S$  of  $+3.9 \pm 0.1$  kcal mol<sup>-1</sup>) (Table 2). The F34M mutation represents a notable variation between dmU1A(F37M/F77M) and the TBP6.9 triple mutant. The F34M mutation was added in the hydrophobic core to produce a TBP6.9 protein with seven methionines for parity with dmU1A(F37M/F77M) (Figure 3a). This was necessary because the lab-evolution and selection processes used to identify TAR binding entailed saturation mutagenesis in the  $\beta 2$ - $\beta 3$  loop [87]. The resulting TBP6.9 sequence did not retain M51, in contrast to the parental U1A protein (Figure 3a). Accordingly, we choose to integrate the F34M variant at site I of the hydrophobic core (Figure 2a) for reasons delineated above. Overall, our binding analyses demonstrated that

dmU1A(F37M/F77M) and TBP6.9(F34M/F37M/F77M) are capable of high-affinity binding to their respective RNA targets in the presence of seven selenomethionines.

### 3.3. Evidence that TBP6.9(F34M/F37M/F77M) Targets a TAR Sequence Without the Apical Loop

To employ TBP6.9(F34M/F37M/F77M) as a crystallization and phasing module, we asked whether the UGG sequence that composes the TAR apical loop could be replaced with a GNRA tetraloop (Figure 4a). Our rationale was that TBP6.9, and related TBPs, appear to recognize the internal bulged loop of TAR in co-crystal structures but not the apical loop, as illustrated in Figure 1b [51,52]. Accordingly, we first evaluated the affinity of parental TBP6.9 for GNRA-TAR. The results revealed a  $K_D$  of  $87.8 \pm 5.0$  nM corresponding to a  $\Delta G$  of  $-9.5 \pm 0.04$  kcal mol<sup>-1</sup> (Figure 4b and Table 2). Interestingly, the enthalpy of  $-7.8 \pm 0.4$  kcal mol<sup>-1</sup> and entropy of  $-1.7 \pm 0.5$  kcal mol<sup>-1</sup> indicate that both terms contribute favorably to binding (Table 2). This observation pinpoints the TAR apical loop as a region that contributes an unfavorable loss of entropy during TBP6.9 binding to TAR, consistent with the known dynamics of this region in the absence of RNA binding partners (reviewed in [34]). A comparable binding analysis to GNRA-TAR was conducted next for selenomethionine substituted TBP6.9(F34M/F37M/F77M). The results yielded a  $K_D$  of  $69.8 \pm 2.9$  nM, which is nearly indistinguishable from parental TBP6.9 binding to GNRA-TAR (Figure 4c and Table 2); favorable entropic and enthalpic contributions to binding were observed as well. The observation that TBP6.9 recognizes GNRA-TAR with high affinity albeit with 5-fold lower affinity compared to selenomethionine-labeled TBP6.9(F34M/F37M/F77M) binding to TAR is supported by prior experiments conducted on TBP6.7, which differs from TBP6.9 by a Q48R variation in the  $\beta 2$ - $\beta 3$  loop. Specifically, TBP6.7 showed no appreciable loss in TAR affinity based on ELISA analysis when GG in the apical loop was mutated to CC [87]. In contrast, alterations in the closing C30–G34 base pair of the apical loop, or changes at G36, reduced TBP6.7 binding. As such, bases in the apical loop appear dispensable for TBP6.9 binding to TAR. The observed  $\Delta G$  values range from +1.0 to +1.4 kcal mol<sup>-1</sup> when comparing TBP6.9 binding to TAR versus GNRA-TAR (Table 2), which is energetically equivalent to two hydrogen bonds or a S– $\pi$  interaction [51,84]. Although we have not yet proven that TBP6.9(F34M/F37M/F77M) is efficacious as a means to promote crystallization and phasing, our ITC results lead us to speculate that the internal bulged loop of TAR (Figure 1b) can be transplanted into duplex regions capped by essential helical or loop sequences (e.g., loop 4 of the preQ<sub>1</sub>-III riboswitch, which forms an H-type pseudoknot [88]) to facilitate crystallographic analysis.

### 3.4. Crystallization of dmU1A(F37M/F77M) Alone and in Complex with RNA

We next addressed whether the dmU1A(F37M/F77M) variant was amenable to crystallization. U1A containing various mutations has been crystallized alone [57,90,91], with hpII RNA [56,57] and in complex with various target RNAs modified to contain the hpII loop (Figure 1a) in place of functionally dispensable loop sequences (Table 1). Based on a prior report of dmU1A crystallization from malonate in the absence of RNA, and experience within the community using acetate as a crystallization agent [92,93], we identified a trigonal crystal form of the selenomethionine-substituted dmU1A(F37M/F77M) variant grown from 2.7–3.1 M Na-acetate solutions buffered in a pH range from 6.5–7.0

(Figure 5a). Crystals reached a maximum dimension of 60  $\mu\text{m}$  in 1 week at 20  $^{\circ}\text{C}$ , appeared free from growth defects and showed orientation-dependent birefringence. Although acetate at high concentration can serve directly as a cryoprotectant like other organic acid salts such as malonate or Tacsimate<sup>TM</sup> [89] we supplemented the 3.0 M Na-acetate mother liquor with glycerol to produce a 5% ( $v/v$ ) solution prior to flash freezing. In this manner, crystals were preserved for x-ray diffraction analysis (described below).

Our next goal was to crystallize dmU1A(F37M/F77M) in complex with the hpII RNA 21-mer (Figure 1a), as described by Nagai and co-workers for dmU1A and various hpII constructs [57]. As reported for the dmU1A-hpII 21-mer complex, we observed crystals under numerous conditions including some with growth deformations, ‘scissor’ morphologies, and polycrystallinity [57]. These were attained from a variety of conditions including high  $(\text{NH}_4)_2\text{SO}_4$ , Tris-HCl pH 7.0, and spermine-HCl, as reported [57]. However, our best single crystals grew from solutions of polyethylene glycol (PEG) 4000 buffered by MOPS pH 7.0, and low concentrations of NaCl. Crystals also grew from PEG6000 buffered at pH 7.5 by Tris-HCl. The latter crystals formed elongated parallelepipeds that appeared in  $\sim 1$  week at 20  $^{\circ}\text{C}$  and were of sufficient size for x-ray diffraction experiments (Figure 5b). Unfortunately, this crystal habit showed no detectable diffraction on our in-house source. Nagai and co-workers similarly noted that co-crystals of the dmU1A-hpII complex prepared from solutions of PEG “exhibited no diffraction, so this precipitant was rarely used in screens” [57]. Although this is a negative result, it suggests that dmU1A(F37M/F77M) behaves similarly to parental dmU1A and that the crystals shown in Figure 5b are not salt, which would likely show very strong Bragg reflections in a range between  $\sim 10$  to 1  $\text{\AA}$  resolution [80].

We also grew cubic crystals of selenomethionine-substituted dmU1A(F37M/F77M) in complex with a *Lactobacillus rhamnosus* preQ<sub>1</sub>-II riboswitch 84-mer from high concentrations of lithium sulfate (Figure 5c), based on an in-house crystallization screen [66]. The hpII stemloop was substituted into the non-conserved P1 hairpin of the wildtype riboswitch [63] to allow dmU1A(F37M/F77M) binding. Crystals grew in  $\sim 1$  week at 20  $^{\circ}\text{C}$  and lacked detectable birefringence, consistent with a high-symmetry space group. Indeed, x-ray diffraction analysis showed that crystals belong to cubic space group  $F23$  with cell dimensions  $a = b = c = 240.2 \text{ \AA}$  with  $\alpha = \beta = \gamma = 90^{\circ}$ . The estimated solvent content is 55–70% [94], suggesting 2 or 3 complexes per asymmetric unit. Crystals showed diffraction to 4.5  $\text{\AA}$  resolution, but the data were complete to only 5.5  $\text{\AA}$  resolution (Supplementary Table S1). The overall  $R_{\text{sym}}$  was 10.7% (54.3%) with an  $I/\sigma(I)$  of 10.7 (1.7), where parenthetical values indicate the highest resolution shells. Although the poor resolution could be attributable to the small crystal size of  $50 \mu\text{m} \times 50 \mu\text{m} \times 20 \mu\text{m}$ , it is also likely that the crystallization construct requires further optimization. Indeed, methods that employ the dmU1A protein for co-crystallization with RNAs containing hpII call for optimization steps that include varying the length of the hpII helix in single base pair increments to improve x-ray diffraction [53–55]. This process can be labor and resource intensive, and we did not undertake such optimization efforts here. Nonetheless, our results demonstrate that the dmU1A(F37M/F77M) variant labeled with selenomethionine crystallizes in the absence and presence of RNA.

### 3.5. Structure Determination and Quality-control Analysis of the dmU1A(F37M/F77M) Variant

We next sought to understand how the F37M and F77M substitutions influenced the structure of the dmU1A hydrophobic core. Accordingly, we collected X-ray diffraction data on flash cooled crystals depicted in Figure 5a and determined the structure by molecular replacement using the dmU1A component of the dmU1A-hpII complex [56]. X-ray diffraction data were collected to 2.20 Å resolution. The data set was complete and showed a high signal-to-noise ratio of 61.2 (20.2 in the highest resolution shell) with 10-fold redundancy (Table 3). The quality of the data was indicated further by an  $R_{\text{merge}}$  of 2.9% (7.6% in the highest resolution shell) and a correlation coefficient between half data sets ( $CC_{1/2}$ ) of 0.995 in the highest resolution shell. The dmU1A(F37M/F77M) structure was refined to an overall  $R_{\text{cryst}}$  of 16.8%, with a 4.7% difference between  $R_{\text{work}}$  and  $R_{\text{free}}$ , signifying the model is not overfit [99]. Representative electron density reveals a well-ordered core at the site of the F37M and F77M mutations (Figure 6a). All amino acids are positioned in the most-favored regions of the Ramachandran plot and the overall average  $B$ -factor was 20.2 Å<sup>2</sup>. Although some electron density was visible for the first methionine, the polypeptide mainchain was not appreciably visible until V3.

An ordered sodium was a notable feature of electron density maps as well (Figure 6b). This cation is present at 3.0 M levels in the crystallization medium and mediates crystal contacts about the proper threefold axis. The ion receives octahedral coordination from the crystallographically-related carbonyl oxygens of R70, R70', and R70'', as well as three symmetry-related water molecules. The use of ligand groups without formal charges is consistent with the single positive charge carried by this ion [100]. Oxygen-to-sodium distances are 2.4 to 2.5 Å, in accord with prior accounts of alkali metal interactions in proteins [100,101]. The sodium ion is also positioned at the carboxylic end of helix B where it could benefit from the local helix dipole [102]. The observed details of the Na<sup>+</sup> ion coordination sphere are a further indicator of quality for the structural model.

### 3.6. Comparison of the dmU1A(F37M/F77M) Variant to a dmU1A Structure Prepared from Malonate

The preparation of our dmU1A(F37M/F77M) crystals from sodium acetate was reminiscent of a prior study of dmU1A wherein crystals were grown from sodium malonate [90]. To identify similarities and differences in these structures each prepared from organic acid salt we performed an all-atom superposition of these molecules (Figure 6c, yellow versus salmon-colored molecules). The average  $xyz$  displacement was 2.07 Å with the greatest difference observed at helix C, which moves closer to helix B and farther away from the RNA binding surface in the malonate structure. Removal of helix C from the comparison reduced the average displacement to 0.94 Å (i.e., 88–98). We then compared the helix C conformation in our Na-acetate structure to that of the dmU1A-hpII complex [56]. We found that the position of helix C in the dmU1A(F37M/F77M) variant is oriented nearly perpendicularly compared to that of the dmU1A-hpII complex (average displacement of 1.78 Å for all atoms and 0.882 Å when helix C is removed). The helix appears to move as a rigid body to accommodate RNA binding (Figure 6c, yellow versus purple structures), which is necessary to avert a clash between hpII and the helix C conformation observed in our Na-

acetate structure. A take-home message is that organic acid salts do not promote a specific helix C conformation and that the C-terminal region is malleable for RNA recognition, as demonstrated by mainchain displacements as great as 10 Å. Interestingly, the dmU1A-derived TBP6.9 protein does not use its C-terminus to recognize HIV TAR RNA (Figure 1b). In fact, this region of the polypeptide chain was disordered in all TBP-TAR complexes [51,59]. The average all-atom displacement of TBP6.9 superimposed on the dmU1A(F37M/F77M) variant was 1.33 Å for residues 5–91. The greatest conformational difference was localized in the lab-evolved  $\beta 2$ – $\beta 3$  loop that recognizes TAR RNA.

Comparison of the dmU1A(F37M/F77M) variant to the dmU1A structure prepared from Na-malonate revealed that each structure coordinates a cation near R70 of helix B (Figure 6c). Remarkably, both structures employ the same R70 carbonyl oxygens and comparable waters for octahedral ion coordination (e.g., Figure 6b), which is centered on a crystallographic 3-fold axis present in each different space groups. Although a Na<sup>+</sup> ion was assigned to the dmU1A(F37M/F77M) variant grown from 3.0 M Na-acetate, Mg<sup>2+</sup> was modeled for the dmU1A structure prepared from 2.2 M Na-malonate. Both structures show similar ion-to-oxygen coordination distances of  $2.46 \pm 0.10$  for dmU1A and  $2.50 \pm 0.05$  Å for dmU1A(F37M/F77M). Given the identical number of electrons for Mg<sup>2+</sup> and Na<sup>+</sup>, crystallization from 2.2 M Na-malonate, and the observation of slightly longer ion-to-oxygen coordination distances for Na<sup>+</sup> ( $2.53 \pm 0.17$  Å) [103] compared to Mg<sup>2+</sup> ( $2.06 \pm 0.02$  Å) [104,105], it appears that R70 is a nexus for Na<sup>+</sup> coordination in the dmU1A structure. It has been noted that Na<sup>+</sup> is frequently mistaken for Mg<sup>2+</sup> in crystal structures, and the greater formal charge of alkaline earth metals is frequently associated with coordination by one or more acidic ligands [100]. In contrast, no evidence was observed for comparable cation coordination in complexes of dmU1A-hpII [56] or TBP6.9-TAR [52,59]. Nonetheless, we posit that inclusion of Na<sup>+</sup> ions could prove helpful to promote crystallization of dmU1A-RNA complexes.

### 3.7. Site I of dmU1A(F37M/F77M) Maintains the Core Fold but Makes new S– $\pi$ Interactions

We next asked how the F37M and F77M mutations were accommodated in the site I hydrophobic core of U1A. We compared our double mutant to the dmU1A structure grown from Na-malonate. Interestingly, F37M and F77M seem well accommodated by environments that evolved to contain aromatic groups (Figure 6d). The selenomethionine side chain at position 37 appears coplanar with the corresponding phenylalanine aromatic ring of dmU1A with only the  $\epsilon$ -methyl group of the former residue breaking planarity. Nearby M72 appeared to adjust in response to the smaller girth of the F37M change by rotating its  $\epsilon$ -methyl group by  $\sim 180^\circ$  about its  $\chi^3$  dihedral angle. This variation allows a closer positioning of the F75 aromatic ring, which appears to form a new, above-plane Se– $\pi$  contact with F37M. M82 similarly rotates its  $\epsilon$ -methyl group by  $\sim 180^\circ$  to maintain a S– $\pi$  contact with F34. F77M reveals another way to accommodate the selenomethionine sidechain within the volume of an aromatic ring. In this instance, the sidechain atoms do not superpose onto those of phenylalanine but the Se $\delta$  and Ce atoms adopt a coplanar orientation relative to the dmU1A aromatic ring. This conformation allows F77M to engage in a new Se– $\pi$  contact with the edge of Y78. The results demonstrate the remarkable

adaptability of the U1A core and its ability to make numerous small compensatory interactions. These observations provide confidence that the TBP6.9(F34M/F37M/F77M) triple mutant will be accommodated similarly.

### 3.8. Substructure Determination and SAD Phasing of the dmU1A(F37M/F77M) Variant

To be useful as a phasing tool, we next assessed the ability to detect an anomalous signal from the selenomethionine-labeled dmU1A(F37M/F77M) variant. Our in-house data-collection strategy was designed to record the anomalous diffraction data for the Se atoms ( $f''$  of  $1.27 e^-$  at  $1.5418 \text{ \AA}$ , which is 2.3-fold higher than  $f''$  for sulfur at  $0.56 e^-$ ). As an indicator of the resulting anomalous signal quality, we plotted the  $CC_{\text{anom}}$  and  $CC_{\text{anomfit}}$  correlation coefficients for the diffraction data (Figure 7a). A  $CC_{\text{anom}} = 0.15$  is evidence of an anomalous signal. The resolution limit of the useful anomalous signal was estimated based on a linear fit [97], which suggested a cutoff of  $\sim 2.58 \text{ \AA}$  resolution. In contrast, the  $R_{\text{CR}_{\text{anom}}}$  metric indicates a significant anomalous signal when the fit value is  $>1$  [86]. Our analysis suggests a cutoff of  $\sim 2.20 \text{ \AA}$  resolution (Figure 7a), where the  $R_{\text{CR}_{\text{anom}}}$  signal is 1.11 in the highest resolution shell (Table 3).

We next asked whether the Se substructure could be located using a hybrid substructure search as implemented in Phenix [73,76,106]. Indeed, six of the seven Se atoms were located, including those resulting from the F37M and F77M mutations (Figure 7b,c). Sites were input into Phaser-EP [72] for refinement and SAD phasing in both substructure hands. The resulting anomalous difference peaks for Se ranged from  $7.5\sigma$  to  $15.5\sigma$ ; the overall figure of merit was 0.43 and the log-likelihood gain was 329 to  $2.20 \text{ \AA}$  resolution. When the new Se sites from the F37M and F77M mutations were excluded from the phase calculation, the resulting figure of merit and log likelihood gain values were diminished to 0.35 and 141 to  $2.20 \text{ \AA}$  resolution. Although the Se resulting from the F77M mutation produced the lowest anomalous difference peak height, its signal was still comparable to that of Mse97. In contrast, the Se resulting from the F37M mutation produced the second highest anomalous signal after that from Mse82 (Figure 7b,c). Electron density maps calculated using phases from each hand of the Se substructure were subjected next to noise filtering [73,77]. Visual inspection of electron-density maps revealed the correct hand based on clearly recognizable secondary-structure features that readily accommodated the refined dmU1A(F37M/F77M) model (Figure 7d). The experimental SAD phases were used for autobuilding [75], which produced a model with  $R_{\text{work}}$  and  $R_{\text{free}}$  values of 26 and 31%, respectively.

In the model resulting from MR, the six Se atoms refined to an average  $B$ -factor of  $27.4 \text{ \AA}^2$ , including Mse97 at the C-terminus, which is often disordered (i.e., in cases when only three Se atoms were reported in Table 1). The Se  $B$ -factors ranged from  $18.3 \text{ \AA}^2$  for the F37M variant to  $48.1 \text{ \AA}^2$  for Mse97 (Figure 7b). The Se at site F77M yielded a  $B$ -factor of  $38.7 \text{ \AA}^2$ ; this value is comparable to that from Mse51 in site II, which refined to  $39.4 \text{ \AA}^2$ . Overall, this analysis demonstrates that the added selenomethionines are ordered, provide detectable anomalous diffraction signals, and that these sites were useful for phasing dmU1A(F37M/F77M). This variant should prove useful to improve the measurable anomalous and dispersive diffraction signals required for SAD and MAD phasing of dmU1A(F37M/F77M)-RNA complexes.

## 4. Conclusions

Structured RNA molecules can be difficult to crystallize in a form that yields high-resolution diffraction, making it challenging to accurately measure anomalous or dispersive intensity differences required for de novo SAD and MAD phasing. Indeed, numerous instances have been reported in which dmU1A was successful in promoting crystallization but additional experimental phasing approaches were needed beyond MAD or SAD including: Ir(NH<sub>3</sub>)<sub>6</sub>(III)-based phasing [42,48], MR and MAD [46], or MAD with phase combination from other heavy-atom derivatives [40]. A survey of dmU1A-RNA complexes determined by MAD or SAD phasing revealed that only three or four selenomethionines are ordered typically in the protein-RNA complex (Table 1). Assuming an  $f''_{\max}$  of  $4e^-$  and an interwavelength  $f'$  of  $-7e^-$  [107,108], we calculated that the anomalous and dispersive signals from a theoretical dmU1A-RNA 50-mer complex would yield Bijvoet and dispersive differences of 2.3–2.7% and 2.0–2.3% [62], based on the diffraction ratios from 3–4 Se sites. These intensity differences are close to the detectable signal limit for most RNA data sets based on  $R_{\text{sym}}$  metrics. Addition of two ordered selenium atoms to the dmU1A crystallization module as shown here could enhance the Bijvoet and dispersive differences for a hypothetical U1A-RNA 50-mer complex by as much as 20 to 30% [62]. The dmU1A(F37M/F77M) variant could also facilitate phasing of larger U1A-RNA complexes.

We further demonstrated that a dmU1A variant containing the F37M/F77M double mutant retained binding to hpII RNA and could be crystallized in the presence and absence of RNA. The resulting crystal structure indicated that the F37M and F77M mutations were well packed in the hydrophobic core and produced compensatory Se- $\pi$  interactions that were accompanied by modest sidechain adjustments to nearby hydrophobic residues. In total, six ordered selenium atoms were observed in electron density maps, and each produced a measurable anomalous difference signal. The useful limit of the anomalous signal was between 2.58 to 2.20 Å based on  $CC_{\text{anomfit}}$  and  $RCR_{\text{anom}}$  calculations (Table 3), as well as Se SAD phasing from the anomalous differences (Figure 7). Based on the core packing of the dmU1A(F37M/F77M) variant, it is conceivable that additional methionines could be added to augment the selenium diffraction ratio. Indeed, the F34M mutation was included in the TBP6.9(F34M/F37M/F77M) triple mutant to provide six ordered selenium atoms for phasing in the context of the TAR RNA internal bulged loop motif. It is conceivable that sufficiently close Se atoms resulting from core mutations at site I would appear as super-selenium when analyzed at low resolution. This concept is similar to disulfides used in SAD phasing, which appear as single super-sulfurs in anomalous difference Patterson maps [109]; albeit the closest selenium atoms in the dmU1A(F37M/F77M) structure are a distant 5.0 Å apart. Nevertheless, the presence of six ordered selenium atoms is expected to improve the phasing capacity of the already successful dmU1A crystallization and phasing module. Although untested, we posit that the TBP6.9(F34M/F37M/F77M) variant will prove useful for experimental MAD and SAD phasing of structures judiciously engineered to contain the TAR internal bulged loop within non-conserved duplex regions of the target RNA sequence.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments:

We thank members of the Wedekind lab for assistance and helpful comments. We thank Dr. Mohammad Salim for help with RNA synthesis and purification. We thank Maya Lippard for assistance with protein purification and ITC. We thank Dr. Richard Gillilan for assistance with X-ray data collection at the Cornell High Energy Synchrotron Source.

**Funding:** This research was funded by National Institutes of Health grants from the National Institutes of Allergy and Infectious Diseases (R01 AI150462) and the National Institute of General Medical Sciences (R01 GM063162) to J.E.W. X-ray instrumentation was supported by a National Institute of General Medical Sciences award (S10 RR026501) to J.E.W. R.B.B. was supported by a National Institute of General Medical Sciences training grant (T32 GM135134).

## References

1. Cech TR; Steitz JA The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* 2014, 157, 77–94. [PubMed: 24679528]
2. Pennisi E Encode project writes eulogy for junk DNA. *Science* 2012, 337, 1159–1161. [PubMed: 22955811]
3. Djebali S; Davis CA; Merkel A; Dobin A; Lassmann T; Mortazavi A; Tanzer A; Lagarde J; Lin W; Schlesinger F; et al. Landscape of transcription in human cells. *Nature* 2012, 489, 101–108. [PubMed: 22955620]
4. Warner KD; Hajdin CE; Weeks KM Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov* 2018, 17, 547–558. [PubMed: 29977051]
5. Hopkins AL; Groom CR The druggable genome. *Nat. Rev. Drug Discov* 2002, 1, 727–730. [PubMed: 12209152]
6. Cooper TA; Wan L; Dreyfuss G RNA and disease. *Cell* 2009, 136, 777–793. [PubMed: 19239895]
7. Stelzer AC; Frank AT; Kratz JD; Swanson MD; Gonzalez-Hernandez MJ; Lee J; Andricioaei I; Markovitz DM; Al-Hashimi HM Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol* 2011, 7, 553–559. [PubMed: 21706033]
8. Disney MD; Dwyer BG; Childs-Disney JL Drugging the RNA world. *Cold Spring Harb. Perspect. Biol* 2018, 10, a034769. [PubMed: 30385607]
9. McCown PJ; Corbino KA; Stav S; Sherlock ME; Breaker RR Riboswitch diversity and distribution. *RNA* 2017, 23, 995–1011. [PubMed: 28396576]
10. Shi H; Moore PB The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: A classic structure revisited. *RNA* 2000, 6, 1091–1105. [PubMed: 10943889]
11. Fourmy D; Recht MI; Blanchard SC; Puglisi JD Structure of the a site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* 1996, 274, 1367–1371. [PubMed: 8910275]
12. Pley HW; Flaherty KM; McKay DB Three-dimensional structure of a hammerhead ribozyme. *Nature* 1994, 372, 68–74. [PubMed: 7969422]
13. Wedekind JE; McKay DB Crystal structure of a lead-dependent ribozyme revealing metal binding sites relevant to catalysis. *Nat. Struct. Biol* 1999, 6, 261–268. [PubMed: 10074945]
14. Theimer CA; Blois CA; Feigon J Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell* 2005, 17, 671–682. [PubMed: 15749017]
15. Brown JA; Bulkley D; Wang J; Valenstein ML; Yario TA; Steitz TA; Steitz JA Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol* 2014, 21, 633–640. [PubMed: 24952594]
16. Toor N; Keating KS; Taylor SD; Pyle AM Crystal structure of a self-spliced group II intron. *Science* 2008, 320, 77–82. [PubMed: 18388288]
17. Cate JH; Gooding AR; Podell E; Zhou K; Golden BL; Kundrot CE; Cech TR; Doudna JA Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* 1996, 273, 1678–1685. [PubMed: 8781224]



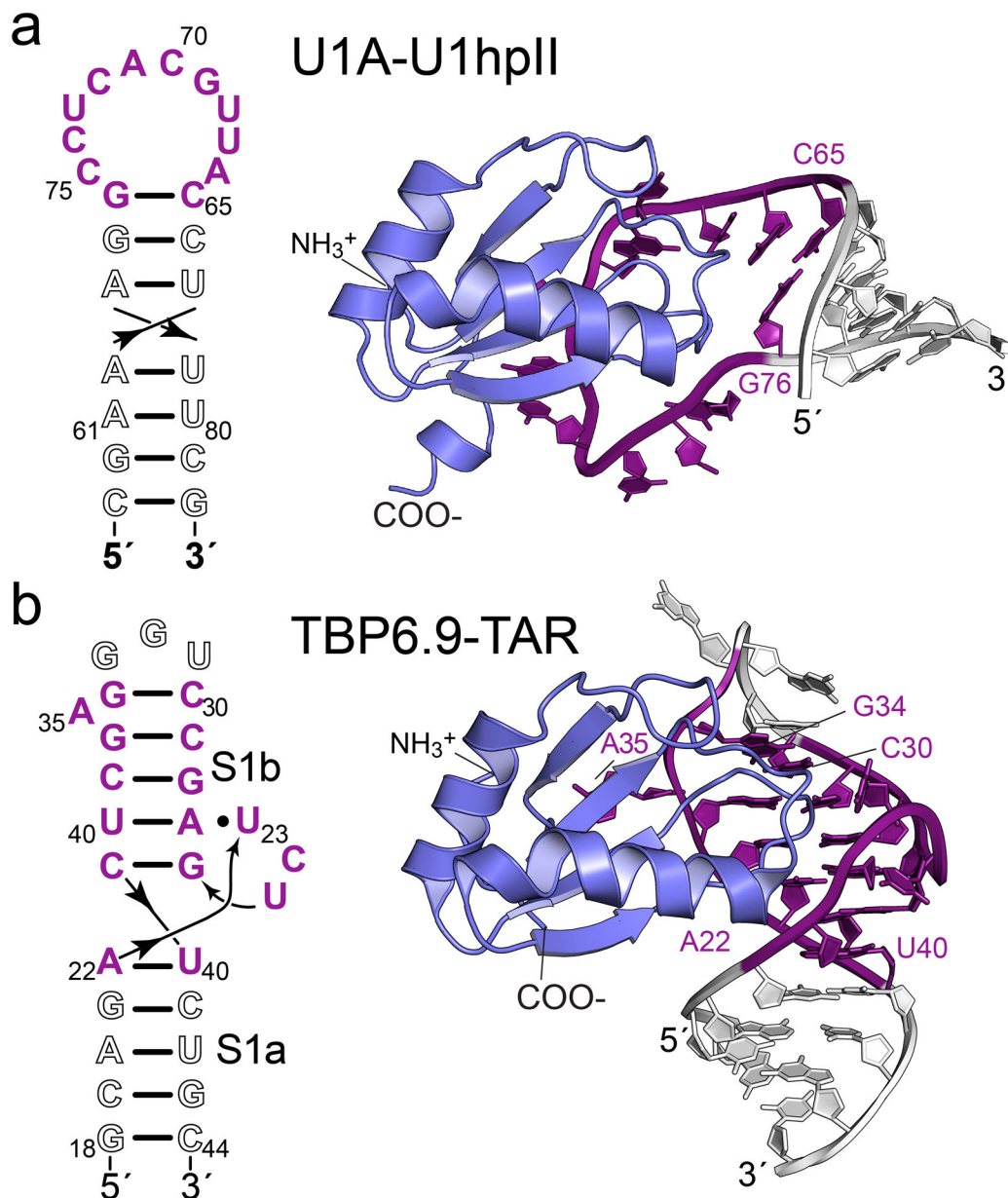
18. Correll CC; Freeborn B; Moore PB; Steitz TA Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* 1997, 91, 705–712. [PubMed: 9393863]
19. Correll CC; Munishkin A; Chan YL; Ren Z; Wool IG; Steitz TA Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. Natl. Acad. Sci. USA* 1998, 95, 13436–13441. [PubMed: 9811818]
20. Golden BL; Kim H; Chase E Crystal structure of a phage twort group I ribozyme-product complex. *Nat. Struct. Mol. Biol* 2005, 12, 82–89. [PubMed: 15580277]
21. Krasilnikov AS; Yang X; Pan T; Mondragon A Crystal structure of the specificity domain of ribonuclease P. *Nature* 2003, 421, 760–764. [PubMed: 12610630]
22. Torres-Larios A; Swinger KK; Krasilnikov AS; Pan T; Mondragon A Crystal structure of the RNA component of bacterial ribonuclease P. *Nature* 2005, 437, 584–587. [PubMed: 16113684]
23. Krasilnikov AS; Mondragon A On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA* 2003, 9, 640–643. [PubMed: 12756321]
24. Nissen P; Ippolito JA; Ban N; Moore PB; Steitz TA RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc. Natl. Acad. Sci. USA* 2001, 98, 4899–4903. [PubMed: 11296253]
25. Hendrix DK; Brenner SE; Holbrook SR RNA structural motifs: Building blocks of a modular biomolecule. *Q. Rev. Biophys* 2005, 38, 221–243. [PubMed: 16817983]
26. Lilley DM The k-turn motif in riboswitches and other RNA species. *Biochim. Biophys. Acta* 2014, 1839, 995–1004. [PubMed: 24798078]
27. Brown JA Unraveling the structure and biological functions of RNA triple helices. *Wiley Interdiscip. Rev. RNA* 2020, 11, e1598. [PubMed: 32441456]
28. Cate JH; Gooding AR; Podell E; Zhou K; Golden BL; Szewczak AA; Kundrot CE; Cech TR; Doudna JA RNA tertiary structure mediation by adenosine platforms. *Science* 1996, 273, 1696–1699. [PubMed: 8781229]
29. Kang M; Peterson R; Feigon J Structural insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol. Cell* 2009, 33, 784–790. [PubMed: 19285444]
30. Schroeder GM; Dutta D; Cavender CE; Jenkins JL; Pritchett EM; Baker CD; Ashton JM; Mathews DH; Wedekind JE Analysis of a preQ<sub>1</sub>-I riboswitch in effector-free and bound states reveals a metabolite-programmed nucleobase-stacking spine that controls gene regulation. *Nucleic Acids Res.* 2020, 48, 8146–8164. [PubMed: 32597951]
31. Pley HW; Flaherty KM; McKay DB Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* 1994, 372, 111–113. [PubMed: 7526219]
32. Ennifar E; Walter P; Ehresmann B; Ehresmann C; Dumas P Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol* 2001, 8, 1064–1068. [PubMed: 11702070]
33. Alam S; Grum-Tokars V; Krucinska J; Kundracik ML; Wedekind JE Conformational heterogeneity at position U37 of an all-RNA hairpin ribozyme with implications for metal binding and the catalytic structure of the S-turn. *Biochemistry* 2005, 44, 14396–14408. [PubMed: 16262240]
34. Chavali SS; Bonn-Breach R; Wedekind JE Face-time with TAR: Portraits of an HIV-1 RNA with diverse modes of effector recognition relevant for drug discovery. *J. Biol. Chem* 2019, 294, 9326–9341. [PubMed: 31080171]
35. Shao Y; Huang H; Qin D; Li NS; Koide A; Staley JP; Koide S; Kossiakoff AA; Piccirilli JA Specific recognition of a single-stranded RNA sequence by a synthetic antibody fragment. *J. Mol. Biol* 2016, 428, 4100–4114. [PubMed: 27593161]
36. Sun A; Gasser C; Li F; Chen H; Mair S; Krasheninina O; Micura R; Ren A Sam-VI riboswitch structure and signature for ligand discrimination. *Nat. Commun* 2019, 10, 5728. [PubMed: 31844059]
37. Chen H; Egger M; Xu X; Flemmich L; Krasheninina O; Sun A; Micura R; Ren A Structural distinctions between NAD<sup>+</sup> riboswitch domains 1 and 2 determine differential folding and ligand binding. *Nucleic Acids Res.* 2020, 48, 12394–12406. [PubMed: 33170270]

38. Ke A; Zhou K; Ding F; Cate JH; Doudna JA A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* 2004, 429, 201–205. [PubMed: 15141216]
39. Ferre-D'Amare AR; Zhou K; Doudna JA Crystal structure of a hepatitis delta virus ribozyme. *Nature* 1998, 395, 567–574. [PubMed: 9783582]
40. Adams PL; Stahley MR; Kosek AB; Wang J; Strobel SA Crystal structure of a self-splicing group I intron with both exons. *Nature* 2004, 430, 45–50. [PubMed: 15175762]
41. Ren A; Xue Y; Peselis A; Serganov A; Al-Hashimi HM; Patel DJ Structural and dynamic basis for low-affinity, high-selectivity binding of L-glutamine by the glutamine riboswitch. *Cell Rep.* 2015, 13, 1800–1813. [PubMed: 26655897]
42. Butler EB; Xiong Y; Wang J; Strobel SA Structural basis of cooperative ligand binding by the glycine riboswitch. *Chem. Biol* 2011, 18, 293–298. [PubMed: 21439473]
43. Jones CP; Ferre-D'Amare AR Crystal structure of a c-di-AMP riboswitch reveals an internally pseudo-dimeric RNA. *EMBO J.* 2014, 33, 2692–2703. [PubMed: 25271255]
44. Kulshina N; Baird NJ; Ferre-D'Amare AR Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. *Nat. Struct. Mol. Biol* 2009, 16, 1212–1217. [PubMed: 19898478]
45. Smith KD; Lipchock SV; Ames TD; Wang J; Breaker RR; Strobel SA Structural basis of ligand binding by a c-di-GMP riboswitch. *Nat. Struct. Mol. Biol* 2009, 16, 1218–1223. [PubMed: 19898477]
46. Xiao H; Murakami H; Suga H; Ferre-D'Amare AR Structural basis of specific tRNA aminoacylation by a small *in vitro* selected ribozyme. *Nature* 2008, 454, 358–361. [PubMed: 18548004]
47. Xiao H; Edwards TE; Ferre-D'Amare AR Structural basis for specific, high-affinity tetracycline binding by an *in vitro* evolved aptamer and artificial riboswitch. *Chem. Biol* 2008, 15, 1125–1137. [PubMed: 18940672]
48. Shechner DM; Grant RA; Bagby SC; Koldobskaya Y; Piccirilli JA; Bartel DP Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science* 2009, 326, 1271–1275. [PubMed: 19965478]
49. Cochrane JC; Lipchock SV; Smith KD; Strobel SA Structural and chemical basis for glucosamine 6-phosphate binding and activation of the glmS ribozyme. *Biochemistry* 2009, 48, 3239–3246. [PubMed: 19228039]
50. Huang L; Lilley DM Structure of a rare non-standard sequence k-turn bound by L7ae protein. *Nucleic Acids Res.* 2014, 42, 4734–4740. [PubMed: 24482444]
51. Belashov IA; Crawford DW; Cavender CE; Dai P; Beardslee PC; Mathews DH; Pentelute BL; McNaughton BR; Wedekind JE Structure of HIV TAR in complex with a lab-evolved RRM provides insight into duplex RNA recognition and synthesis of a constrained peptide that impairs transcription. *Nucleic Acids Res.* 2018, 46, 6401–6415. [PubMed: 29961805]
52. Chavali SS; Mali SM; Jenkins JL; Fasan R; Wedekind JE Co-crystal structures of HIV TAR RNA bound to lab-evolved proteins show key roles for arginine relevant to the design of cyclic peptide TAR inhibitors. *J. Biol. Chem* 2020, 295, 16470–16486. [PubMed: 33051202]
53. Ferre-D'Amare AR Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. *Methods* 2010, 52, 159–167. [PubMed: 20554048]
54. Ferre-D'Amare AR; Doudna JA Crystallization and structure determination of a hepatitis delta virus ribozyme: Use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol* 2000, 295, 541–556. [PubMed: 10623545]
55. Ferre-D'Amare AR Use of the U1A protein to facilitate crystallization and structure determination of large RNAs. *Methods Mol. Biol* 2016, 1320, 67–76. [PubMed: 26227038]
56. Oubridge C; Ito N; Evans PR; Teo CH; Nagai K Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 1994, 372, 432–438. [PubMed: 7984237]
57. Oubridge C; Ito N; Teo CH; Fearnley I; Nagai K Crystallisation of RNA-protein complexes. II. The application of protein engineering for crystallisation of the U1A protein-RNA complex. *J. Mol. Biol* 1995, 249, 409–423. [PubMed: 7783201]

58. Hendrickson WA Anomalous diffraction in crystallographic phase evaluation. *Q. Rev. Biophys* 2014, 47, 49–93. [PubMed: 24726017]
59. Chavali SS; Cavender CE; Mathews DH; Wedekind JE Arginine forks are a widespread motif to recognize phosphate backbones and guanine nucleobases in the RNA major groove. *J. Am. Chem. Soc* 2020, 142, 19835–19839. [PubMed: 33170672]
60. Hendrickson W Maturation of MAD phasing for the determination of macromolecular structures. *J. Synchrotron Radiat* 1999, 6, 845–851.
61. Hendrickson WA; Ogata CM [28] Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* 1997, 276, 494–523. [PubMed: 27799111]
62. Smith JL Multiwavelength anomalous diffraction in macromolecular crystallography. In *Direct Methods for Solving Macromolecular Structures NATO ASI Series; Fortier S, Ed.; Series C: Mathematical and Physical Sciences; Springer: Dordrecht, The Netherlands, 1998; pp. 211–225.*
63. Liberman JA; Salim M; Krucinska J; Wedekind JE Structure of a class II preQ<sub>1</sub> riboswitch reveals ligand recognition by a new fold. *Nat. Chem. Biol* 2013, 9, 353–355. [PubMed: 23584677]
64. Doublet S Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* 1997, 276, 523–530.
65. Kapust RB; Tozser J; Fox JD; Anderson DE; Cherry S; Copeland TD; Waugh DS Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* 2001, 14, 993–1000. [PubMed: 11809930]
66. Lippa GM; Liberman JA; Jenkins JL; Krucinska J; Salim M; Wedekind JE Crystallographic analysis of small ribozymes and riboswitches. *Methods Mol. Biol* 2012, 848, 159–184. [PubMed: 22315069]
67. Liberman JA; Bogue JT; Jenkins JL; Salim M; Wedekind JE ITC analysis of ligand binding to preQ(1) riboswitches. *Methods Enzymol.* 2014, 549, 435–450. [PubMed: 25432759]
68. Fernando H; Chin C; Rosgen J; Rajarathnam K Dimer dissociation is essential for Interleukin-8 (IL-8) binding to CXCR1 receptor. *J. Biol Chem* 2004, 279, 36175–36178. [PubMed: 15252057]
69. Wiseman T; Williston S; Brandts JF; Lin LN Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem* 1989, 179, 131–137. [PubMed: 2757186]
70. Turnbull WB; Daranas AH On the value of c: Can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc* 2003, 125, 14859–14866. [PubMed: 14640663]
71. Benvenuti M; Mangani S Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nat. Protoc* 2007, 2, 1633–1651. [PubMed: 17641629]
72. McCoy AJ; Grosse-Kunstleve RW; Adams PD; Winn MD; Storoni LC; Read RJ Phaser crystallographic software. *J. Appl. Crystallogr* 2007, 40, 658–674. [PubMed: 19461840]
73. Adams PD; Afonine PV; Bunkoczi G; Chen VB; Davis IW; Echols N; Headd JJ; Hung LW; Kapral GJ; Grosse-Kunstleve RW; et al. Phenix: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr* 2010, 66, 213–221. [PubMed: 20124702]
74. Emsley P; Lohkamp B; Scott WG; Cowtan K Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr* 2010, 66, 486–501. [PubMed: 20383002]
75. Afonine PV; Grosse-Kunstleve RW; Echols N; Headd JJ; Moriarty NW; Mustyakimov M; Terwilliger TC; Urzhumtsev A; Zwart PH; Adams PD Towards automated crystallographic structure refinement with Phenix.Refine. *Acta Crystallogr. D Biol. Crystallogr* 2012, 68, 352–367. [PubMed: 22505256]
76. Grosse-Kunstleve RW; Adams PD Substructure search procedures for macromolecular structures. *Acta Crystallogr. D Biol. Crystallogr* 2003, 59, 1966–1973. [PubMed: 14573951]
77. Terwilliger T Solve and Resolve: Automated structure solution, density modification and model building. *J. Synchrotron Radiat* 2004, 11, 49–52. [PubMed: 14646132]
78. Winn MD; Ballard CC; Cowtan KD; Dodson EJ; Emsley P; Evans PR; Keegan RM; Krissinel EB; Leslie AG; McCoy A; et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr* 2011, 67, 235–242. [PubMed: 21460441]
79. Kabsch W A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A* 1978, 32, 922–923.

80. McPherson A Introduction to protein crystallization. *Methods* 2004, 34, 254–265. [PubMed: 15325645]
81. Spitale RC; Torelli AT; Krucinska J; Bandarian V; Wedekind JE The structural basis for recognition of the preQ<sub>0</sub> metabolite by an unusually small riboswitch aptamer domain. *J. Biol. Chem* 2009, 284, 11012–11016. [PubMed: 19261617]
82. Otwinowski Z; Minor W Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 1997, 276, 307–326.
83. Gassner NC; Baase WA; Hausrath AC; Matthews BW Substitution with selenomethionine can enhance the stability of methionine-rich proteins. *J. Mol. Biol* 1999, 294, 17–20. [PubMed: 10556025]
84. Valley CC; Cembran A; Perlmutter JD; Lewis AK; Labello NP; Gao J; Sachs JN The methionine-aromatic motif plays a unique role in stabilizing protein structure. *J. Biol. Chem* 2012, 287, 34979–34991. [PubMed: 22859300]
85. Reid KSC; Lindley PF; Thornton JM Sulphur-aromatic interactions in proteins. *FEBS Lett.* 1985, 190, 209–213.
86. Katsamba PS; Bayramyan M; Haworth IS; Myszka DG; Laird-Offringa IA Complex role of the beta 2-beta 3 loop in the interaction of U1A with U1 hairpin II RNA. *J. Biol. Chem* 2002, 277, 33267–33274. [PubMed: 12082087]
87. Crawford DW; Blakeley BD; Chen PH; Sherpa C; Le Grice SF; Laird-Offringa IA; McNaughton BR An evolved RNA recognition motif that suppresses HIV-1 Tat/TAR-dependent transcription. *ACS Chem. Biol* 2016, 11, 2206–2215. [PubMed: 27253715]
88. Liberman JA; Suddala KC; Aytenfisu A; Chan D; Belashov IA; Salim M; Mathews DH; Spitale RC; Walter NG; Wedekind JE Structural analysis of a class III preQ<sub>1</sub> riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. *Proc. Natl. Acad. Sci. USA* 2015, 112, E3485–E3494. [PubMed: 26106162]
89. Blose JM; Proctor DJ; Veeraraghavan N; Misra VK; Bevilacqua PC Contribution of the closing base pair to exceptional stability in RNA tetraloops: Roles for molecular mimicry and electrostatic factors. *J. Am. Chem. Soc* 2009, 131, 8474–8484. [PubMed: 19476351]
90. Rupert PB; Xiao H; Ferre-D'Amare AR U1A RNA-binding domain at 1.8 Å resolution. *Acta Crystallogr. D Biol. Crystallogr* 2003, 59, 1521–1524. [PubMed: 12876372]
91. Rosenbach H; Victor J; Borggrafe J; Biehl R; Steger G; Eitzkorn M; Span I Expanding crystallization tools for nucleic acid complexes using U1A protein variants. *J. Struct. Biol* 2020, 210, 107480. [PubMed: 32070773]
92. Wedekind JE; McKay DB Purification, crystallization, and x-ray diffraction analysis of small ribozymes. *Methods Enzymol.* 2000, 317, 149–168. [PubMed: 10829279]
93. McPherson A A comparison of salts for the crystallization of macromolecules. *Protein Sci.* 2001, 10, 418–422. [PubMed: 11266627]
94. Weichenberger CX; Rupp B Ten years of probabilistic estimates of biocrystal solvent content: New insights via nonparametric kernel density estimate. *Acta Crystallogr. D Biol. Crystallogr* 2014, 70, 1579–1588. [PubMed: 24914969]
95. Weiss MS Global indicators of x-ray data quality. *J. Appl Cryst* 2001, 34, 130–135.
96. Karplus PA; Diederichs K Linking crystallographic model and data quality. *Science* 2012, 336, 1030–1033. [PubMed: 22628654]
97. Evans PR; Murshudov GN How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr* 2013, 69, 1204–1214. [PubMed: 23793146]
98. Chen VB; Arendall WB 3rd; Headd JJ; Keedy DA; Immormino RM; Kapral GJ; Murray LW; Richardson JS; Richardson DC Molprobity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr* 2010, 66, 12–21. [PubMed: 20057044]
99. Kleywegt GJ; Brunger AT Checking your imagination: Applications of the free R value. *Structure* 1996, 4, 897–904. [PubMed: 8805582]
100. Zheng H; Cooper DR; Porebski PJ; Shabalin IG; Handing KB; Minor W Checkmymetal: A macromolecular metal-binding validation tool. *Acta Crystallogr. D Struct. Biol* 2017, 73, 223–233. [PubMed: 28291757]

101. Wedekind JE; Trame CB; Dorywalska M; Koehl P; Raschke TM; McKee M; FitzGerald D; Collier RJ; McKay DB Refined crystallographic structure of *Pseudomonas aeruginosa* exotoxin A and its implications for the molecular mechanism of toxicity. *J. Mol. Biol* 2001, 314, 823–837. [PubMed: 11734000]
102. Aqvist J; Luecke H; Quioco FA; Warshel A Dipoles localized at helix termini of proteins stabilize charges. *Proc. Natl. Acad. Sci. USA* 1991, 88, 2026–2030. [PubMed: 2000410]
103. Wedekind JE; Liberman JA; Jenkins JL; Salim M Metal dependence of ligand binding and heavy-atom derivatization of evolutionarily distinct preQ<sub>1</sub> riboswitches. In *Chemical Biology of Nucleic Acids: Fundamentals and Clinical Applications*; Erdmann V, Markiewicz W, Barciszewski J, Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 423–440.
104. Wedekind JE Metal ion binding and function in natural and artificial small RNA enzymes from a structural perspective. In *Structural and Catalytic Roles of Metal Ions in RNA*; Sigel A, Sigel H, Sigel R, Eds.; Walter de Gruyter GmbH & Co KG: Berlin/Heidelberg, Germany, 2011. pp. 299–345.
105. Bonn-Breach R; Gu Y; Jenkins J; Fasan R; Wedekind J Structure of sonic hedgehog protein in complex with zinc(II) and magnesium(II) reveals ion-coordination plasticity relevant to peptide drug design. *Acta Crystallogr. D Struct. Biol* 2019, 75, 969–979. [PubMed: 31692471]
106. McCoy AJ; Storoni LC; Read RJ Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallogr. D Biol. Crystallogr* 2004, 60, 1220–1228. [PubMed: 15213383]
107. Cromer DT Anomalous dispersion corrections computed from self-consistent field relativistic Dirac–Slater wave functions. *Acta Crystallogr.* 1965, 18, 17–23.
108. Cromer DT; Liberman D Relativistic calculation of anomalous scattering factors of x rays. *J. Chem. Phys* 1970, 53, 1891–1898.
109. Sheldrick GM Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr* 2010, 66, 479–485. [PubMed: 20383001]



**Figure 1.** Schematic and cartoon diagrams of RNA sequences targeted by U1A and the related HIV TAR-binding protein (TBP) variant 6.9. **(a)** (Left) Human U1A RRM1 binds the short hairpin II (hpII) sequence of U1 snRNA. Binding occurs primarily in the single-stranded loop region closed by the canonical C65-G76 base pair. This region is necessary and sufficient for U1A binding. (Right) Cartoon drawing of the double-mutant protein (dm)U1A-hpII 21-mer co-crystal structure [56], emphasizing protein recognition of the hpII loop (PDB entry 1urn); the duplex region (white) is not involved in hpII binding and can be replaced by a target sequence to promote RNA crystallization. **(b)** (Left) Lab-evolved TBP6.9 recognizes TAR RNA at the internal bulged loop, which includes a U23 A27-U40 base triple and bulged A35 within helical stem S1b [51,52]. An arginine fork is a key

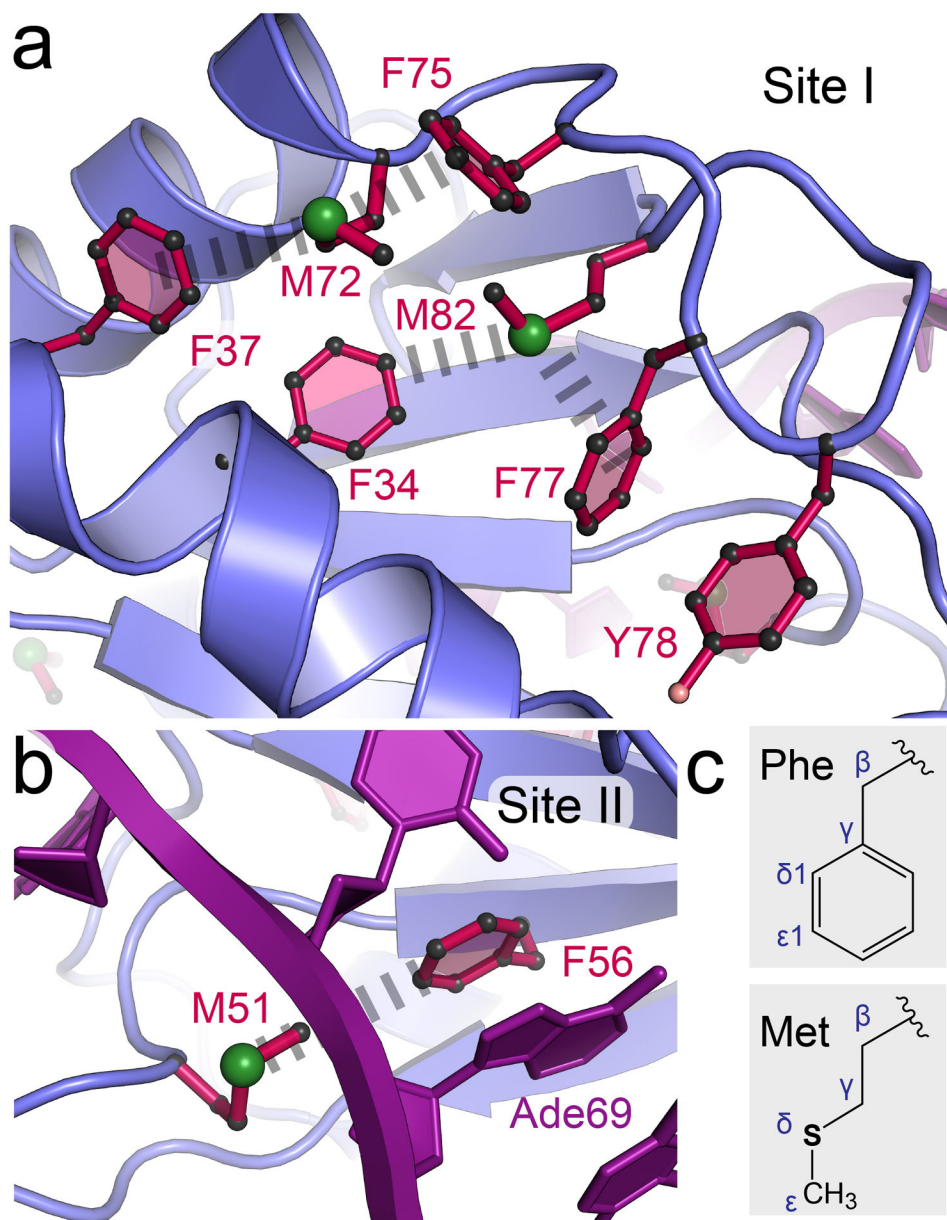
determinant of recognition at the G26 major-groove edge and the U23 backbone [59].  
(*Right*) Cartoon drawing of the TBP6.9-TAR co-crystal structure [52], emphasizing protein recognition of the TAR internal bulged loop (PDB entry 5xh0); the apical loop and S1a (white) are not involved in TAR binding and are hypothesized herein to allow replacement within a duplex target sequence to promote RNA crystallization and phasing.

Author Manuscript

Author Manuscript

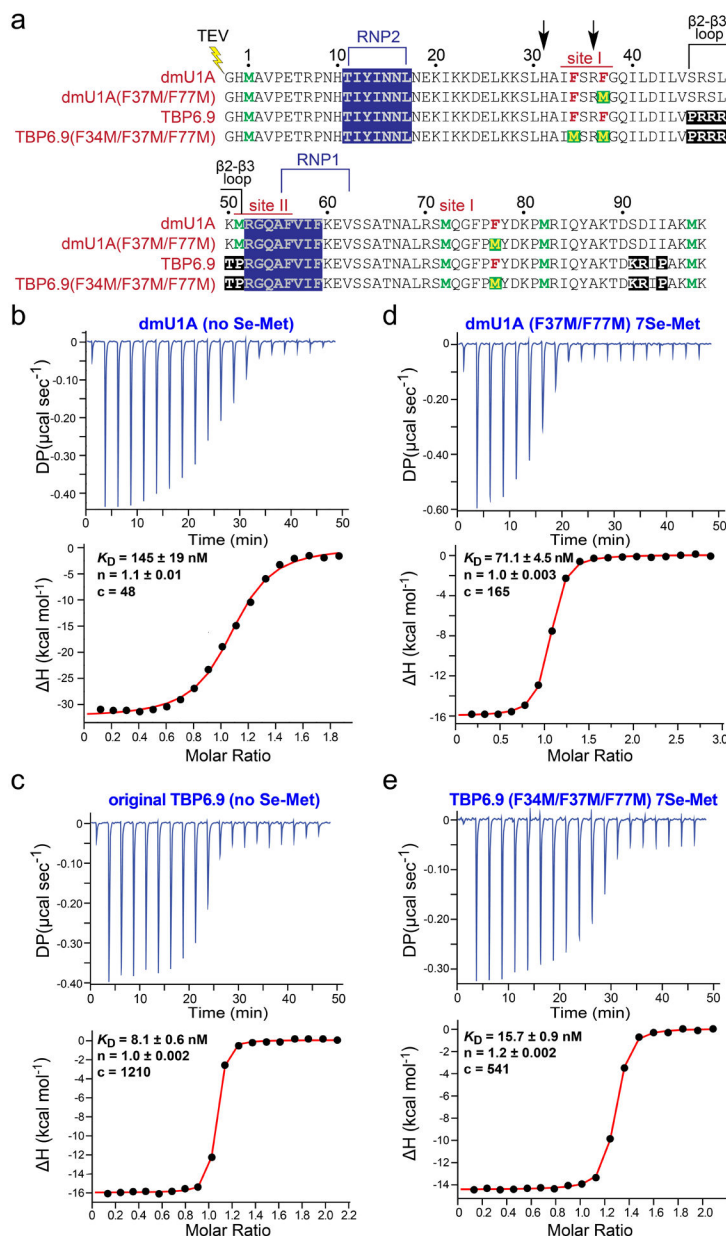
Author Manuscript

Author Manuscript



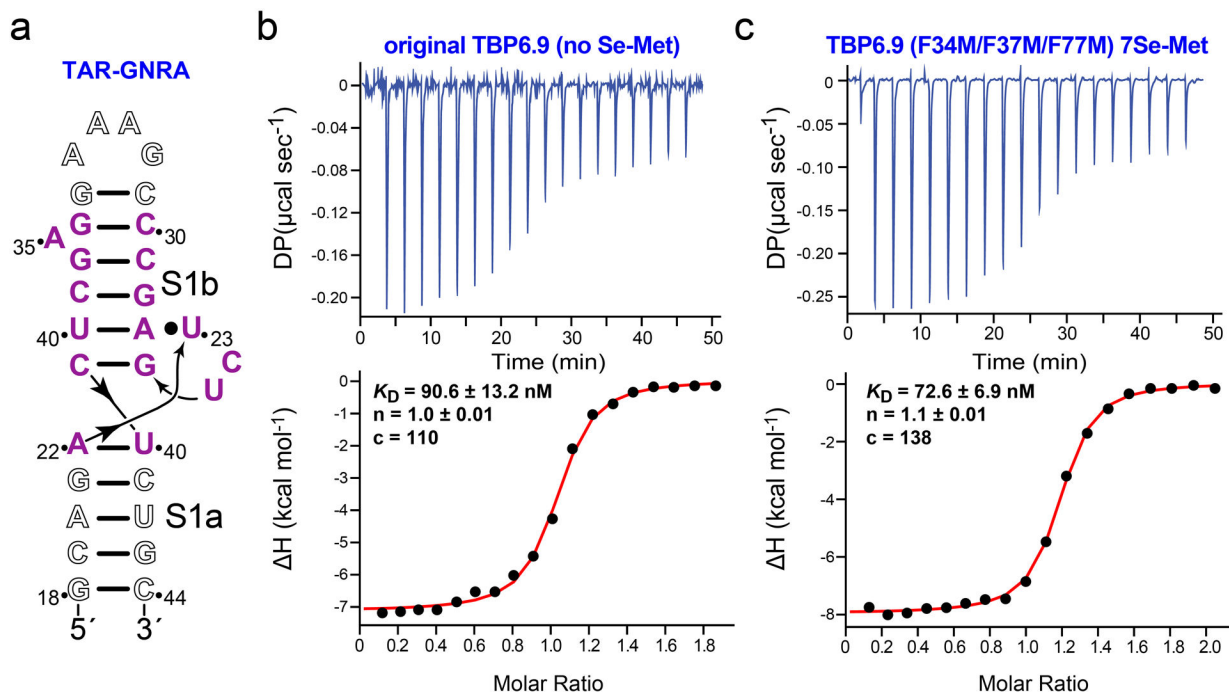
**Figure 2.** Rationale for Phe-to-Met mutations in U1A based on S- $\pi$  interactions in the hydrophobic core. **(a)** Close-up view of the ‘site I’ hydrophobic core depicting aromatic and methionine residues. Hash marks indicate putative S- $\pi$  interactions between specific Phe and Met residues. Here and in panel **b**, protein coordinates were derived from PDB entry 1urn, referred to as dmU1A, due to the double-mutations made to the protein for crystallization [56,57]. **(b)** Close-up view of the site II hydrophobic core on the side of the  $\beta$ -sheet opposite site I. F56 interacts directly with M51 and Ade69 of hpII. **(c)** Drawing of the phenylalanine and methionine sidechains to compare their relative sizes and atomic compositions.



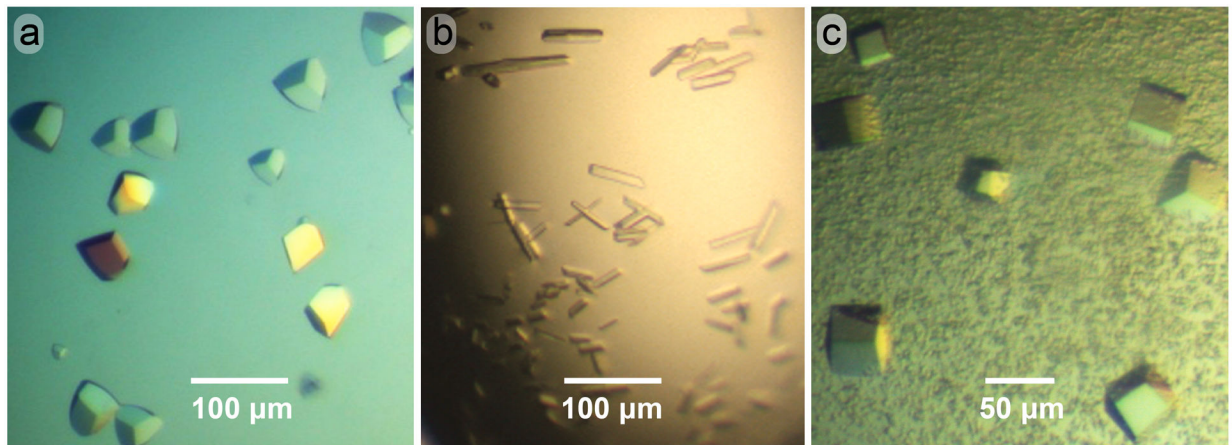


**Figure 3.** Sequences of dmU1A variants and thermograms of RNA binding to dmU1A(F37M/F77M) and TBP6.9(F34M/F37M/F77M). (a) The sequence of human dmU1A crystallized by Nagai [56]; Y31H and Q36R mutants that promote crystallization [57] are depicted by arrows. The conserved RRM binding regions that interact with hpII are labeled RNP2 and RNP1. Met amino acids of parental sequences are green; site I Phe sites described in Figure 2a and those selected for mutagenesis are maroon. Phenylalanine-to-methionine mutations made for this investigation are depicted as green boxes with yellow letters. The lab-evolved β2–β3 loop that promotes HIV TAR RNA binding is labeled. All constructs possess an N-terminal TEV-protease cleavage site (lightning bolt). (b) Representative thermogram of dmU1A binding to hpII RNA. Here and elsewhere the apparent  $K_D$ ,  $n$  (ligand to receptor ratio), and  $c$  value (i.e.,

a unitless parameter that describes the curve shape for quality control) are provided for the titration shown; average values are in Table 2. (c) Representative thermogram of HIV-1 TAR RNA binding to parental TBP6.9 prepared in the absence of selenomethionine. (d) Representative thermogram of hpII RNA dmU1A(F37M/F77M) labeled with selenomethionine. (e) Representative thermogram of HIV-1 TAR RNA binding to TBP6.9(F34M/F37M/F77M) labeled with selenomethionine.

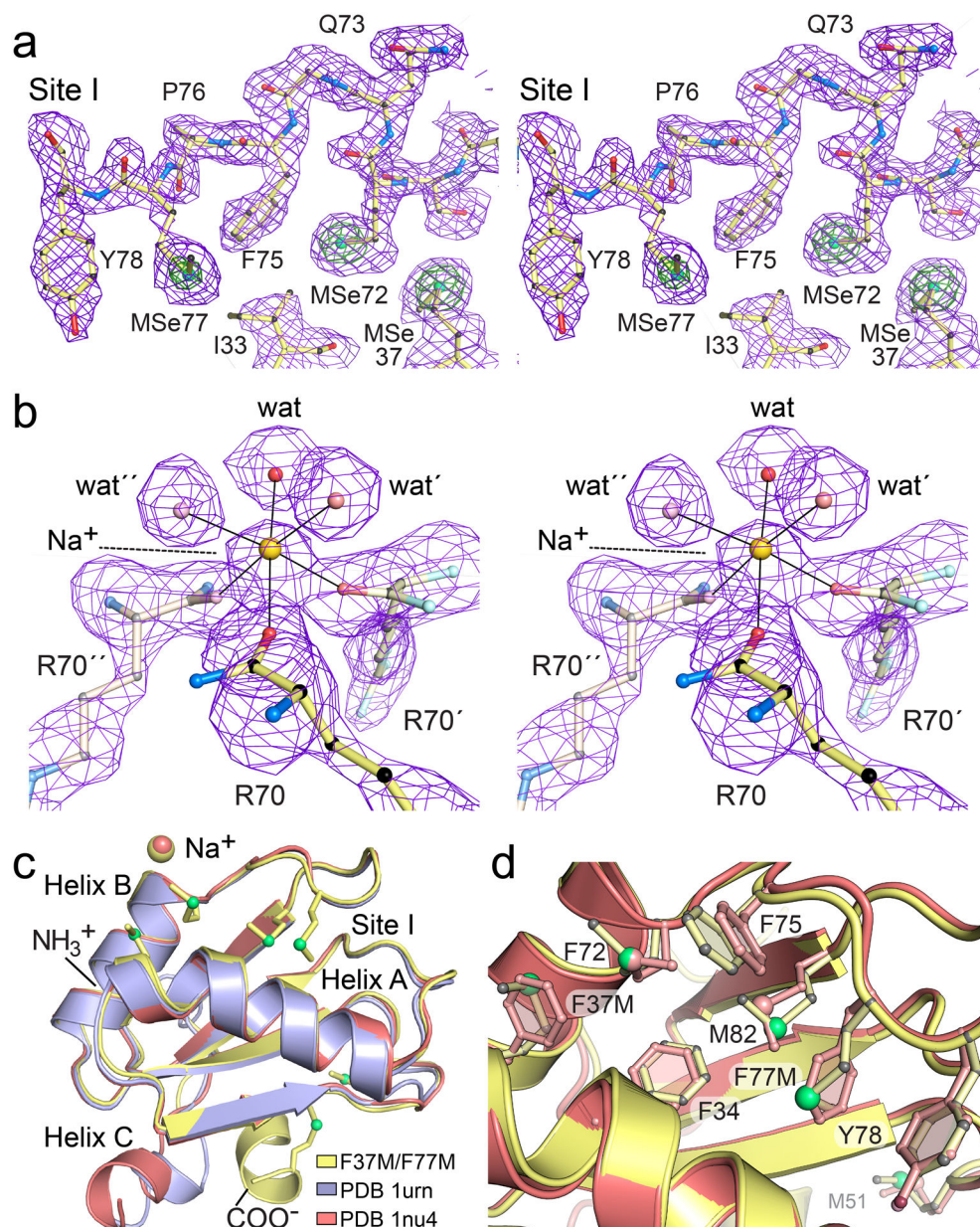


**Figure 4.** Schematic diagram of the TAR core binding motif for TBP and thermograms of TAR-GNRA binding to TBP6.9 and TBP6.9(F34M/F37M/F77M). (a) The TAR internal bulged loop (purple) depicting regions that are non-essential for binding (white) based on the crystal structure and binding experiments [51,52]. The apical loop was substituted with a GAAA tetraloop and a closing C–G pair, which produces optimal thermodynamic stability for this motif [89]. (b) Thermogram of GNRA-TAR RNA binding to the parental TAR TBP6.9 prepared without selenomethionine substitution. Here and elsewhere the average thermodynamic parameters are reported in Table 2. (c) Thermogram of GNRA-TAR RNA binding to TBP6.9(F34M/F37M/F77M) prepared in the presence of selenomethionine.



**Figure 5.**

Representative crystals of dmU1A(F37M/F77M) labeled with selenomethionine. **(a)** Trigonal crystals grown in the absence of RNA from solutions of Na-acetate buffered at neutral pH. Here and elsewhere, crystals were photographed under polarized light. **(b)** Rods of dmU1A(F37M/F77M) bound to the hpII 21-mer RNA. Crystals grew from solutions of polyethylene glycol buffered at neutral pH. **(c)** Cubic crystals of dmU1A(F37M/F77M) in complex with a preQ<sub>1</sub>-II riboswitch 84-mer. Crystals grew from lithium sulfate buffered at pH 8.5.



**Figure 6.** Representative electron density maps of the selenomethionine-labeled dmU1A(F37M/F77M) variant and comparison of dmU1A structures. **(a)** Stereo view of  $2mF_o - DF_c$  electron density (purple) contoured at  $1.0 \sigma$  (purple) and  $5.0 \sigma$  (green) for the site I hydrophobic pocket. Amino acids including F37M and F77M are depicted as ball-and-stick models. **(b)** Stereo view of a simulated-annealing composite omit map depicting a sodium ion on the crystallographic 3-fold axis. The ion is coordinated by three carbonyl oxygens from R70 and three waters (labeled ‘wat’). For clarity, only the R70 amino acid, the  $\text{Na}^+$  ion and an inner sphere water were omitted from the phase calculation; the amide nitrogen of S71 is shown to indicate the peptide bond but was included in the phase calculation. **(c)** Pairwise superpositions of dmU1A (salmon) prepared from sodium malonate [90] and dmU1A from

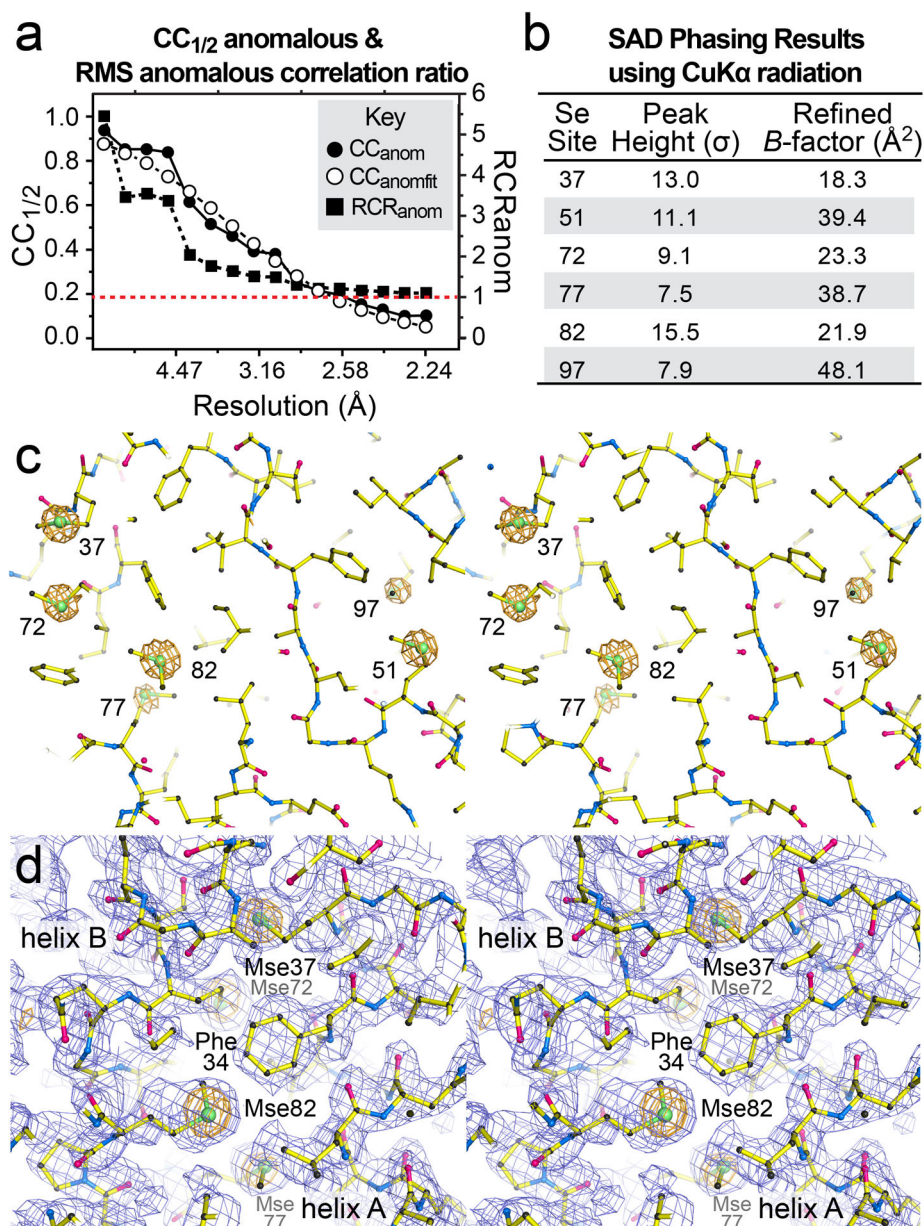
the protein-RNA complex (slate) [56] with the dmU1A(F37M/F77M) mutant of this investigation (yellow) prepared from Na-acetate and labeled with selenomethionine. **(d)** Close-up view of site I from the superimposed dmU1A and dmU1(F37M/F77M) mutant from panel **c**.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 7.** Selenomethionine anomalous signal, substructure determination and representative noise-filtered SAD-phasing electron-density map of dmU1A(F37M/F77M). (a) (*Left ordinate*) Plot of the anomalous difference correlation coefficients  $CC_{anom}$  and  $CC_{anomfit}$  as a function of resolution. (*Right ordinate*) Plot of RMS correlation ratio for anomalous diffraction ( $RCR_{anom}$ ) as a function of resolution. A dashed red line indicates a value of 1.0 for the  $RCR_{anom}$  signal. Parameters are defined as described in Aimless [86]. (b) Results of Se-substructure identification using a hybrid search [76] and values of the refined  $B$ -factors from the corresponding model in Figure 6. The heights of the six selenium anomalous difference Fourier peaks were derived from Phaser EP [72]. (c) Stereo view of the anomalous difference Fourier peaks from panel b, shown in the correct hand and contoured

at  $5\sigma$ ; the final refined model of dmU1A(F37M/F77M) from Figure 6 is shown. **(d)** The noise-filtered SAD phasing map contoured at  $1.0\sigma$  (slate blue) shows recognizable secondary structure features. The refined model from Figure 6 is shown and fits well to the experimental electron density. The anomalous difference Fourier map contoured at  $4\sigma$  is shown to emphasize the locations of the site I Se atoms.



**Table 1.**

Representative Structures in Complex with dmU1A or TBP.

Sample Name	PDB ID	Phasing Method	No. Se Sites	Ref.
SAM-VI riboswitch	6las	SAD <sup>a</sup>	3–4	[36]
NAD <sup>+</sup> -I riboswitch	7d7v	SAD	3 of 4	[37]
HDV ribozyme	1sjf	MAD <sup>b</sup>	4 of 4	[38]
HDV ribozyme	1drz	MAD	4 of 4	[39]
group I intron	1u6b	MAD <sup>c</sup>	4 of 4	[40]
glutamine riboswitch	5ddp	MR <sup>d</sup>	n/a	[41]
glycine riboswitch	3p49	MAD <sup>e</sup>	n/a	[42]
cyclic-di-AMP riboswitch <i>ydaO</i>	4w90	MAD	3 of 4	[43]
c-di-GMP riboswitch	3iwn	MR <sup>d</sup>	n/a	[44]
c-di-GMP riboswitch	3irw	MAD <sup>e</sup>	n/a	[45]
aminoacyl tRNA synthetase ribozyme	3cun	MAD/MR <sup>c</sup>	3 of 4	[46]
tetracycline artificial riboswitch	3egz	MAD	3 of 4	[47]
class I ligase ribozyme	3hhn	MAD <sup>e</sup>	n/a	[48]
pre-cleaved <i>glmS</i> ribozyme	3g8S	MR <sup>d</sup>	n/a	[49]
K-turn-L7Ae complex	4c4w	MR <sup>d</sup>	n/a	[50]
TAR RNA in complex with TAR binding protein 6.7	6cmn	MR <sup>f</sup>	n/a	[51]
TAR RNA in complex with TAR binding protein 6.9	6xh0	MR <sup>g</sup>	n/a	[52]

<sup>a</sup>Single-wavelength anomalous diffraction.<sup>b</sup>Multiwavelength anomalous diffraction.<sup>c</sup>Additional phasing approaches were required beyond the use of the Se-Met atoms in U1A.<sup>d</sup>Molecular replacement (MR) using the dmU1A complex with the hpII loop.<sup>e</sup>MAD phasing was used but Ir(NH<sub>3</sub>)<sub>6</sub> was the basis for phasing.<sup>f</sup>The search model was dmU1A alone.<sup>g</sup>The search model was the TAR-TBP6.7 complex.

**Table 2.**

Isothermal titration calorimetry of wildtype and mutant dmU1A and TBP6.9.

Sample	$K_D$ nM	n	H kcal mol <sup>-1</sup>	-T S kcal mol <sup>-1</sup>	G kcal mol <sup>-1</sup>	$K_{rel}$ <sup>a</sup>
dmU1A <sup>b</sup>	152.0 ± 7.0 <sup>c</sup>	1.0 ± 0.02	-32.0 ± 0.6	22.8 ± 0.6	-9.2 ± 0.03	N/A
dmU1A 7Se-Met (F37M/F77M) <sup>b</sup>	59.7 ± 11.4	0.95 ± 0.05	-19.1 ± 3.0	9.4 ± 2.9	-9.7 ± 0.12	0.39
TBP6.9 (WT) <sup>d</sup>	8.5 ± 0.4	1.0 ± 0.02	-16.0 ± 0.1	5.2 ± 0.01	-10.9 ± 0.1	N/A
TBP6.9 7Se-Met (F34M/F37M/F77M) <sup>d</sup>	13.5 ± 2.3	1.2 ± 0.02	-14.5 ± 0.1	3.9 ± 0.1	-10.6 ± 0.1	1.6
TBP6.9 (WT) <sup>e</sup>	87.8 ± 5.0	0.94 ± 0.04	-7.8 ± 0.4	-1.7 ± 0.5	-9.5 ± 0.04	10.3
TBP6.9 7Se-Met (F34M/F37M/F77M) <sup>e</sup>	69.8 ± 2.9	1.1 ± 0.01	-7.2 ± 0.8	-2.4 ± 0.9	-9.6 ± 0.03	8.2

<sup>a</sup>The ratio of  $K_D$  values of the mutant divided by the matched wildtype (i.e., dmU1A or TBP6.9).<sup>b</sup>The RNA sample in the cell was the hpII 24-mer in Figure 1a.<sup>c</sup>Errors represent standard errors of the mean for multiple measurements.<sup>d</sup>The RNA sample in the cell was the TAR RNA 27-mer in Figure 1b.<sup>e</sup>The RNA sample in the cell was the GNRA-TAR RNA 30-mer depicted in Figure 4a.

**Table 3.**

Diffraction and refinement statistics of dmU1A(F37M/F77M) Se-Met.

<b>Data Collection</b>	
Wavelength (Å)	1.5418
Resolution range (Å)	27.88–2.20 (2.27–2.20)
Space group	<i>P</i> 3 2 1
<i>a</i> = <i>b</i> , <i>c</i> (Å)	64.4, 46.8
$\alpha$ = $\beta$ , $\gamma$ (°)	90.0, 120
Unique reflections	10,961
Multiplicity	10.0 (4.8)
Completeness (%)	100.0 (100.0)
Mean <i>I</i> $\sigma$ ( <i>I</i> )	61.2 (20.2)
$R_{\text{merge}}$ (%) <sup>b</sup>	2.9 (7.6)
$R_{\text{p.i.m.}}$ (%) <sup>c</sup>	0.9 (3.8)
CC <sub>1/2</sub> (%) <sup>d</sup>	1.00 (0.995)
CC <sub>anom</sub> (%) <sup>e</sup>	0.46 (0.10)
RCR <sub>anom</sub> <sup>f</sup>	1.65 (1.11)
<b>Refinement</b>	
No. reflections (all/work/test)	11519/10,961/558
$R_{\text{cryst}}/R_{\text{work}}/R_{\text{free}}$ (%) <sup>g</sup>	16.8/16.6/21.3
No. of atoms:	
protein	776
waters	85
Na/Acetate/ $\beta$ ME	13
Se	6
R.M.S. deviations:	
bonds (Å)	0.003
angles (°)	0.53
ML coordinate error (Å)	0.19
Clashscore <sup>h</sup>	0.0
Molprobrity score <sup>h</sup>	0.50
Ramachandran Plot (%):	
most favored	100
outliers (%)	0
<i>B</i> -factors overall (Å <sup>2</sup> ):	
protein	19.6
waters	24.3
Na/Acetate/ $\beta$ ME	30.8
Se atoms	27.4

<sup>a</sup>Statistics for the highest-resolution shell are shown in parentheses.

<sup>b</sup> $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i - \langle I \rangle| / \sum_{hkl} \sum_i I_i$  where  $I_i$  is an intensity  $I$  for the  $i^{\text{th}}$  measurement of a reflection with indices  $hkl$  and  $\langle I \rangle$  is the weighted mean of all measurements of  $I$ .

<sup>c</sup> $R_{\text{p.i.m.}} = \sum_{hkl} (1/(n-1)) \sum_i |I_i - \langle I \rangle| / \sum_{hkl} \sum_i I_i$  where  $n$  is the number of observations of intensity  $I_i$  [95].

<sup>d</sup> $CC_{1/2}$ , correlation coefficient between intensities of random half-dataset [96].

<sup>e</sup> $CC$  between anomalous pairs measured in the intensity data, as defined in Aimless [97].

<sup>f</sup>RMS Correlation Ratio (RCR) defined in Aimless [97].

<sup>g</sup> $R_{\text{cryst}} = \sum_{hkl} |F_{\text{obs}}(hkl) - F_{\text{calc}}(hkl)| / \sum_{hkl} F_{\text{obs}}(hkl)$  for all reflections,  $R_{\text{work}}$  is for ~95% of reflections, and  $R_{\text{free}}$  is for ~5% of the reflections chosen randomly and excluded from refinement.

<sup>h</sup>Calculated using the program Molprobit [98].