# Supporting Information

# Crystal Structure, Stability and Ago2 Affinity of Phosphoro-dithioate-Modified RNAs

**Pradeep S. Pallan,[a] Xianbin Yang,*[,b] Malgorzata Sierant,[c] Nuwan Abeydeera,[b] Tom Hassell,[d] Carlos Martinez,[d] Barbara Nawrot,[c] and Martin Egli*[,a]**

[a]*Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN 37232, USA.*

[b]*AM Biotechnologies LLC, 12521 Gulf Freeway, Houston, TX 77034, USA.*

[c]*Department of Bioorganic Chemistry, Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Lodz, Poland.*

[d]*Sigma Life Science, 9186 Six Pines, The Woodlands, Texas 77380, USA.*

*E-mail: martin.egli@vanderbilt.edu; Fax: 1-615-322-7122; Tel: 1-615-343-8070, or E-mail: Xianbin.yang@thioaptamer.com Fax: 1-832-476-0294; Tel: 1-832-379-2175.

## Supporting Information Contents

**Synthesis of native and Me-, PS-, PS2- and MePS2-modified siRNAs**

The siRNA and modified RNAs were synthesized on an Expedite 8909 DNA/RNA Synthesizer using commercial 5′-DMT-2′-*O*-TBDMS nucleoside ($A^{Bz}$, $C^{Ac}$, $G^{Ac}$, and U) phosphoramidite monomers as well as 2′-*O*Me-thiophosphoramidites. The average stepwise coupling efficiency of all phosphoramidites including thiophosphoramidites was about 97% as estimated by the DMT-cation assay. After deprotection, all of the modified RNAs were isolated by FPLC according to a previously described protocol for purifying PS2-DNAs.[1,2] The PS2-RNAs were desalted using reverse-phase HPLC to yield the PS2-RNA final products. The representative structures of the MePS2-RNAs were confirmed by ESI-MS. Assembly of siRNA duplexes was performed in ammonium acetate buffer (pH 7.4) by heating the equivalent mixture of RNA oligonucleotides coding for the sense and antisense strands of siRNA at 90°C for 2 min, followed by slow cooling to room temperature (over 2 h). Assembly of the resulting duplexes was confirmed by electrophoresis using a 4% agarose gel.

**$T_m$ measurements**

All absorption measurements were performed in a 1-cm path length cell with a Cintra 4040 spectrophotometer equipped with a Peltier Thermocell (GBC, Dandenong, Australia). Complementary RNA strands were mixed in 10 mM Tris-HCl buffer (pH 7.4) with 100 mM NaCl, and 0.1 mM EDTA at a final concentration of 2 μM (0.8 OD of both mixed strands). The first step of the analysis was an annealing step from 85 to 15°C, with a temperature gradient of 1.5°C/min. Melting profiles were measured with a temperature gradient of 1°C/min, from 15 to 85 °C, with the detector set at 260 nm. Each $T_m$ reported was an average of values from three to six independent experiments.

**Ago2 affinity assays**

The antisense strand siRNA is biotinylated at its 3′-terminus, allowing immobilization of an annealed duplex on to a streptavidin (SA) coated sensor surface. This enables kinetic analysis of siRNA duplex binding to Ago2 on a fortéBIO Octet Red96 instrument set at 30°C. Samples were agitated at 1000 rpm. The annealing method, binding buffer condition, siRNA loading concentration, siRNA loading time, Ago2 titer range, appropriate blocking reagent, blocking time, Ago2 stability over assay duration, and octet baseline correction methods were thoroughly investigated and optimized in the current study. Ago2 in HSCMT was prepared as a dilution series (0, 12.5, 25, 50, 100, and 200 nM) along with HSCMT buffer blanks (unloaded sensor controls). Association was monitored for 300 sec and dissociation was followed for 300 sec. If necessary, the dissociation was stretched to at least 900 sec to verify tight

binding. The data were fit to a 1:1 binding model using fortéBIO Octet data analysis software. Kinetic constants were determined by integration of the experimental data using the differential rate equation $dR/dt = k_{on} \cdot C \cdot (R_{max}-R) - k_{off} \cdot R$ to obtain both the $k_a$ and $k_d$ values ($R$ = observed response, $R_{max}$ = maximum response upon saturation, $C$ = analyte concentration, $k_{on}$ = association rate constant, $k_{off}$ = dissociation rate constant). The ratio between $k_{off}$ and $k_{on}$ corresponds to the reported dissociation constants ($k_{off}/k_{on}=K_D$). The goodness of the fit was judged by the reduced $\chi^2$ and $R^2$ values.

**CD measurements**

CD spectra were recorded on a CD6 dichrograph (Jobin-Yvon, Longjumeau, France) using cells with 0.5 cm path length, 2 nm bandwidth, and 1-2 sec integration time. Each spectrum was smoothed with a 25-point algorithm (included in the manufacturer′s software, version 2.2.1) after averaging of at least three scans. The spectra from 200 nm to 340 nm were recorded at 25°C in the same buffer used for the melting experiments. The concentration of the two complementary RNA oligonucleotides was ca. 2 μM.

**Modeling of PAZ:siRNA complexes**

Coordinates for the crystal structure of a self-complementary RNA with a 3′-dCT overhang bound to the human Ago2 PAZ domain[3] were obtained from the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do; PDB ID 1si2). Using the program UCSF Chimera[4] non-canonical Py:Py and Pu:Pu pairs were replaced by Watson-Crick pairs according to the EphA2 siRNA sequence (**Fig. S1**, **Table 2**), and the 2′-deoxy-C in the 3′-overhang was replaced by T. In addition all water molecules were removed and hydrogen atoms were added. Following rebuilding of the RNA, the native complex was subjected to minimization by steepest descent and conjugate gradient using the Amber ff12SB force field in Chimera. In the complex with the MePS2-modified sense strand, two phosphates adjacent to the 3′-overhang were replaced by phosphorodithioate moieties (taken from the crystal structure of $C_{PS2}$G-RNA). In addition, methyl groups were added to the 2′-oxygen atoms, such that the conformation of the $CH_3$-O2′-C2′-C3′ torsion angle is in the *antiperiplanar* range. The complex was subsequently subjected to refinement as described above for the native PAZ:siRNA complex.

**Crystallization experiments**

The modified RNA dodecamers r(CGC$_{PS2}$GAAUUCGCG) [$C_{PS2}$G-RNA], r(CGCGA$_{MePS2}$AUUAGCG) [$A_{MePS2}$A-RNA] and r(CGCGAAUUA$_{MePS2}$GCG) [$A_{MePS2}$G-RNA] were dissolved in deionized water and the stock concentrations adjusted to ca. 1.2 mM. Crystallization trials were performed with the

hanging drop vapor diffusion technique, using the 24 conditions of the Nucleic Acid Miniscreen (Hampton Research Inc., Aliso Viejo, CA).[5] Droplets (1 μL) of modified RNA dodecamer were mixed with droplets of equal volume of the individual sparse matrix screen solutions and equilibrated against 0.6 mL of a 35% v/v solution of 2-methyl-2,4-pentanediol (MPD) at 18°C. The optimal crystallization conditions for the three oligonucleotides were as follows. $C_{PS2}$G-RNA: condition 20; sodium cacodylate buffer, pH 7.0, 80 mM sodium chloride, 20 mM barium chloride, 12 mM spermine tetrahydrochloride, and 10% v/v MPD. $A_{MePS2}$A-RNA: condition 15; 40 mM sodium cacodylate pH 7.0, 12 mM spermine tetrahydrochloride, 80 mM potassium chloride, and 10% v/v MPD. $A_{MePS2}$G-RNA: condition 8; 40 mM sodium cacodylate, pH 6.0, 80 mM sodium chloride, 12 mM spermine tetrahydrochloride, and 10% v/v MPD.

**X-ray data collection, structure determination and refinement**

Crystals were mounted in nylon loops, flash-frozen in liquid nitrogen without further cryo-protection and stored in liquid nitrogen prior to data collection. Diffraction data were collected on the 21-ID-F or 21-ID-G beam lines of the Life Sciences Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS), located at Argonne National Laboratory (Argonne, IL), using MARCCD 225/300 detectors. The wavelength was 0.98 Å and crystals were kept at 100K during data collection. Diffraction data were integrated, scaled and merged using HKL2000.[6] Selected data collection and refinement statistics are listed in **Table S1**. The structure was determined by the molecular replacement method with the program MOLREP[7] in the CCP4 suite of crystallographic software,[8] using the native RNA dodecamer as the search model (PDB ID 2Q1R[9]; http://www.rcsb.org). Following initial positional and isotropic temperature factor refinements with the program REFMAC,[10] the R-work dropped to the mid 20s and R-free to the upper 20s. After further refinement in REFMAC, refinements were continued with SHELX.[11] The phosphorodithioate and 2'-*O*-methyl moieties (the asymmetric unit contains a single strand) were built into Fourier ($2F_o$-$F_c$) sum and ($F_o$-$F_c$) difference electron density maps that were visualized with COOT[12] and the refinement was continued after adaptation of the dictionary files. Further refinements were carried out and after each refinement the nucleotides and water molecules were checked for possible alternate conformations/positions and corrections were made where necessary. A final refinement was carried out with all nucleic acid atoms and water molecules being treated with anisotropic B-factors and these refinement parameters are summarized in **Table S1**. All figures were generated with the program UCSF Chimera.[4]

**Table 1.** Selected crystal data and refinement parameters for PS2- and MePS2-modified RNAs

| **Structure** | $C_{PS2}$G-RNA | $A_{MePS2}$A-RNA | $A_{MePS2}$G-RNA |
|---|---|---|---|
| **Crystal data** | | | |
| Space group | | Monoclinic *C*2 | |
| Unit cell constants | | | |
| *a* [Å] | 40.93 | 41.18 | 40.91 |
| *b* [Å] | 35.02 | 35.05 | 34.84 |
| *c* [Å] | 31.87 | 31.85 | 32.27 |
| $\beta$ [deg] | 128.8 | 128.7 | 127.5 |
| No. of strands per asym. unit | 1 | 1 | 1 |
| No. of unique reflections | 11,133 | 11,706 | 12,267 |
| Resolution [Å] (last shell) | 1.19 (1.23-1.19) | 1.18 (1.22-1.18) | 1.13 (1.17-1.13) |
| Completeness [%] (last shell) | 97.7 (88.5) | 99.9 (100) | 95.5 (90.7) |
| R-merge [%] (last shell) | 5.0 (4.2) | 8.2 (21.6) | 5.3 (25.5) |
| **Refinement** | | | |
| R-work / R-free | 0.139 /0.183 | 0.160/0.201 | 0.181/0.208 |
| No. of RNA atoms | 256 | 258 | 258 |
| No. of water molecules | 97 | 70 | 52 |
| No. of metal ions | 2 $Sr^{2+}$ [a] | - | - |
| R.m.s. deviations: | | | |
| Bond lengths [Å] | 0.012 | 0.011 | 0.012 |
| Bond angles [Å] [b] | 0.03 | 0.036 | 0.040 |
| **Data deposition** | | | |
| PDB ID | 4RBY | 4RBZ | 4RC0 |

[a] Refinement suggested $Sr^{2+}$ ions with occupancy 0.3. This is in contrast to the crystallization conditions.
[b] 1…3 distance based on SHELX refinement.

Sense 5'-UGACAUGCCGAUCUACAU$_{MePS2}$G$_{MePS2}$TT-3'
Antisense 3'-TTACUGUACGGCUAGAUGUA C-5'

**Fig. S1.** Sequence of the EphA2 siRNA with two sense-strand MePS2 modifications.

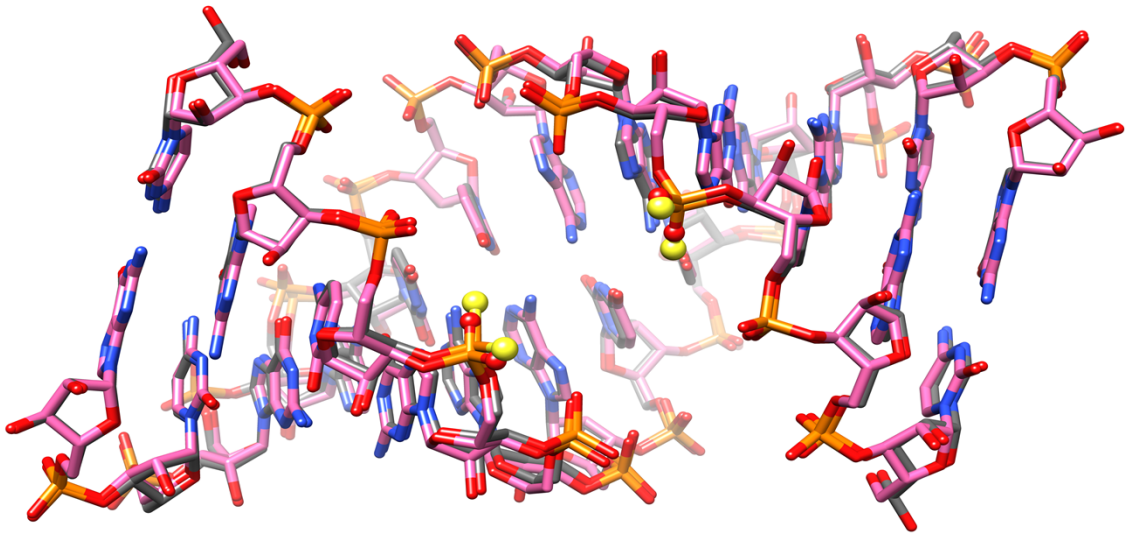

**Fig. S2.** Superimposition of the native (pink carbon atoms; PDB ID 2q1r) and $C_{PS2}G$ RNA dodecamer crystal structures (gray carbon atoms).

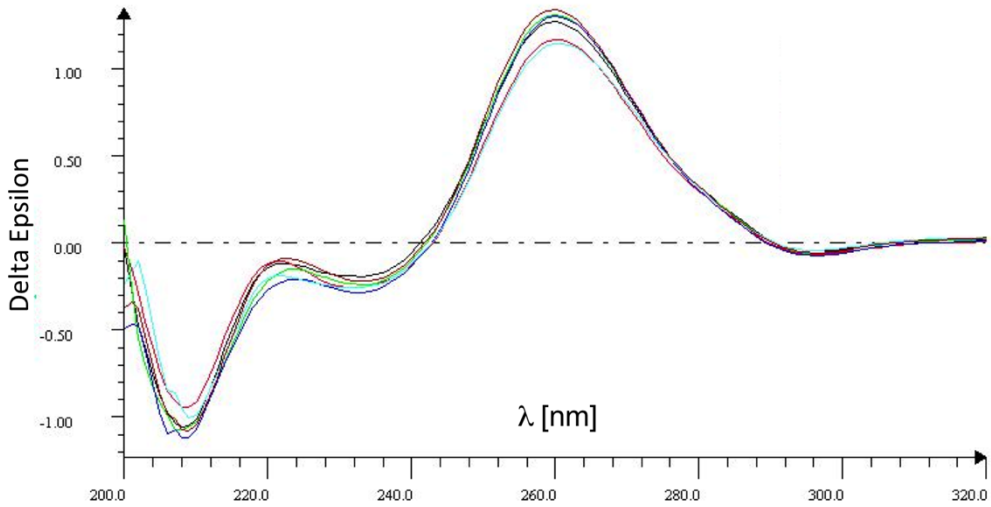


**Fig. S3.** CD spectra of the native RNA duplex siRNA-1 (for sequences see **Table 1**; purple), and MePS2-modified RNA duplexes MePS2-1 (black), MePS2-3 (red), MePS2-4 (green), MePS2-9 (blue), and MePS2-11 (cyan).

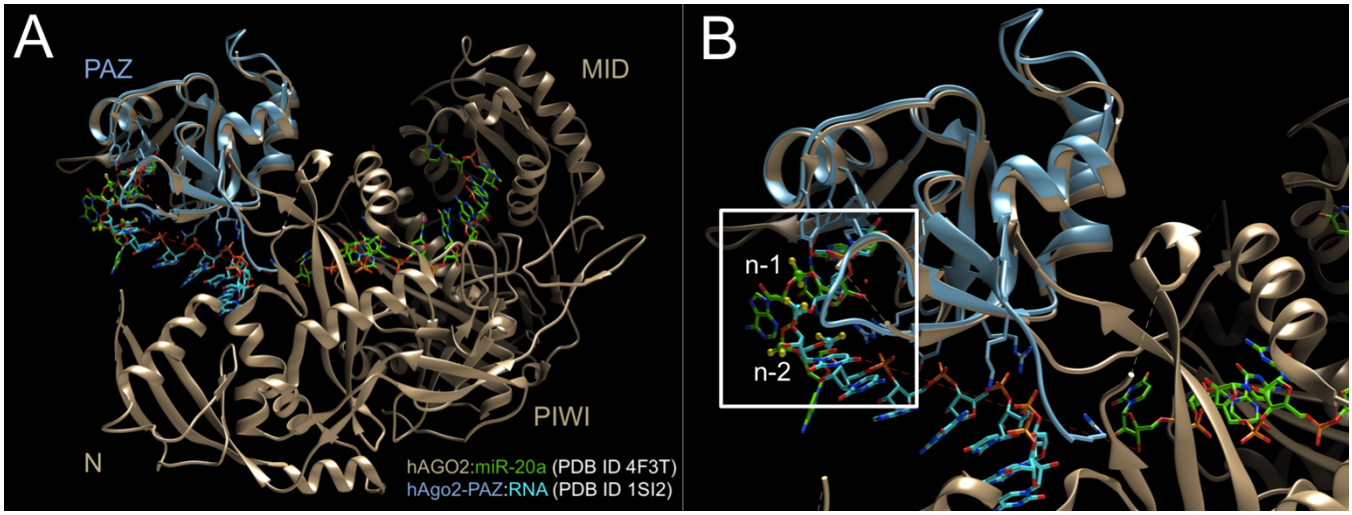


**Fig. S4.** (A) Overall and (B) close-up views of superimposed crystal structures of the complexes between the human AGO2 PAZ domain and a self-complementary 9mer (cyan) with a 3′-dCT overhang[3] (PDB ID 1si2), and between human AGO2 and miR-20a RNA (green)[13] (PDB ID 4f3t). The images illustrate that the interactions between the 3′-end of siRNAs and the PAZ domain can be analyzed in an isolated fashion. The region of the structure analyzed in the context of the MePS2 modifications at the n-1 and n-2 phosphates and shown in **Fig. 4** in the main paper is boxed in panel B and phosphates are labeled. The mid-section of the green RNA strand (dotted red line) is not visible in electron density maps of the Ago2:miR-20a complex.

**References**


1. Yang, X.; Sierant, M.; Janicka, M.; Peczek, L.; Martinez, C.; Hassell, T.; Li, N.; Li, X.; Wang. T.; Nawrot, B. *ACS Chem. Biol*. **2012**, *7*, 1214-1220.
2. Wu, S. Y.; Yang, X.; Gharpure, K.; Hatakeyama, H.; Egli, M.; McGuire, M.; Nagaraja, A. S.; Miyake, T.; Rupaimoole, R.; Pecot, C. V.; Taylor, M.; Pradeep, S.; Sierant, M.; Rodriguez-Aguayo, C.; Choi, H. J.; Previs, R. A.; Armaiz-Pena, G. N.; Huang, L.; Martinez, C.; Hassell, T.; Ivan, C.; Sehgal, V.; Singhania, R.; Han, H.-D.; Su, C.; Kim, J.-H.; Dalton, H.; Kowali, C.; Keyomarsi, K.; McMillan, N. A. J.; Overwijk, W. W.; Liu, J.; Lee, J.-S.; Baggerly, K.; Lopez-Berestein, G.; Ram, P.; Nawrot, B.; Sood, A. K. *Nat. Comm.* **2014**, *5*, 3459.
3. Ma, J.-B.; Ye, K.; Patel, D. J. *Nature*, **2004**, *429*, 318-322.
4. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comp. Chem.* **2004**, *25*, 1605-1612.
5. Berger, I.; Kang, C. H.; Sinha, N.; Wolters, M.; Rich, A. *Acta Cryst. D* **1996**, *52*, 465-468.
6. Otwinowski, Z.; Minor, W. *Meth. Enzymol.* **1997**, *276*, 307-326.
7. Vagin, A.; Teplyakov, A. *J. Appl. Crystallogr.* **1997**, *30*, 1022-1025.
8. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Cryst. D*, **1994**, *50*, 760-763.
9. Li, F.; Pallan, P. S.; Maier, M. A.; Rajeev, K. G.; Mathieu, S. L.; Kreutz, C.; Fan, Y.; Sanghvi, J.; Micura, R.; Rozners, E.; Manoharan, M.; Egli, M. *Nucleic Acids Res.* **2007**, *35,* 6424-6438.
10. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Cryst. D* **1997**, *53*, 240-255.
11. Sheldrick, G. M. Methods Enzymol. **1997,** 276, 628-641.
12. Emsley, P.; Cowtan, K. *Acta Cryst. D* **2004**, *60*, 2126-2132.
13. Elkayam, E.; Kuhn, C. D.; Tocilj, A.; Haase, A. D.; Greene, E. M.; Hannon, G. J.; Joshua-Tor, L. *Cell* **2012**, *150*, 100-110.


**Acknowledgements**


The project upon which this publication is based was performed pursuant to an agreement with AM Biotechnologies, LLC which in turn was supported by a Grant Number 2R44 GM086937-03 from the National Institutes of Health. Use of the Advanced Photon Source (APS) was supported by the U. S. Dept. of Energy, Office of Science, Office of Basic Energy Sciences, Contract No. DE-AC02-06CH11357.