### Title

A New Variant of Emissive RNA Alphabets




### Journal

Chemistry - A European Journal, 28(13)

### ISSN

0947-6539



### Authors

Ludford, Paul T
Yang, Shenghua
Bucardo, Marcela S
et al.



### Publication Date

2022-03-01

### DOI

10.1002/chem.202104472

Peer reviewed

HHS Public Access

Author manuscript

*Chemistry*. Author manuscript; available in PMC 2023 March 01.

Published in final edited form as:

*Chemistry*. 2022 March 01; 28(13): e202104472. doi:10.1002/chem.202104472.

# A New Variant of Emissive RNA Alphabets

**Paul Ludford 3rd**[1], **Shenghua Yang**[1], **Marcela S Bucardo**[1], **Yitzhak Tor**[2]

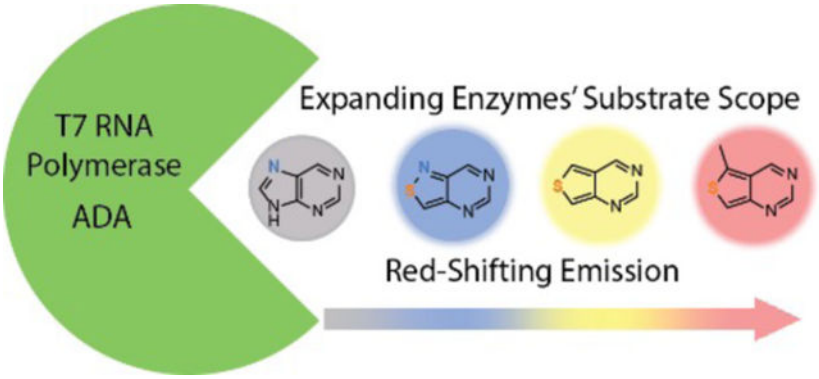
[1]University of California San Diego, Chemistry and Biochemistry, UNITED STATES.

[2]University of California, San Diego, Department of Chemistry and Biochemistry, 9500 Gilman Drive, MC 0358, 92093-0358, La Jolla, UNITED STATES.

## Abstract

A new fluorescent ribonucleoside alphabet ($^{mth}$N) consisting of pyrimidines and purines analogues, all derived from methylthieno[3,4-*d*]pyrimidine as the heterocyclic core, is described. Large bathochromic shifts and high microenvironmental susceptibility of their emission relative to previous alphabets derived from thieno[3,4-*d*]pyrimidine ($^{th}$N) and isothiazole[4,3-*d*]pyrimidine ($^{tz}$N) scaffolds are observed. Subjecting the purine analogues to adenosine deaminase, guanine deaminase and T7 RNA polymerase indicate that, while varying, all but one enzyme tolerate the corresponding $^{mth}$N/$^{mth}$NTP substrates. The robust emission quantum yields, high photophysical responsiveness and enzymatic accommodation suggest that the $^{mth}$N alphabet is a biophysically viable tool and can be used to probe the tolerance of nucleoside/tide-processing enzymes to structural perturbations of their substrates.

## Graphical Abstract




The substrate scope of T7 RNA Polymerase and adenosine deaminase (ADA) are expanded by the introduction of a new ribonucleoside alphabet based on a methyltheino[3,4-*d*]pyrimidine core. The majority of the new analogues display significantly red shifted emission relative to the corresponding thieno[3,4-*d*]pyrimidine and isothiazolo[4,3-*d*]pyrimidine based analogues.

ytor@ucsd.edu .

Supporting information for this article is given via a link at the end of the document.

**Keywords**

Emissive Nucleosides; Fluorescent Probes; Fluorescence; Ribonucleosides; RNA

Fluorescent ribonucleosides, which have been developed to address the lack of practical emission levels in native ribonucleosides, have been used to great success as probes for biophysical analyses, RNA structural and recognition analyses, and discovery assays.[1–14] Some prized characteristics of many synthetic fluorescent ribonucleosides include red shifted emission and high brightness as well as sensitivity of their photophysical parameters to the microenvironment, endowing them with "reporting" traits. To maintain structural similarity, or isomorphicity,[1,15] researchers have attempted to minimize changes to the Watson-Crick pairing face and, in more recent years, even the Hoogsteen pairing face.[16–26] Such efforts have often met with undesired consequences, such as diminished emission quantum yield.[22,23] This may imply that regressing structural similarity and deviating away from the canonical nucleobase structures can endow such molecules with desirable photophysical features, but will likely consequently compromise their utility. As many analogues are ultimately intended for application in biochemical/biological contexts, a better understanding of the tolerance of enzymes to systematically introduced modifications would be helpful for fine-tuning future probes and their applications.

Our lab has previously developed two isomorphic and "isofunctional"[27] ribonucleoside alphabets based on thieno[3,4-*d*]pyrimidine and isothiazolo[4,3-*d*]pyrimidine cores.[22,23] The reinstallation of the "N7" in the isothiazolo-based alphabet resulted in improved reception by enzymes, best exemplified by the reaction of the adenosine analogues with adenosine deaminase (ADA),[24,25] but generally resulted in hypsochromic shifts and diminished emission quantum yields.[22,23] We hypothesized that the introduction of a nonpolar moiety at the equivalent position, such as a methyl group (Figure 1), might offset such adverse photophysical consequence, albeit at the conceivable cost of reduced isomorphicity. This, in turn, would allow us to systematically probe progressive structural alterations on both the photophysical features and enzymes tolerance for such modifications.

Herein, we set out to synthesize and scrutinize a ribonucleoside analogue alphabet based on a methylthieno[3,4-*d*]pyrimidine core (Figure 1), with the following goals: (a) analyze the photophysical characteristics of the new ribonucleoside alphabet, (b) compare its features to previously developed emissive alphabets, and (c) assess the tolerance level of appropriate enzymes for such Hoogsteen face perturbation. We start by establishing a synthetic route to a key precursor and proceed to synthesize analogues of U, C, A, and G. Next, we photophysically characterize each ribonucleoside by analyzing the photophysical parameters, including extinction coefficients, emission quantum yields, Stokes shifts, and evaluating their sensitivity to changes in polarity. Finally, we explore the reactions between various enzymes, particularly deaminases and RNA polymerase, and compare the performance to that of previously developed emissive analogues. Our observations suggest that the new $^{mth}$N emissive alphabet displays robust photophysical features and, despite the perceived structural perturbation introduced into the Hoogsteen face, the new nucleosides can be used to probe several RNA related processes.

Synthesis of the methylthieno[3,4-*d*]pyrimidine-based analogues began with construction of **2**, a key intermediate. A conjugate addition between methyl thioglycolate and methyl crotonate was followed by a Claisen condensation of the resulting thioether to yield **1** (Scheme 1).[28] The cyclic sulfide **1** was reacted with hydroxylamine hydrochloride in methanol to yield an oxime (Scheme 1).[28] The oxime was then stirred in 2M HCl for 24 h to yield the key intermediate **2** (Scheme 1).[28] Synthesis of $^{mth}$**G**, $^{mth}$**A**, $^{mth}$**U**, and $^{mth}$**C** in parallel routes all began with compound **2** (Figure 1, Schemes 1, S1–4).

We will briefly describe the synthesis of one analogue, $^{mth}$**G**,[29] which began by treating **2** with chloroformamidine (Scheme 1). The resulting nucleobase $^{mth}$**G**$_{n}$ was then protected as a dimethylformamidine to yield **3** (Scheme 1, S1). Compound **3** was regioselectively *C*-glycosylated producing **4** as a single isomer, which was deprotected in methanolic ammonia (65$^{O}$C 16 h) to yield $^{mth}$**G** (Scheme 1, S1). To prepare $^{mth}$**GTP**, the corresponding triphosphate, $^{mth}$**G** was treated with phosphoryl chloride in trimethyl phosphate, followed by bis(tetrabutylammonium) pyrophosphate and tributylamine (Schemes 1, S2). The remaining ribonucleoside analogues syntheses are reported in the supporting information.[29]

Crystal structures of $^{mth}$**U** and $^{mth}$**I** confirmed their beta configuration, thus securing the anomeric configuration of $^{mth}$**U**, $^{mth}$**C**, and $^{mth}$**A** (Figures 2, S1, S2). In the solid state, $^{mth}$**U** and $^{mth}$**I** were found to be in the anti-orientation about the glycosidic linkage similar to uridine and inosine (Figures 2, S1, S2).[30,31] As expected of the beta anomeric configuration, 2D NOESY $^{1}$H NMR revealed through space correlations between the H1' and 2'-OH for $^{mth}$**U**, $^{mth}$**C**, $^{mth}$**G**, $^{mth}$**A**, and $^{mth}$**I** (Figures S23, S35, S39, S41 and S43).[32]. To assess the impact of the structural modification on solution conformation, the North vs. South sugar puckering pseudorotational equilibrium was assessed using the (HH) $^{3}J$-coupling constants between H1' and H2' and H3' and H4' (Table S3). As expected, the modified pyrimidines $^{mth}$**U** and $^{mth}$**C** suffer larger impact compared to their native counterparts (%S 56 and 54, respectively) but the values obtained for the purine analogs $^{mth}$**A** and $^{mth}$**G** (63%, 71% respectively), were very similar to the averaged value (60%) reported for the native purines.[33] While recognizing the potential imperfections in correlating values for *C*- and *N*-nucleosides, the solid state and solution analyses suggest the purine surrogates can access native-like conformations.

Absorption and emission spectra of $^{mth}$**A**, $^{mth}$**G**, $^{mth}$**U**, $^{mth}$**C**, and $^{mth}$**I** were recorded in water, dioxane, and mixtures of the two and the extinction coefficients, emission quantum yields, Stokes Shifts, and sensitivities toward microenvironmental polarity were obtained (Table 1, S4, Figure 3, S3).[34] All analogues except $^{mth}$**G** displayed significantly bathochromic-shifted emission maxima relative to our previously developed ribonucleoside alphabets and their native counterparts ranging from 414 ($^{mth}$**I**) to 467 ($^{mth}$**A**) nm in water and 382 ($^{mth}$**I**) to 445 ($^{mth}$**C**) nm in dioxane.[22,23] In water, $^{mth}$**A**, $^{mth}$**U**, $^{mth}$**C**, and $^{mth}$**I** displayed bathochromic shifts in emission maxima of 47 and 57 nm, 18 and 35 nm, 26 and 44 nm, and 23 and 37 nm relative to their thieno- and isothiazolo-based counterparts, respectively.[35,36]

Absorption spectra of $^{mth}$**A**, $^{mth}$**G**, $^{mth}$**U**, $^{mth}$**C**, and $^{mth}$**I** were recorded in buffered solutions with pH values ranging from 1 to 13. Fitting a Boltzmann sigmoidal curve to a plot

of the absorption maxima versus pH yielded inflection points corresponding to p$K_a$ values for each nucleoside (Figure S4, Tables 1 and S4). Deprotonation of the N3 of **$^{mth}$A** resulted in a slight hypsochromic shift in its absorption maximum from pH 4 to 8 and yielded a p$K_a$ value of 6.5, which was above the reported values for adenosine (p$K_a$ = 3.6–4.2).[23] **$^{mth}$G** displayed two bathochromic shifts in absorption maximum between pH 3 and 7 and pH 9 and 12 corresponding to deprotonation of the N3 and N1, respectively, and from which two p$K_a$ values of 5.0 and 10.4 were determined, respectively. These were slightly greater than the reported values of guanosine (p$K_a$ = 3.2–3.3, 9.2–9.6).[23] The deprotonation of **$^{mth}$U** was characterized by a red shift of its absorption maximum between pH 10 and 12 and p$K_a$ of 10.8 which was slightly higher than that of uridine (p$K_a$ = 9.20–9.25).[23] **$^{mth}$C** was characterized by a blue shift in its absorption maximum upon deprotonation of the N1 between pH 2 and 6. Further examination yielded a p$K_a$ of 3.9 which was near the value reported for cytidine (p$K_a$ = 4.2).[37]

The emission quantum yields of all methylthieno-based analogues in water are greater than the isothiazolo-based counterparts (Table S5). The quantum yields of all methylthieno-based analogues in water except **$^{mth}$A**, which is similar to **$^{th}$A**, are lower than their thieno-based counterparts (Table S5). The quantum yields of all methylthieno-based analogues in dioxane are higher than both the isothiazolo- and thieno-based counterparts, except for **$^{mth}$C** which was slightly lower than **$^{tz}$C** (Table S5). **$^{mth}$A**, **$^{mth}$G**, **$^{mth}$U**, **$^{mth}$C**, and **$^{mth}$I** displayed modest relative quantum yields of 0.21, 0.42, 0.30, 0.24, and 0.56 in water, respectively, and 0.23, 0.61, 0.13, 0.03, and 0.11 in dioxane, respectively (Figure 3a, S3). Interestingly, the emission intensity for **$^{mth}$A**, **$^{mth}$U**, and **$^{mth}$C** was highest in 60 v/v% dioxane in water (Figure 3c, S3). The absorption spectra of **$^{mth}$G**, **$^{mth}$A**, and **$^{mth}$C** in dioxane were bathochromically shifted relative to their spectra in water (Figure 3a–c).[38] The absorption spectra of **$^{mth}$U** in dioxane was hypsochromically shifted relative to its absorption spectra in water.[38] The emission spectra of all analogues in water were bathochromically shifted compared to their emission spectra in dioxane (Table 1, S4).[38] The spectral shifts of the analogues and the solvent dependency of their maxima suggest an enhanced charge transfer character in their excited states.

All methylthieno-based analogues, except **$^{mth}$A**, which was similar to its thieno-based counterpart, showed higher sensitivity to microenvironmental polarity than the two previous emissive RNA alphabets (quantified by the steepness of the slope of the linear fit of Stokes Shift versus Reichard's $E_T(30)$).[22,23] The isothiazolo[4,3-*d*]pyrimidine based alphabet, in particular, appears much less sensitive to environmental polarity. For example, **$^{mth}$U**, **$^{mth}$C**, and **$^{mth}$A** showed sensitivities to polarity which were 2–3 fold higher than that of **$^{tz}$U**, **$^{tz}$C**, and **$^{tz}$A**, respectively. We tentatively attribute this distinct photophysical characteristic of the new **$^{mth}$N** alphabet to the impact of the hydrophobic methyl group on the susceptibility of the ground and excited states to solvation and its dynamics.[39]

To preliminarily demonstrate the utility of the new analogues, we elected to monitor the reaction of adenosine deaminase with **$^{mth}$A** and compare it to the native substrate and to previously reported emissive A analogues. ADA-mediated deamination of A, **$^{tz}$A**, **$^{th}$A**, and **$^{mth}$A** was monitored over time via absorption at 260, 340, 340, and 353 nm, respectively (Table 2, S8, Figure 4). Additionally, the enzymatic reactions of **$^{tz}$A**, **$^{th}$A**, and **$^{mth}$A** were

monitored over time by emission at 410, 391, and 420 nm, respectively, (upon excitation at 322, 318, and 332 nm, respectively) (Tables 2, S8, Figure 4). The resulting photophysical data were plotted against time and fitted to first order curves to extract the $k_{app}$ and $t_{1/2}$ values (Tables 2, S8). Calculated from absorption spectroscopy data, $^{tz}$**A** had the shortest $t_{1/2}$ at 32 ± 2 seconds followed by adenosine at 57 ± 1, $^{th}$**A** at 980 ± 60, and $^{mth}$**A** at 1.1 ± 0.2×10$^4$ seconds.[40] Calculated $t_{1/2}$ values from changes in emission spectra for $^{tz}$**A**, $^{th}$**A**, and $^{mth}$**A** were 31 ± 2, 960 ± 20, and 1.1 ± 0.1×10$^4$ sec, respectively, matching the absorption-monitored kinetics well. Not unexpectedly, longer reaction times were observed for $^{mth}$**A** as a bulkier moiety is installed at the "N7" position.[40] Nonetheless, while perhaps surprising, $^{mth}$**A** is recognized as a substrate by ADA making it a useful tool for potential assay development.[40]

We also tested $^{mth}$**G$_n$** and $^{mth}$**C** with guanine deaminase (GDA) and cytidine deaminase (CDA), respectively (Figure S5, Table S9).[41] The former did not react with GDA but the latter reacted faster than $^{tz}$**C** and cytidine with CDA. Docking simulations of $^{mth}$**G$_n$** in the binding pocket of GDA compared to previously reported simulations of G, $^{tz}$**G$_n$** and $^{th}$**G$_n$** corroborated these findings (see also Table S10, Figures S6–S9).[42] Docking simulations of $^{mth}$**C** and previously reported C analogues within the binding pocket of CDA suggest a void in the active site pocket, where such structural alterations of the native substrate can be accommodated (Figures S10–S13).[43]

To further examine the biochemical compatibility of the new analogues, $^{mth}$**GTP** was used for T7 RNA polymerase-mediated in vitro transcription reactions.[45,46] To compare to the native nucleosides and the previously reported emissive ribonucleoside alphabets, GTP, $^{tz}$**GTP**, and $^{th}$**GTP** were also used for in vitro transcription reactions generating the same representative RNA transcripts.[45] The resulting strands from in vitro transcription reactions were separated by gel electrophoresis and compared under UV light (Figure 5, Lanes 1–4). All reactions with emissive GTP analogues showed fluorescent bands upon 365 nm irradiation, indicating they were successfully incorporated. Truncated products, which are common for in vitro transcription of such short strands,[45,46] were also observed in all lanes and emitted under similar excitation. The yields of the extracted target strands were estimated using their absorbance at 260 nm and were found to be in ~1–2×10$^2$-fold excess of the corresponding template DNA duplex.[47] The yields of target strands of $^{tz}$**GTP**, $^{th}$**GTP**, $^{mth}$**GTP** strand reactions relative to reactions with native GTP were 1.1 ± 0.1, 0.9 ± 0.1, and 0.9 ± 0.1, respectively. MALDI TOF MS analysis of the extracted bands confirmed the desired strands were obtained.[48] This indicated T7 RNA polymerase can initiate transcription and elongate transcripts with $^{mth}$**GTP**, suggesting the Me substituent at the Hoogsteen face was well tolerated by this RNA polymerase.

We have previously reported that T7 RNA polymerase can also initiate transcription with $^{th}$**G** and elongates the nascent transcript strand using native NTPs, when the former is added to transcription reactions in large excess over the canonical triphosphates (5–8 × NTP concentrations).[49] We thus sought to test whether $^{mth}$**G** would also be tolerated by T7 RNA Polymerase in transcription initiation. To compare to the native nucleoside and across all emissive ribonucleoside alphabets, G, $^{tz}$**G**, $^{th}$**G**, and $^{mth}$**G** were added in high concentration (10 mM) to in vitro transcription reactions of the same representative DNA

template as above (Figure 5). The transcripts obtained were separated by gel electrophoresis and compared under UV illumination as above (Figure 5). Similar to the reactions with the emissive nucleotide analogues, all lanes containing reactions with the corresponding emissive G nucleoside analogues showed fluorescent bands under illumination at 365 nm, indicating the analogues were successfully incorporated. Fluorescent truncated products were also observed in all lanes, as expected for **$^{mth}$G**-initiated transcription. The yields of the extracted target strands were estimated using absorbance at 260 nm and were found to be in ~50–60-fold excess of the corresponding template DNA duplex.[47] The yields of target strands of G-, **$^{tz}$G**-, **$^{th}$G**-, and **$^{mth}$G**-initiated strand reactions relative to GTP strand reactions were 0.2 ± 0.1, 0.2 ± 0.1, 0.3 ± 0.1, and 0.3 ± 0.1, respectively. MALDI TOF MS analyses of the extracted bands confirmed the desired strands were obtained.[48] This indicated T7 RNA polymerase can initiate transcription with **$^{mth}$G** corroborating our findings from the **$^{mth}$GTP** reaction that the Me substituent is tolerated by this RNA polymerase.

In summary, we introduce a third family of emissive ribonucleoside analogues, based on a methylthieno[3,4-*d*]pyrimidine scaffold. This structural modification provides an additional opportunity for probing the tolerance of enzymes to structural perturbations centered at the imidazole core of the purine scaffold.[9,50,51] Reactions of the corresponding substrate analogue with ADA and GDA indicate widely varying tolerance for this methyl modification. T7 RNA polymerase was found to initiate transcription of RNA strands with both **$^{mth}$GTP** and **$^{mth}$G** and efficiently elongate the nascent transcripts with **$^{mth}$GTP**. This constitutes two separate methods for the generation of custom emissive RNA strands either via initiation and elongation by the former or via initiation by the latter. These results suggest that despite the perceived structural perturbation introduced into the nucleosides' Hoogsteen face, the new emissive methylthieno[3,4-*d*]pyrmidine alphabet can be used to advance the biophysical probing of certain RNA related processes. In particular, the guanosine analogue may prove useful for developing an emissive and charge-neutral 5'-cap mimic. Additionally, due to their relatively high emission quantum yields and responsiveness, these nucleosides may find utility in developing fluorescence-based activity and inhibition assays for metabolic enzymes as well as for advancing emissive analogs of secondary messengers.[52,53]

## Crystallographic Details

Deposition Numbers 2074931 (for **$^{mth}$U**) and 2075448 (for **$^{mth}$I**) contain the supplementary crystallographic data for this paper. These data are provided free of charge by the joint Cambridge Crystallographic Data Centre and Fachinformationszentrum Karlsruhe Access Structures service.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank the National Institutes of Health for generous support (through grant R35 GM139407). We thank the Chemistry & Biochemistry MS Facility. We thank M. Gembicky and J. Bailey at the UCSD X-ray Crystallography Facility for determining the crystal structures. We thank A. Mrse and the UCSD NMR Facility for assistance.

## References

[1]. Sinkeldam RW, Greco NJ, Tor Y, Chem. Rev 2010, 110, 2579–2619. [PubMed: 20205430]

[2]. Wilhelmsson M, Q. Rev. Biophys 2010, 43, 159–182. [PubMed: 20478079]

[3]. Hawkins ME, Cell Biochem. Biophys 2001, 34, 257–281. [PubMed: 11898867]

[4]. Wilson JN, Kool ET, Org. Biomol. Chem 2006, 4, 4265–4274. [PubMed: 17102869]

[5]. Okamoto A, Saito Y, Saito I, J. Photochem. Photobiol., C 2005, 6, 108–122.

[6]. Dodd DW, Hudson RHE, Mini-Rev. Org. Chem 2009, 6, 378–391.

[7]. Kimoto M, Cox III RS, Hirao I, Expert Rev. Mol. Diagn 2011, 11, 321–331. [PubMed: 21463241]

[8]. Rist MJ, Marino JP, Curr. Org. Chem 2002, 6, 775–793.

[9]. Wierzchowski J, Antosiewicz JM, Shugar D, Mol. BioSyst 2014, 10, 2756–2774. [PubMed: 25124808]

[10]. Dziuba D, Didier P, Ciaco S, Barth A, Seidel CAM, Mély Y, Chem. Soc. Rev 2021, 50, 7062–7107. [PubMed: 33956014]

[11]. Saito Y, Hudson RHE, Journal of Photochemistry and Photobiology C: Photochemistry Reviews 2018, 36, 48–73.

[12]. Xu W, Chan KM, Kool ET, Nat. Chem 2017, 9, 1043–1055. [PubMed: 29064490]

[13]. Dai N, Kool ET, Chem. Soc. Rev 2011, 40, 5756–5770. [PubMed: 21290032]

[14]. Klöcker N, Weissenboeck FP, Rentmeister A, Chem. Soc. Rev 2020, 49, 8749–8773. [PubMed: 33084688]

[15]. Greco NJ, Tor Y, J. Am. Chem. Soc 2005, 127, 10784–10785. [PubMed: 16076156]

[16]. Bood M, Sarangamath S, Wranne MS, Grotli M, Wilhelmsson LM, Beilstein J. Org. Chem 2018, 14, 114–129. [PubMed: 29441135]

[17]. Gaied BN, Glasser N, Ramalanjaona N, Beltz H, Wolff P, Marquet R, Burger A, Mély Y, Nucleic Acids Res. 2005, 33, 1031–1039. [PubMed: 15718302]

[18]. Nadler A, Strohmeier J, Diederichsen U, Angew. Chem., Int. Ed 2011, 50, 5392–5396.

[19]. Dierckx A, Miannay F-A, Gaied NB, Preus S, Björck M, Brown T, Wilhemlmsson LM, Chem. Eur. J 2012, 18, 5987–5997. [PubMed: 22437923]

[20]. Dumat B, Bood M, Wranne MS, Lawson CP, Foller Larsen A, Preus S, Streling J, Gradén H, Wellner E, Grøtli M, Wilhelmsson LM, Chem. Eur. J 2015, 21, 4039–4048. [PubMed: 25641628]

[21]. Hopkins PA, Sinkeldam RW, Tor Y, Org. Lett 2014, 6, 25–64.

[22]. Shin D, Sinkeldam RW, Tor Y, J. Am. Chem. Soc 2011, 133, 14912–14915. [PubMed: 21866967]

[23]. Rovira AR, Fin A, Tor Y, J. Am. Chem. Soc 2015, 137, 14602–14605. [PubMed: 26523462]

[24]. Sinkeldam RW, McCoy LS, Shin D, Tor Y, Angew. Chem. Int. Ed 2013, 52, 14026–14030.

[25]. Ludford III PT, Rovira AR, Fin A, Tor Y, ChemBioChem 2019, 20, 718–726. [PubMed: 30566279]

[26]. Li Y, Ludford III PT, Fin A, Rovira AR, Tor Y, Chem. Eur. J 2020, 26, 6076–6084. [PubMed: 32157755]

[27]. Defined here as the ability of the analogue to interact or react with an enzyme in a similar capacity to the native counterpart.

[28]. Wang S, Fang K, Dong G, Chen S, Liu N, Miao Z, Yao J, Li J, Zhang W, Sheng C, J. Med. Chem 2015, 58, 6678–6696. [PubMed: 26226379]

[29]. The synthesis of mthA, mthU, mthC, and mthI is described in the Supporting Information (Scheme S3–S4).

[30]. Thewalt U, Bugg CE, Marsh RE, Acta Cryst. 1970, B26, 1089–1101.

[31]. Crystal structure of Uridine: CID6029.

[32]. 2D 1H COSY NMR and 2D 1H NOESY NMR spectra were taken for each of the nucleoside analogues to identify sugar proton interactions and are contained in the Supporting Information.

[33]. Davies DB, Progress in NMR Spectroscopy 1978, 12, 135–225.

[34]. A table containing the error of calculated values is included in the Supporting Information (Table S4).

[35]. In energy units, the bathochromic shifts in emission maxima in water of mthA, mthU, mthC, and mthI relative to their thieno- and isothiazolo-based counterparts are 2400 and 2980 cm–1, 1030 and 2090 cm–1, 1330 and 2350 cm–1, and 1420 and 2370 cm–1, respectively.

[36]. :Gustavsson T, Sarkar N, Bányász Á, Markovitsi D, Improta R, Photochemistry and Photobiology, 2007, 83, 595–599. [PubMed: 17576372]

[37]. Saenger W, Principles of Nucleic Acid Structure; Springer-Verlag: New York, 1984.

[38]. See Figure S3 for absorption and emission spectra of mthA, mthC, mthG, mthU, and mthI in water, dioxane, and mixtures thereof.

[39]. (See: Sinkeldam RW, Tor Y, Org. Biomol. Chem 2007, 5, 2523–2528. [PubMed: 18019524] See also: Sinkeldam RW, Greco N, Tor Y, ChemBioChem 2008, 9, 706–709; [PubMed: 18286575] Sinkeldam RW, Wheat A, Boyaci H, Tor Y, ChemPhysChem 2011, 12, 567–570; [PubMed: 21344595] Sinkeldam RW, Marcus P, Uchenik D, Tor Y, ChemPhysChem 2011, 12, 2260–2265). For a recent study, demonstrating the exquisite environmental sensitivity of thG, [PubMed: 21698743] see: Kuchlyan J, Martines-Fernandes L, Mori M, Gavvala K, Ciaco S, Boudier C, Ludovic R, Didier P, Tor Y, Improta R, Mély Y, J. Am. Chem. Soc 2020, 142, 16999–17014. [PubMed: 32915558]

[40]. :Adamek RN, Ludford P, Duggan SM, Tor Y, Cohen SM, ChemMedChem. 2020, 15, 2151–2156 and [PubMed: 32729197] Ludford PT III, Tor Y, Methods in Enzymology 2020, 639, 71–90.

[41]. Ludford III PT, Li Y, Yang S, Tor Y, Org. Biomol. Chem, 2021, 19, 6237–6243. [PubMed: 34019616]

[42]. Bucardo M, Wu Y, Ludford III PT, Fin A, Tor Y, ACS Chem. Biol, 2021, 16, 1208–1214. [PubMed: 34190533]

[43]. See reference 39 for full analysis.

[44]. Note that the band intensity seen under UV shadowing does not reflect the actual amounts due to differences in absorption maxima and extinction coefficients between the oligonucleotides.

[45]. Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC, Nucleic Acids Res. 1987, 15, 8783–8798. [PubMed: 3684574]

[46]. Milligan JF, Uhlenbeck OC, Methods in Enzymology 1989, 180, 51–62. [PubMed: 2482430]

[47]. Transcription reactions involving T7 RNA Polymerase, the T7 promoter, and a template typically yield 50 equivalents of product or more relative to the template strand. This is due to the enzyme being able to release mature strands and reinitiate transcription at the 5’ end of the template in the T7 promoter region. Triphosphates are typically added in large excess to ensure strand initiation and elongation continues for the lifetime of the protein.

[48]. See supporting information. MALDI TOF MS for the desired products in Lanes 1–8, Lane 1, GTP, observed [M–H]–: 3415.77, calculated: 3413.35; Lane 2, tzGTP, observed [M–H]–: 3484.38, calculated: 3481.35; Lane 3, thGTP, observed [M–H]–: 3480.49, calculated: 3477.23; Lane 4, mthGTP, observed [M–H]–: 3535.63, calculated: 3533.27; Lane 5, G, observed [M–H]–: 3173.47, calculated: 3172.44; Lane 6, tzG, observed [M–H]–: 3189.35, calculated: 3189.44; Lane 7, thG, observed [M–H]–: 3189.62, calculated: 3188.41; Lane 8, mthG, observed [M–H]–: 3203.85, calculated: 3202.42.

[49]. Li Y, Fin A, McCoy L, Tor Y, Angew. Chem. Int. Ed 2017, 56, 1303–1307.See also: Lyon S, Gopalan V, ChemBioChem 2018, 19, 142–146. [PubMed: 29115013]

[50]. Ingale SA, Seela F, J. Org. Chem 2016, 81, 8331–8342. [PubMed: 27529562]

[51]. Zhang A, Kondhare D, Leonard P, Seela F, Chem. Eur. J 2021, 27, 7453–7466. [PubMed: 33443814]

[52]. Rovira AR, Fin A, Tor Y, J. Am. Chem. Soc 2017, 139, 15556–15559. [PubMed: 29043790]

[53]. Hallé F, Fin A, Rovira AR, Tor Y, Angew. Chem. Int. Ed 2018, 57, 1087–1090.

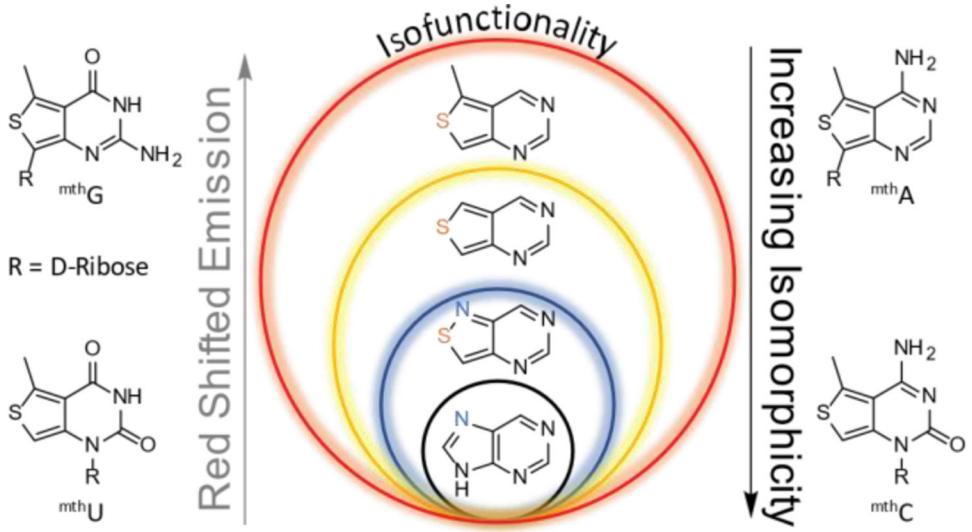



**Figure 1.**

Predicted red shifted emission and diminished structural similarity to the native ribonucleosides associated with the introduction of a methylthieno[3,4-*d*]pyrimidine based ribonucleoside alphabet.

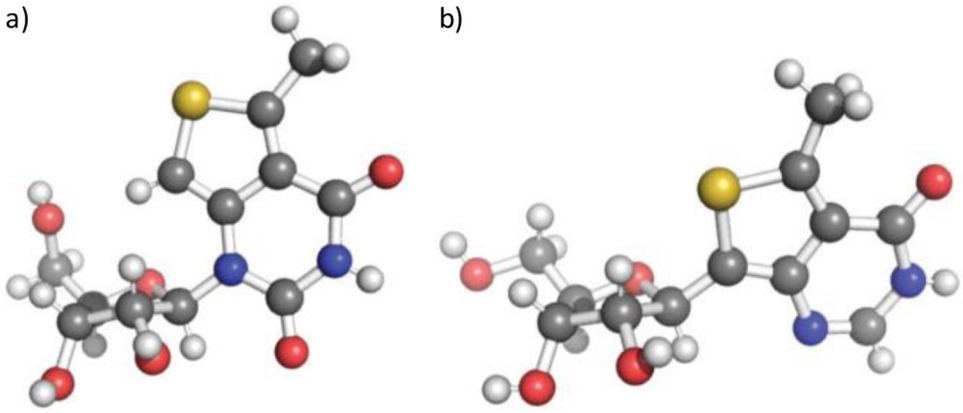



**Figure 2.**
X-ray crystal structures of methylthieno[3,4-*d*]pyrimidine analogues: (a) $^{mth}$**U**, (b) $^{mth}$**I**.

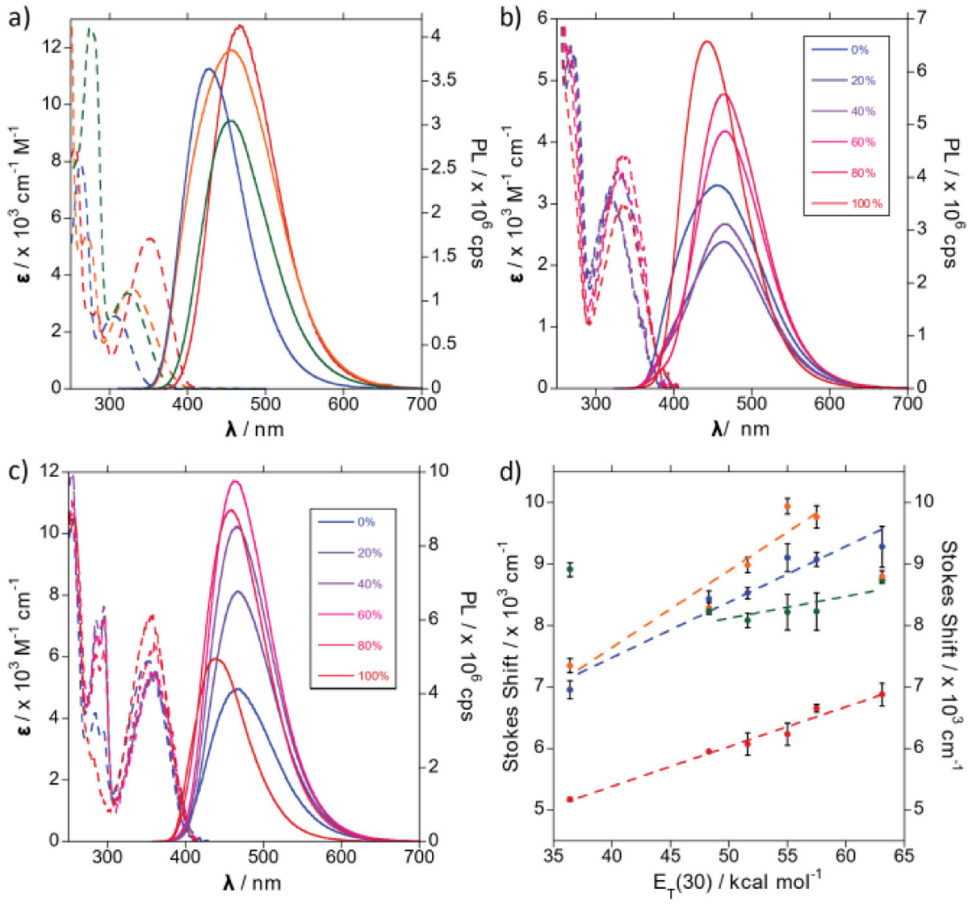



**Figure 3.**

(a) Absorption (dashed lines) and emission (solid lines) spectra of $^{mth}$**A** (red), $^{mth}$**C** (green), $^{mth}$**G** (orange), $^{mth}$**U** (blue) in water. (b) Absorption (dashed lines) and emission (solid lines) spectra of $^{mth}$**G** in water (blue), dioxane (red), and mixtures thereof (20, 40, 60, and 80% v/v dioxane in water). (c) Absorption (dashed lines) and emission (solid lines) spectra of $^{mth}$**A** in water (blue), dioxane (red), and mixtures thereof (20, 40, 60, and 80% v/v dioxane in water). (d) Stokes shift correlation versus solvent polarity ($E_T(30)$) of water/dioxane mixtures for $^{mth}$**A** (red), $^{mth}$**C** (green), $^{mth}$**G** (orange), $^{mth}$**U** (blue).

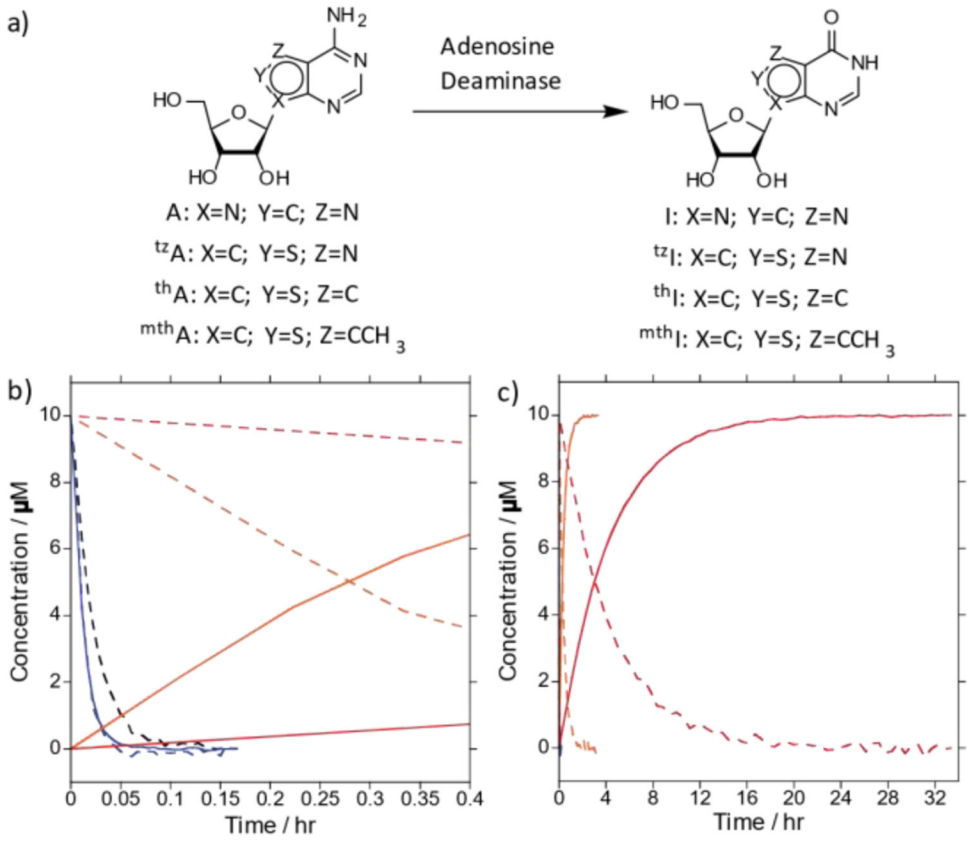



**Figure 4.**

ADA reactions overlay. (a) Deamination of A, $^{tz}$**A**, $^{th}$**A**, and $^{mth}$**A** to I, $^{tz}$**I**, $^{th}$**I**, and $^{mth}$**I** by adenosine deaminase, respectively. (b, c) Enzymatic deamination of A to I (black), $^{tz}$**A** to $^{tz}$**I** (blue), $^{th}$**A** to $^{th}$**I** (orange), and $^{mth}$**A** to $^{mth}$**I** (red) with ADA monitored by absorption (dashed) at 260, 340, 340, and 353 nm, respectively, and monitored by emission at 410, 391, and 420 nm (upon excitation at 322, 318, and 332 nm, respectively), respectively, from (b) 0 to 0.4 h or (c) 0 to 32 h.

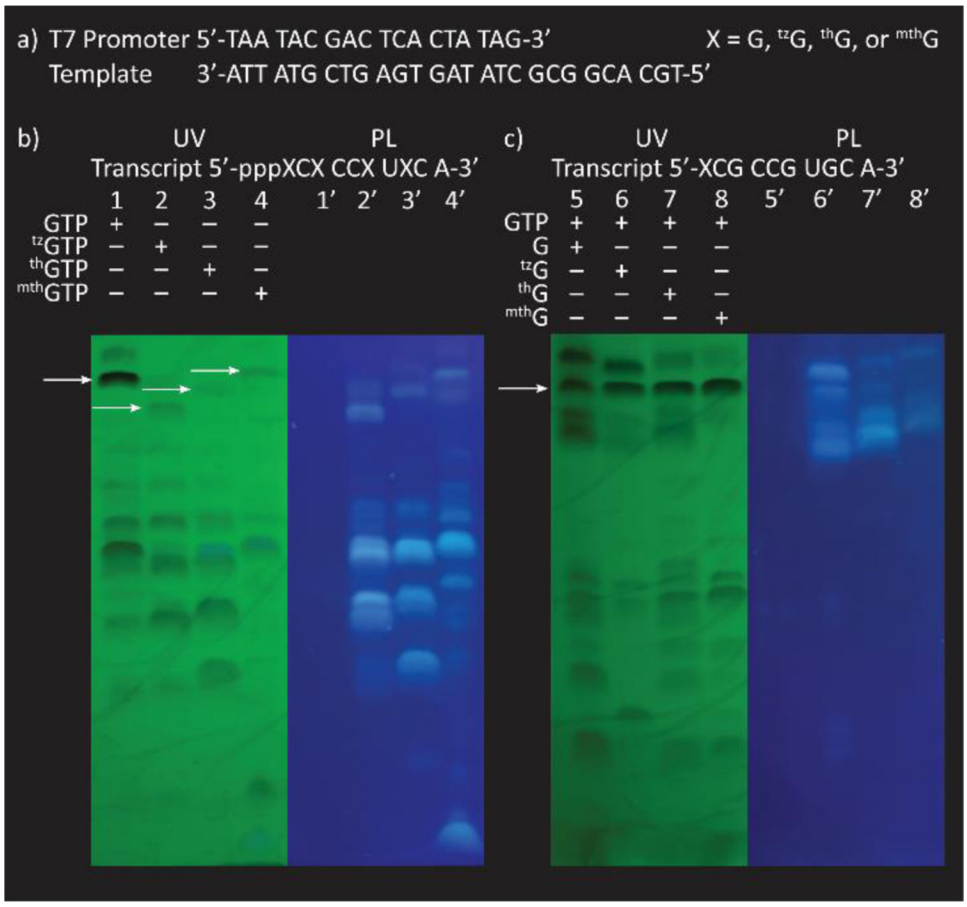
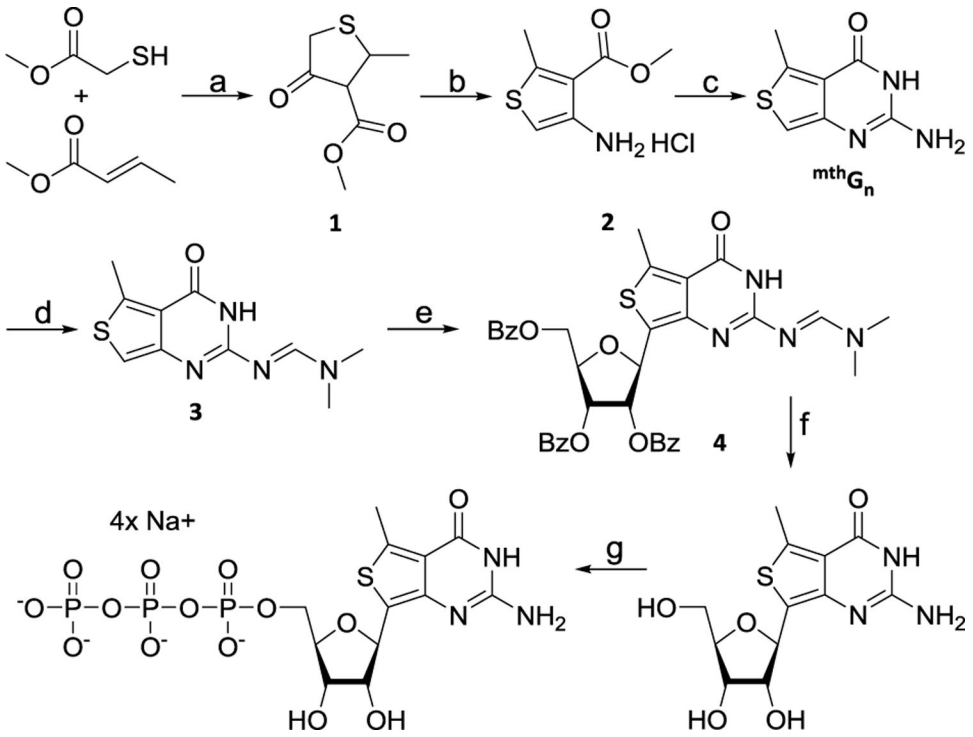



**Figure 5.**

Transcription reactions with T7 RNA Polymerase. a) T7 Promoter and Template strand sequences; b) Transcription reaction using template with 2 mM ATP, UTP, and CTP and 2 mM GTP, **$^{tz}$GTP**, **$^{th}$GTP**, or **$^{mth}$GTP**; c) Transcription reaction using template with 2 mM of all natural NTPs and 10 mM guanosine, **$^{tz}$G**, **$^{th}$G**, or **$^{mth}$G**. The white arrows indicate the target transcript of each reaction. UV shadowing was observed by illumination at 254 nm, photoluminescence (PL) was observed upon excitation at 365 nm.[44]




**Scheme 1.**

Synthetic pathway to the key intermediate **2**, **$^{mth}$G**, and **$^{mth}$GTP**. [a]Reagents and Conditions: (a) i) Piperidine, 50$^{O}$C, 2 h; ii) NaH (60% in mineral oil), THF, 70$^{O}$C, 16 h, 19% over two steps. (b) i) **1**, $BaCO_3$, hydroxylamine hydrochloride, MeOH, 70$^{O}$C, 16 h; ii) 2M HCl in $OEt_2$, $OEt_2$, MeOH, RT, 24 h, 78% over two steps. (c) Chloroformamidine hydrochloride, $DMSO_2$, 125$^{O}$C, 2 h, 99%. (d) Dimethylformamide dimethyl acetal, DMF, RT, 16 h, 95%. (e) β-D-ribofuranose 1-acetate 2,4,5-tribenzoate, tin (IV) chloride, $MeNO_2$, 65$^{O}$C, 16 h, 25%. (f) Ammonia saturated MeOH, 65$^{O}$C, 16 h, 63%. (g) i) $POCl_3$, trimethylphosphate, 0$^{O}$C, 2 h; ii) Bistributylammonium pyrophosphate, tributylamine, DMF, 0$^{O}$C, 1 h; iii) $NaClO_4$ (3 w/v% in acetone), MeOH, 60 seconds, 0$^{O}$C, 5%.

**Table 1.**

Photophysical Properties of Methylthieno[3,4-*d*]pyrimidine Nucleoside Analogues.

| | Solvent | $\lambda_{abs}$(ε)[a] | $\lambda_{em}$(Φ)[a] | Φε | Stokes Shift[a] | Polarity Sensitivity[b] | p$K_a$[c] |
|---|---|---|---|---|---|---|---|
| $^{mth}$G | Water | 327 (3.51) | 456 (0.42) | 1470 | 8650 | 130 | 5.0, 10.4 |
| | Dioxane | 334 (3.60) | 443 (0.61) | 2200 | 7190 | | |
| $^{mth}$A | Water | 353 (5.83) | 467 (0.21) | 1230 | 6920 | 65 | 6.5 |
| | Dioxane | 355 (6.18) | 440 (0.23) | 1660 | 5340 | | |
| $^{mth}$U | Water | 306 (2.55) | 427 (0.30) | 765 | 9260 | 90. | 10.8 |
| | Dioxane | 305 (2.72) | 387 (0.12) | 353 | 6950 | | |
| $^{mth}$C | Water | 323 (3.35) | 455 (0.24) | 804 | 8930 | 36 | 3.9 |
| | Dioxane | 325 (3.15) | 445 (0.03) | 95 | 8300 | | |
| $^{mth}$I | Water | 322 (4.72) | 414 (0.56) | 2660 | 6900 | 69 | 2.0, 11.1 |
| | Dioxane | 322 (5.26) | 382 (0.11) | 580 | 4880 | | |

[a] $\lambda_{abs}$, ε, $\lambda_{em}$, and Stokes Shift are reported in units of nm, $M^{-1}$ $cm^{-1}$, nm, and $cm^{-1}$ respectively.

[b] Sensitivity to polarity reported in $cm^{-1}$/(kcal $mol^{-1}$) is equal to the linear fit in Figure 3d.

[c] p$K_a$ values reflect the average over three independent measurements and are equal to the inflection point determined by the fitting curves in Figure S4. See Table S4 for expanded data including experimental errors.

**Table 2.**

Kinetic Properties of Enzymatic Reactions of A, $^{tz}$**A**, $^{th}$**A**, and $^{mth}$**A** with ADA.

| | Alphabet | $k_{app}$ (× $10^{-4}$ $s^{-1}$) | $t_{1/2}$ (× $10^{2}$ s) |
|---|---|---|---|
| Abs | A | 120 ± 10 | 0.57 ± 0.01 |
| | $^{tz}$A | 220 ± 10 | 0.32 ± 0.02 |
| | $^{th}$A | 7.0 ± 0.5 | 9.8 ± 0.6 |
| | $^{mth}$A | 0.64 ± 0.09 | 110 ± 20 |
| Em | $^{tz}$A | 220 ± 10 | 0.31 ± 0.02 |
| | $^{th}$A | 7.0 ± 0.5 | 9.6 ± 0.2 |
| | $^{mth}$A | 0.64 ± 0.06 | 110 ± 10 |